Koninklijk Nederlands Meteorologisch Instituut
Ministerie van Infrastructuur en Waterstaat

VU VRIJE UNIVERSITEIT AMSTERDAM

Improving sub-seasonal summer temperature forecasts using ML

Maurice Schmeits
KNMI, R&D Weather and climate modeling
VU/IVM, Water and climate risk

Thanks to **Chiem van Straaten**

NWO-project

Title: Improvement of sub-seasonal probabilistic forecasts of European high-impact weather events using machine learning techniques

Period: 2018-2022

PI: Maurice Schmeits

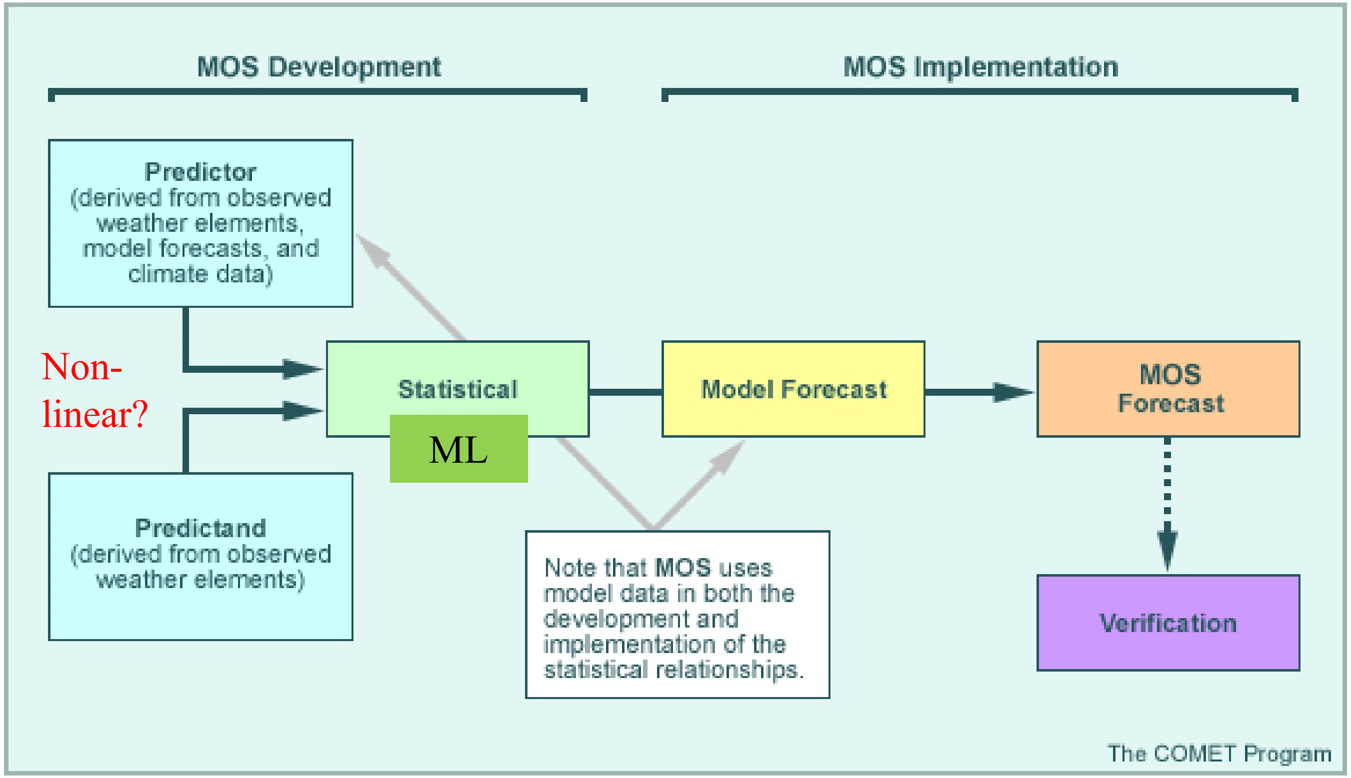
Co-supervisors: Kirien Whan, Dim Coumou, and Bart van den Hurk

PhD student: Chiem van Straaten

Outline

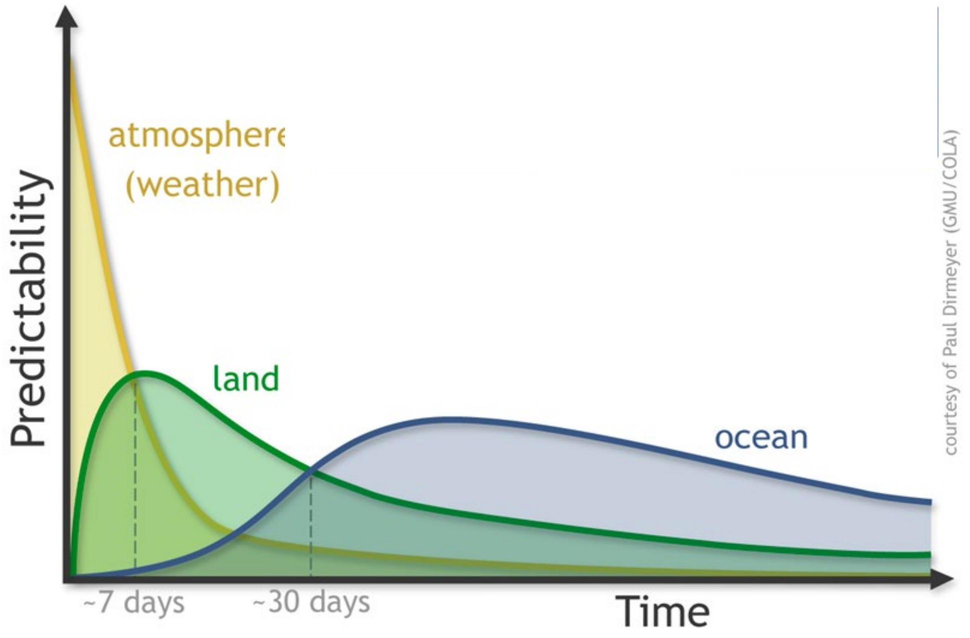
	Project 1 (van Straaten et al, 2022)	**Project 2 (paper in preparation)**
Forecasts	empirical	Post-processing (NWP & empirical)
ML	Random Forests	Shallow neural network
Data	ERA5 (1981-2019)	ERA5 & ECMWF IFS reforecasts (1998-2019)
eXplainable AI	yes	yes
Goal	Discover sources of predictability	Improve skill of IFS, discover IFS shortcomings

Model output statistics (MOS)



