Identifying oceanic-atmospheric controls on hydrology using a machine learning approach

Laura Crocetti[1], Wouter Dorigo[1], Brecht Martens[2], Peter Filzmoser[3]

1) CLIMERS – Climate and Environmental Remote Sensing Research Group, Department of Geodesy and Geoinformation, TU Wien

2) Laboratory of Hydrology and Water Management, Ghent University

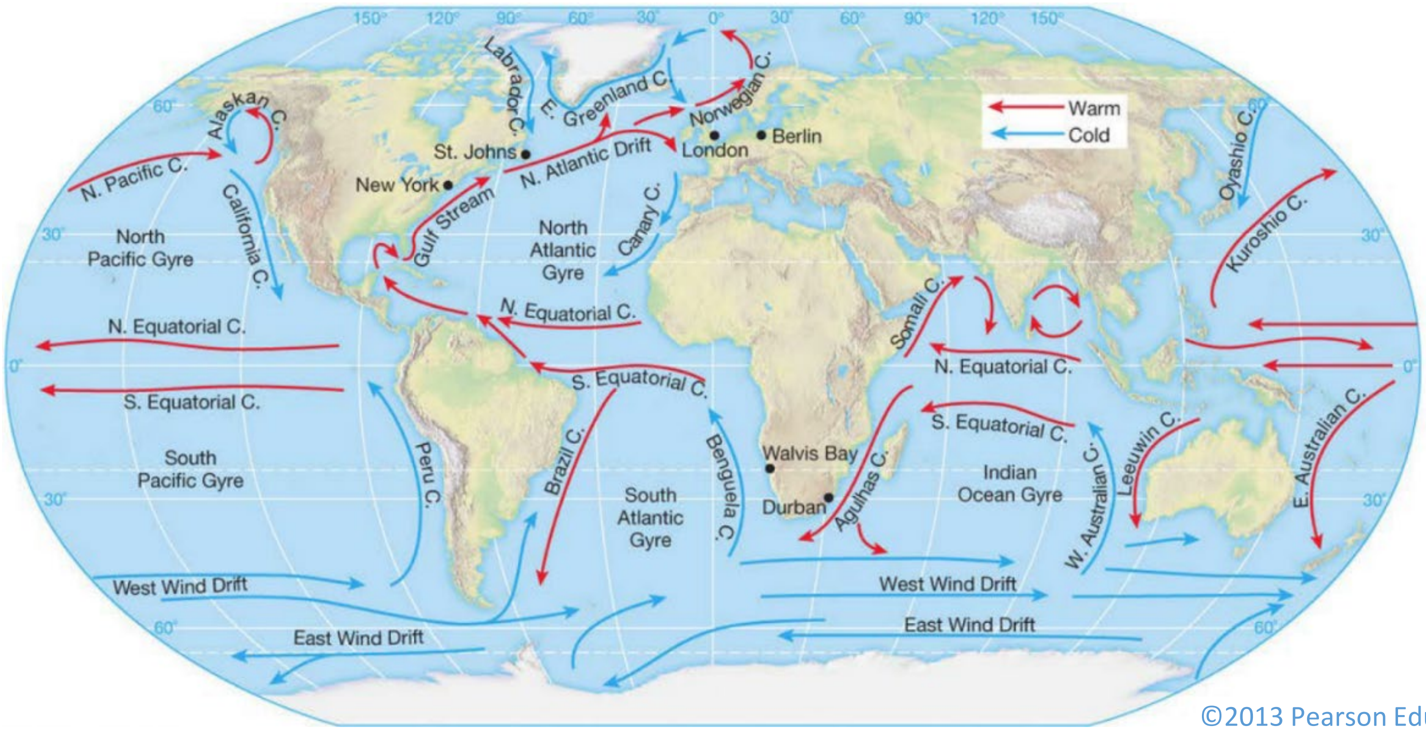
3) Computational Statistics, Institute of Statistics and Mathematical Methods in Economics, TU Wien