Project 1:
Framework

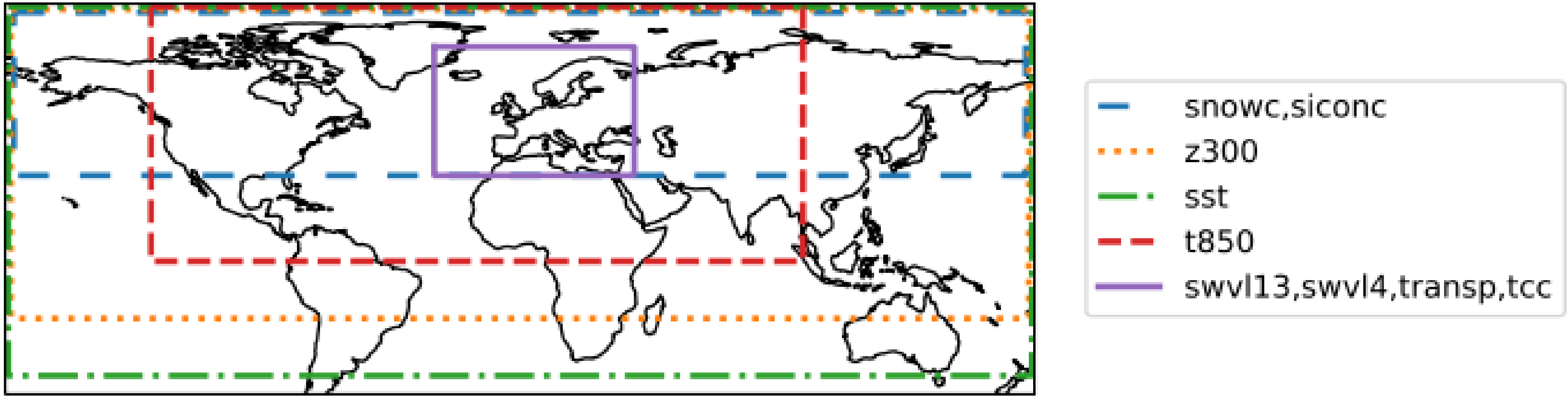
Using explainable machine learning forecasts to **discover sub-seasonal drivers** of high summer temperatures in western and central Europe.

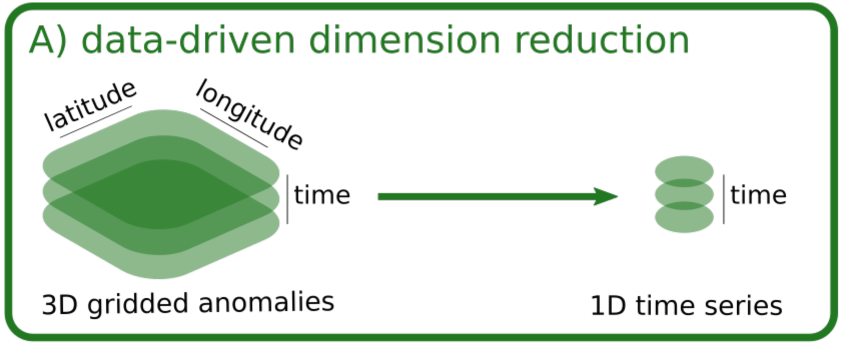


Climate variable → feature → source of predictability

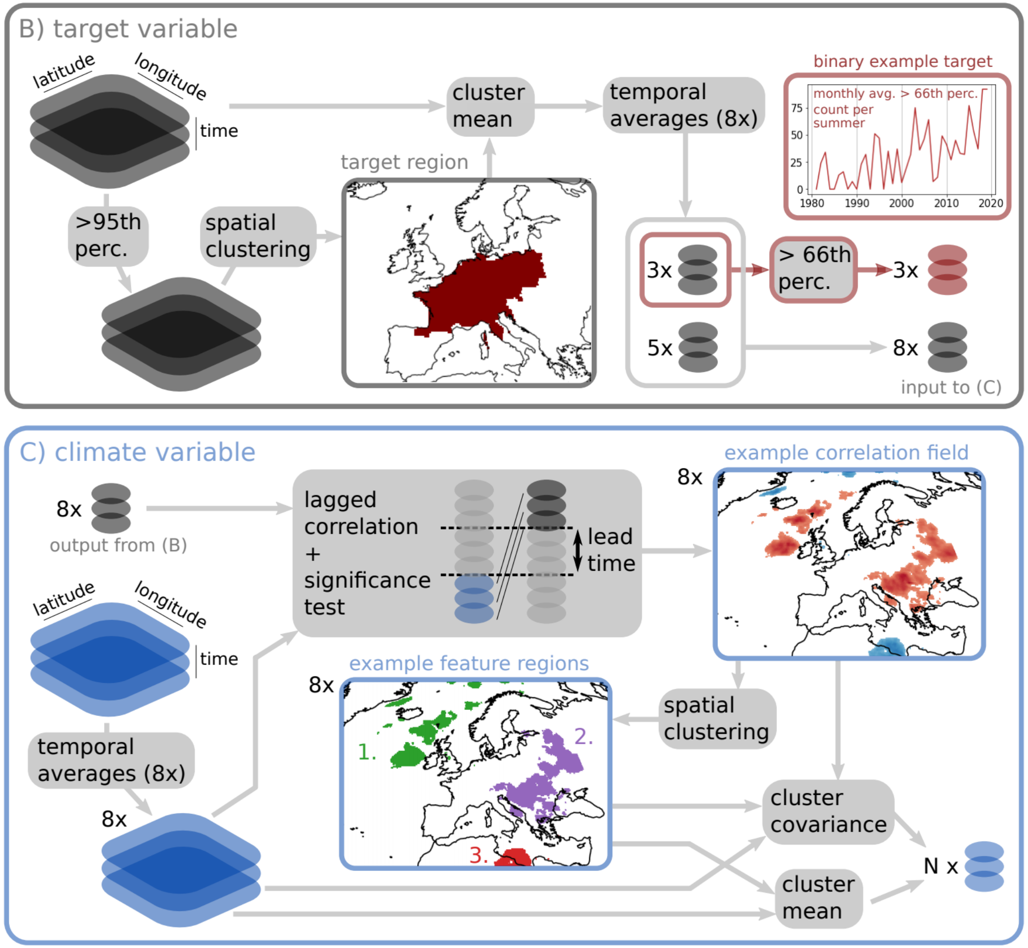
Project 1:
Data

ERA5 & ERA5-Land

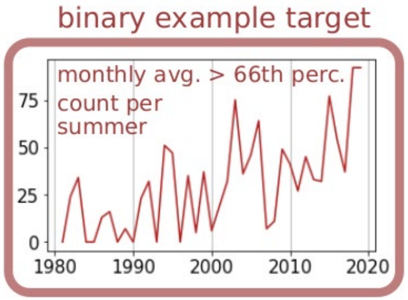


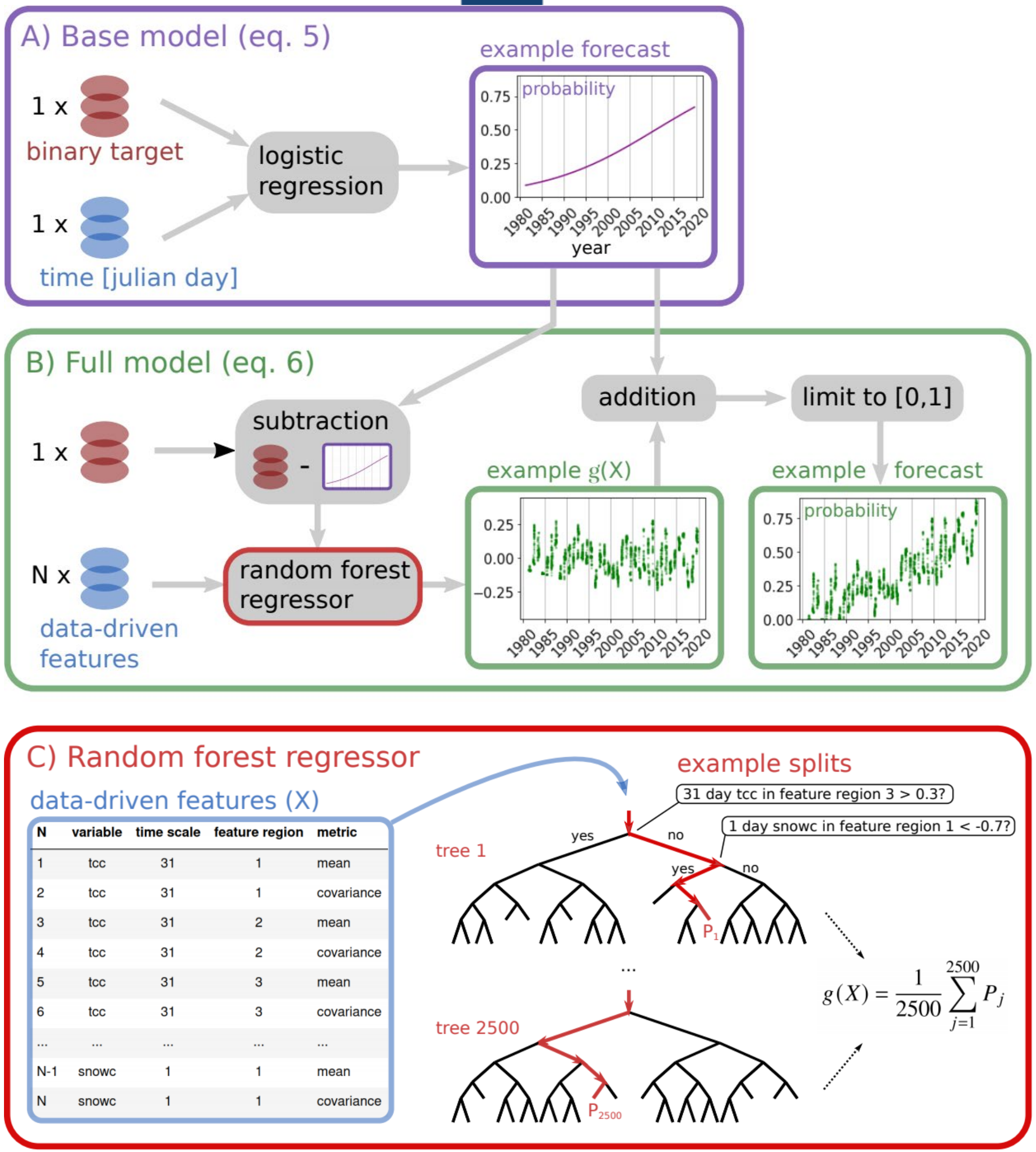


Project 1: Dimension reduction



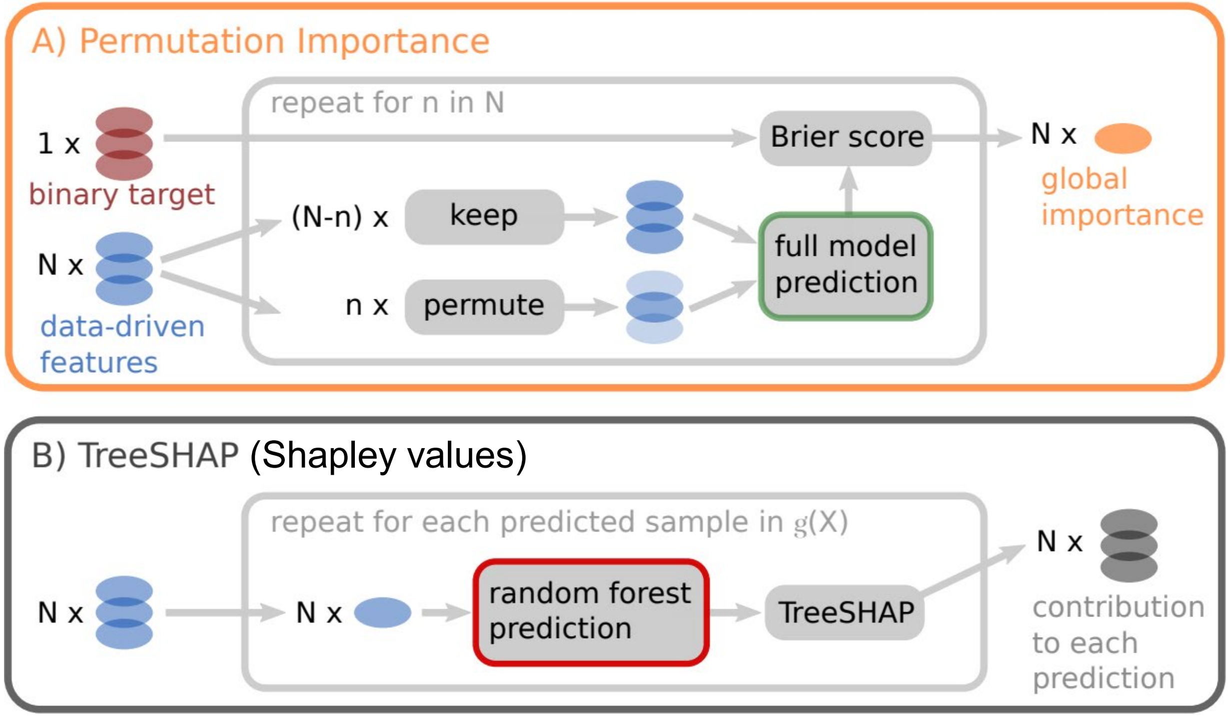
Project 1: ML models



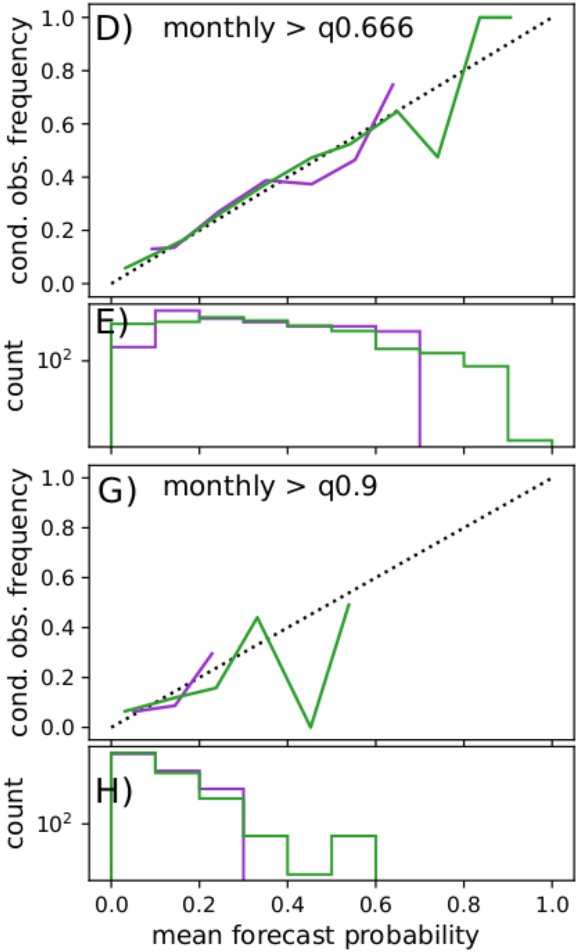


$$p_{full} = p_{base} + g(X) \quad \text{with } p \in [0,1]$$

Project 1: ML explainability



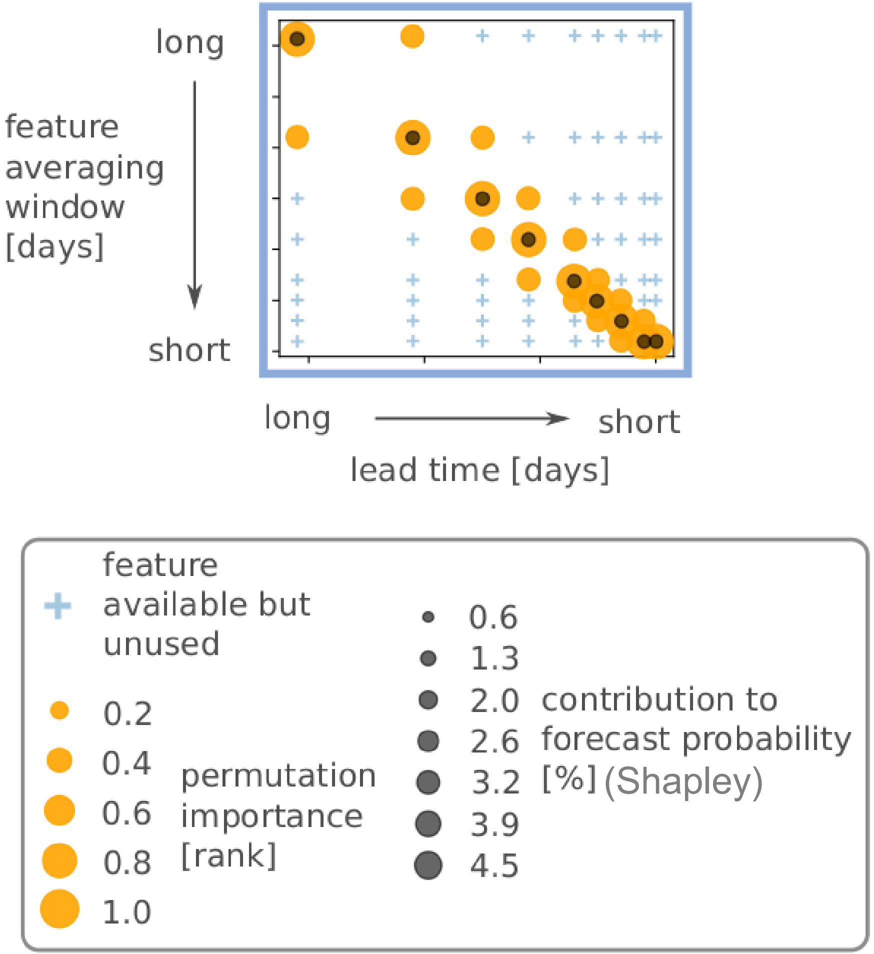
Project 1: Verification



Target: monthly temperature > q….
Lead time: 15 days
Metric: Reliability diagram

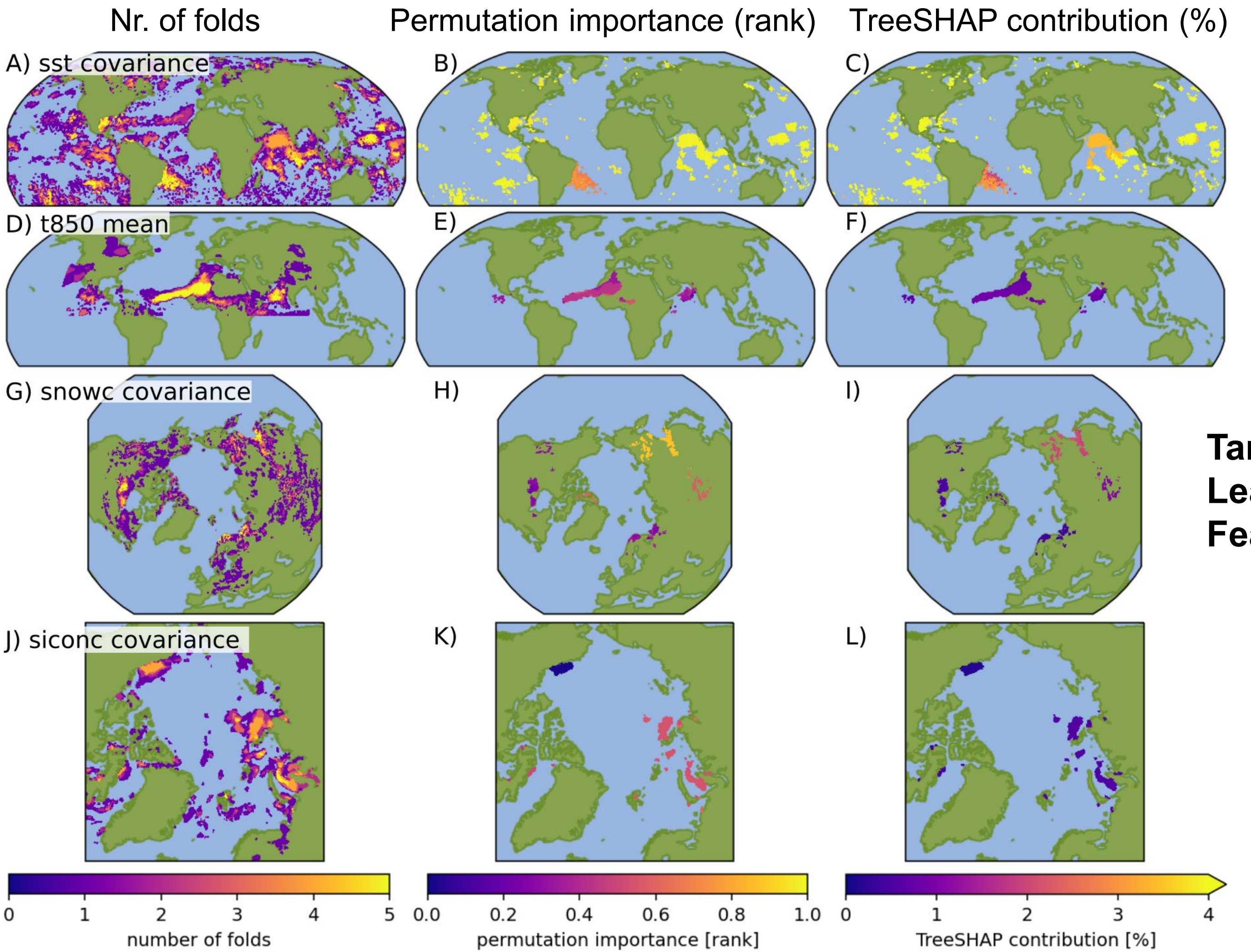
Project 1: Sources of predictability

Example



Project 1: Sources of predictability

Project 1: Sources of predictability



Target: monthly temperature > q0.66
Lead time: 15 days
Feature time scale: monthly

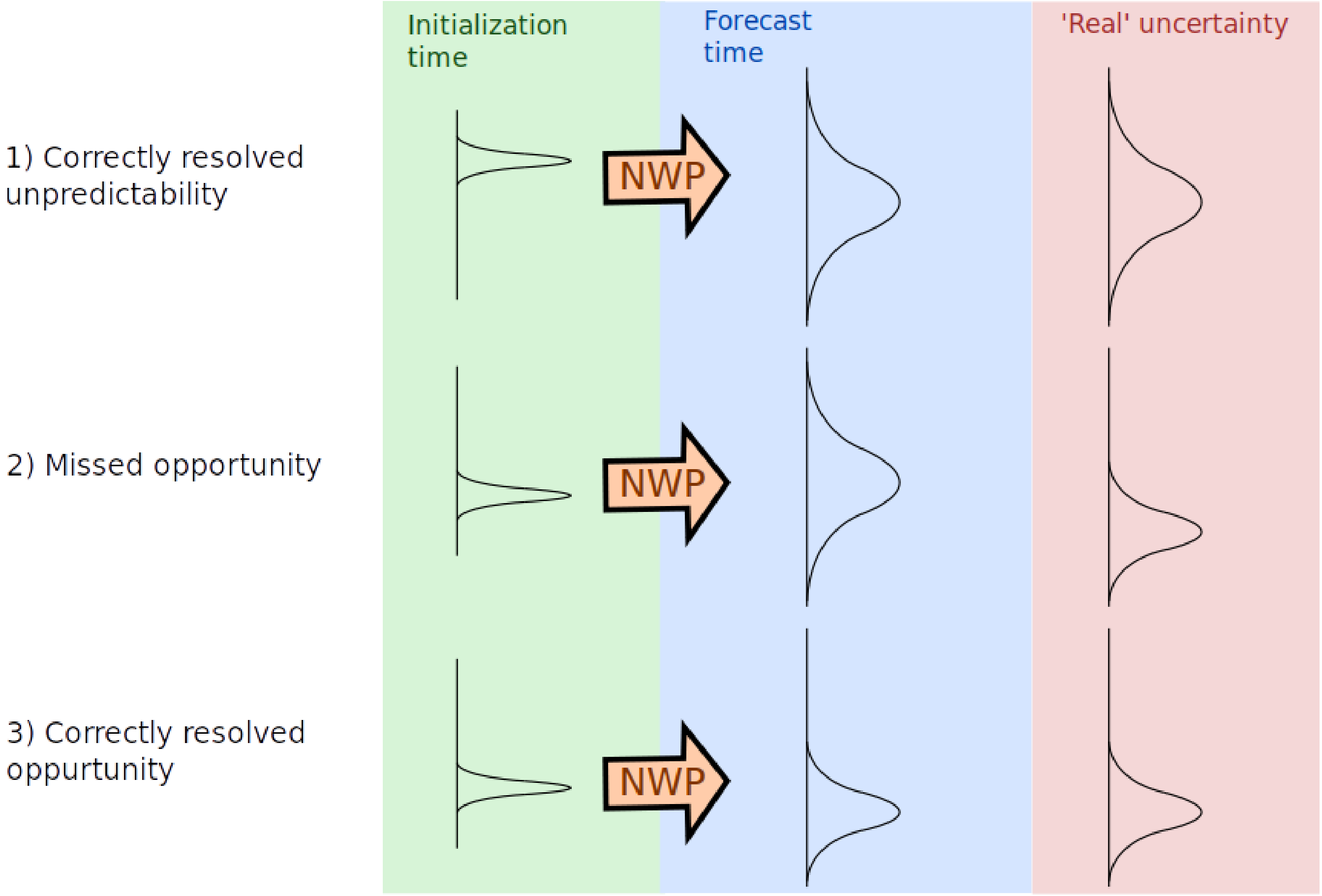
Project 1: Conclusion + challenges

- Identified several sources of predictability in a data-driven way

- Trend
  - provides predictability
  - challenge for random forests

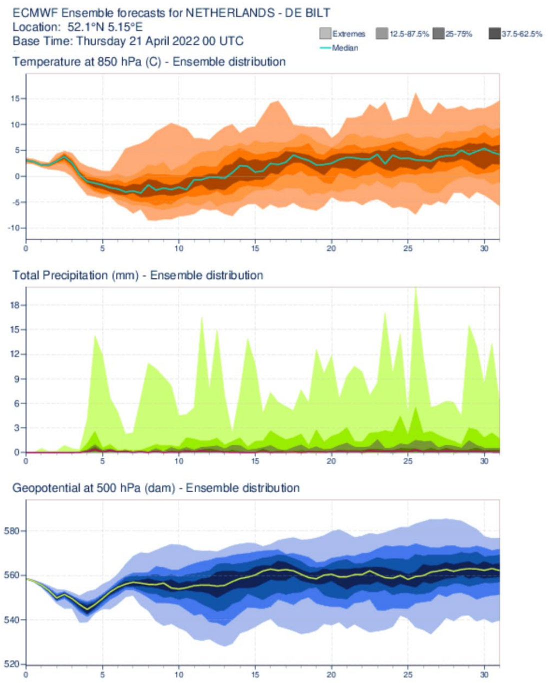
- Variability from fold to fold
  - limited data increases chance of fitting spurious links

Project 2: Framework



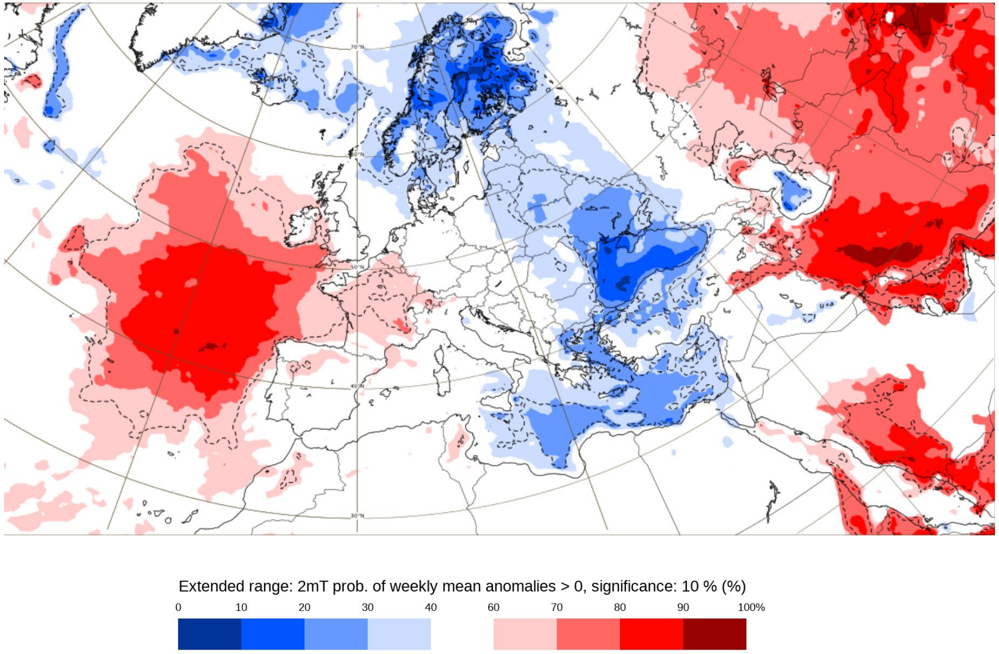
Project 2: ECMWF ensemble forecasts

Monthly forecast plumes - Extended range forecast



2m temperature: Probability of weekly anomaly > 0 (week 3)

Base time: Thu 21 Apr 2022 Valid time: Mon 09 May 2022 - Mon 16 May 2022 (+600h) Area : Europe

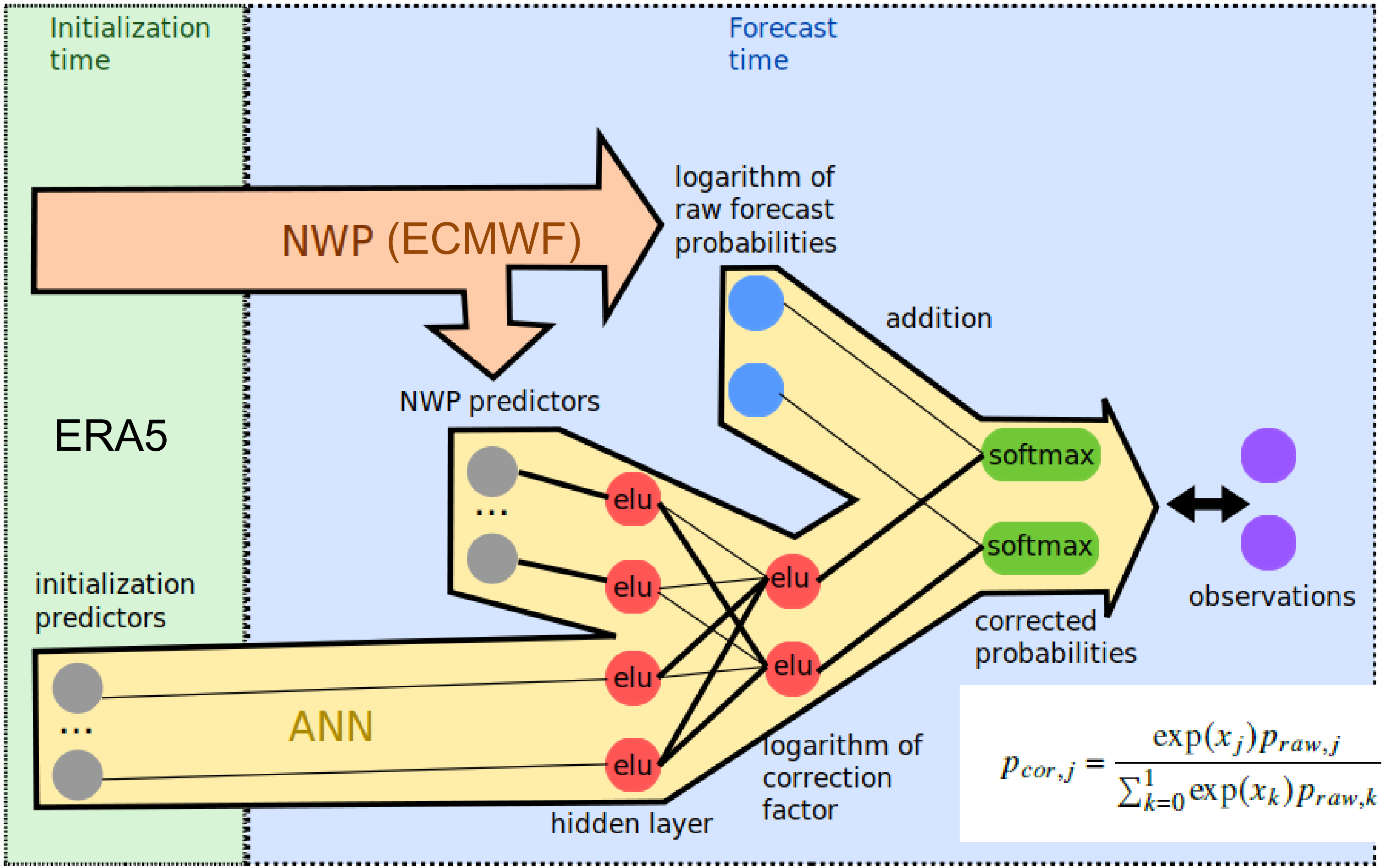


ECMWF

ropean Centre for Medium-Range Weather Forecasts (ECMWF)
ww.ecmwf.int
C-BY-4.0 and ECMWF Terms of Use(https://apps.ecmwf.int/datasets/licences/general/)

ECMWF

Project 2: ANN-based post-processing



$$p_{cor,j} = \frac{\exp(x_j)p_{raw,j}}{\sum_{k=0}^{1}\exp(x_k)p_{raw,k}}, \quad i = 0, 1.$$

Inspired by Scheuerer et al. (2020)

Project 2: ANN training

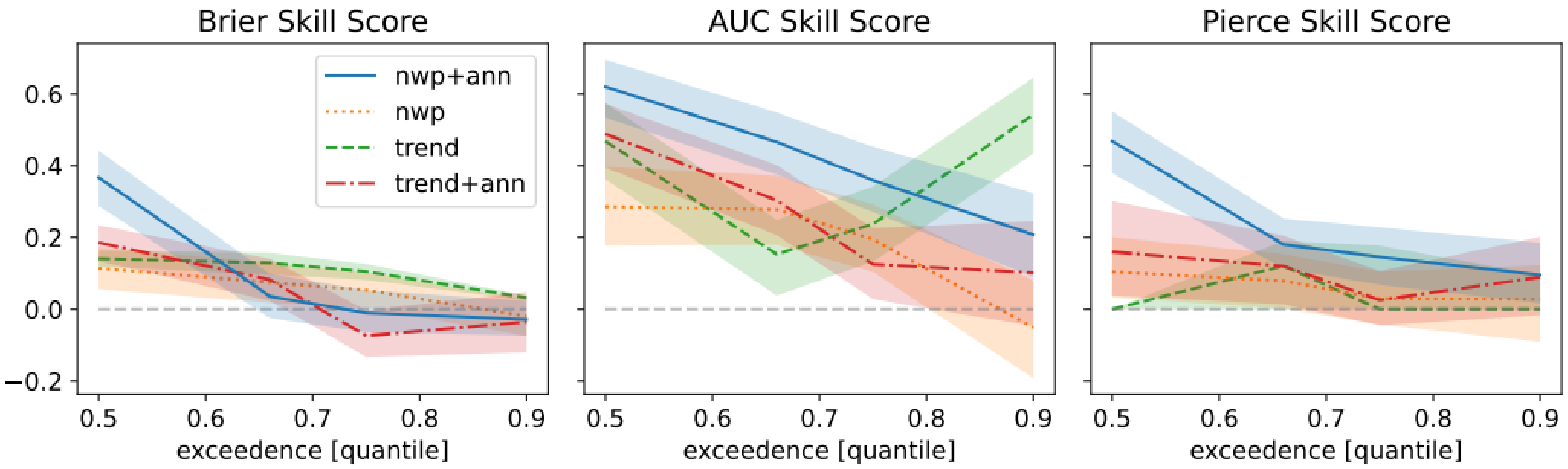
3-fold cross-validation

Hyper parameters

param	value
batch size	32
early stop patience	7
epochs	200
learning rate	0.0014
n hidden layers	1
n hidden nodes	4

Project 2: Verification (incl. benchmarks)

Target: monthly temperature in western Europe > … quantile
Lead time: 12-15 days

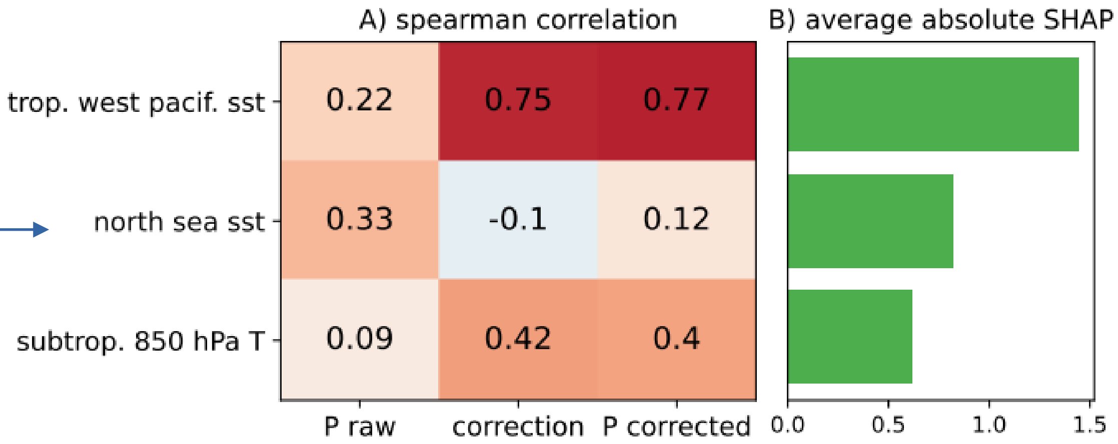


Project 2: Selected predictors

Target: monthly temperature in western Europe > 0.5 quantile
Lead time: 12-15 days

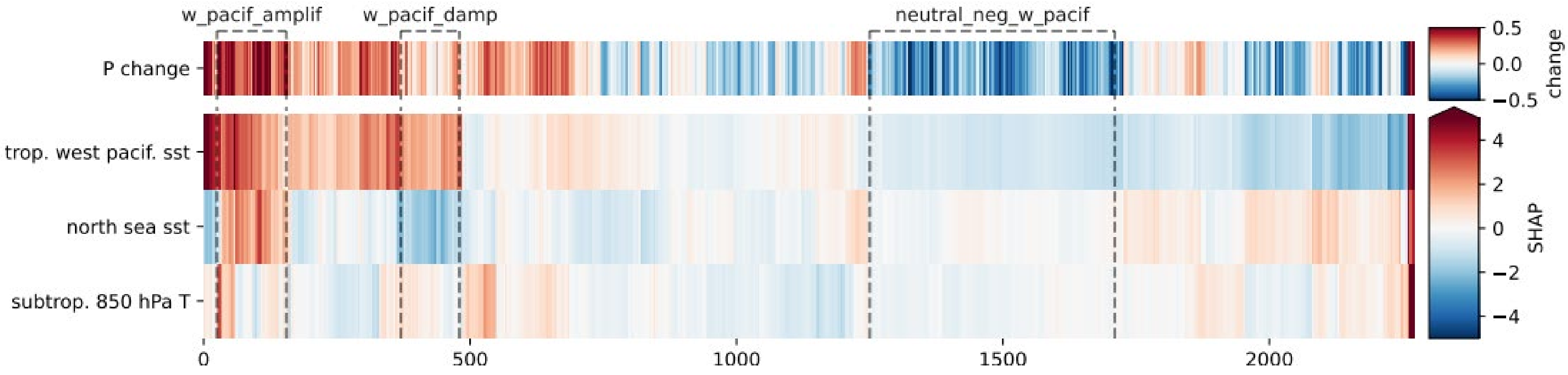
Predictors from:

- Initialization
- Forecast time



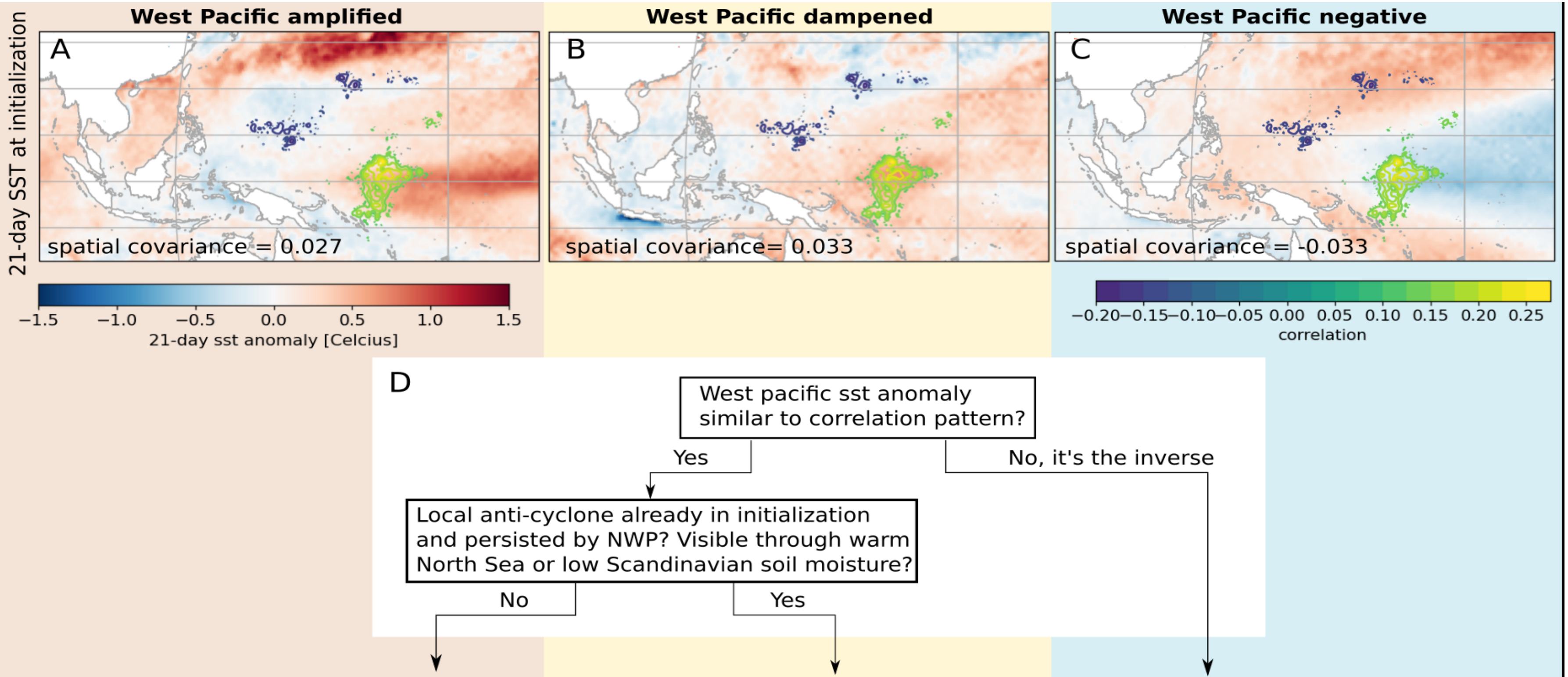
Project 2: Missed opportunities

Target: monthly temperature in western Europe > 0.5 quantile
Lead time: 12-15 days



Samples (ordered by hierarchical clustering)

Missed opportunities



	Under-estimation	More or less correct	Over-estimation
mean P raw	0.53	0.72	0.52
Observed freq.	0.95	0.8	0.18
n samples	130	110	460

Project 2: Conclusion + challenges

- ANN improves forecast skill
  - not for extreme targets (but tropical western Pacific SST pattern is also important for those)

- eXplainable AI shows ‘situations requiring the same correction for the same reason’
  - We discovered a missing teleconnection in the ECMWF model

- How does tropical western Pacific SST pattern influence temperature in western and central Europe?
  - follow-up study

- More elegant incorporation of trend
  - difficulty extrapolating

Thank you for your attention!

Questions?

References:

Scheuerer, M., Switanek, M. B., Worsnop, R. P., & Hamill, T. M. (2020). Using artificial neural networks for generating probabilistic subseasonal precipitation forecasts over California. Monthly Weather Review, 148(8), 3489-3506. https://doi.org/10.1175/MWR-D-20-0096.1

van Straaten, C., Whan, K., Coumou, D., van den Hurk, B., & Schmeits, M. (2022). Using explainable machine learning forecasts to discover sub-seasonal drivers of high summer temperatures in western and central Europe. Monthly Weather Review. https://doi.org/10.1175/MWR-D-21-0201.1