laura.crocetti@geo.tuwien.ac.at

Photography encouraged

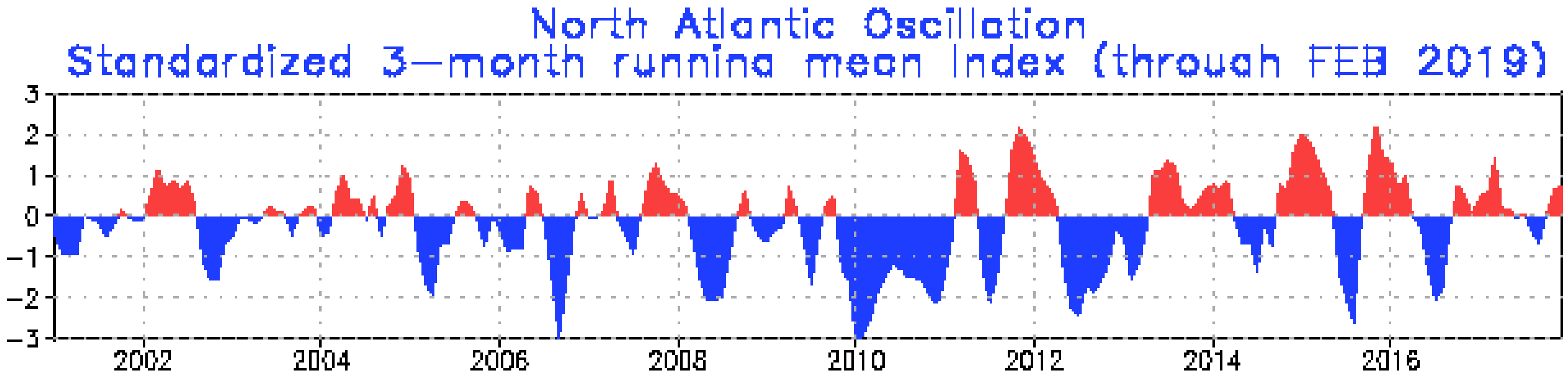
Picture: © NASA Visible Earth

- Climate variability
- Ocean and atmosphere are strongly coupled

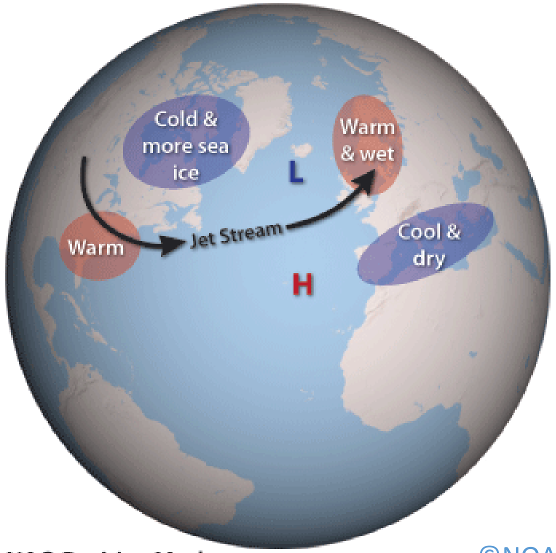


©2013 Pearson Education, Inc.

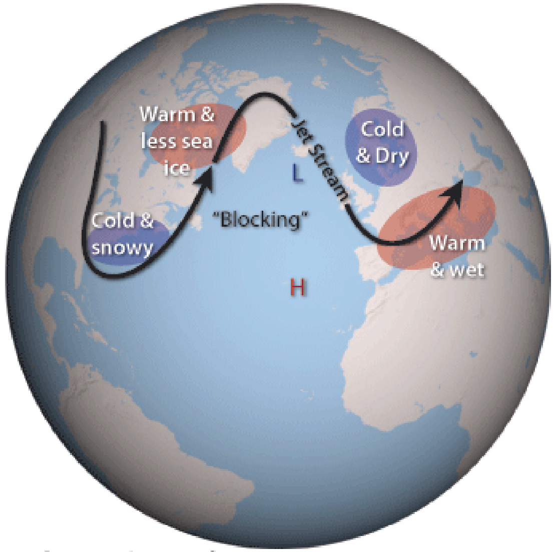
- Climate variability
- Ocean and atmosphere are strongly coupled
- Dominant climate mode in the North Atlantic
  → **N**orth **A**tlantic **O**scillation (NAO)
- NAO index: difference of normalized surface pressure between Lisbon, Portugal and Reykjavik, Iceland



©Climate Prediction Center



NAO Positive Mode ©NOAA



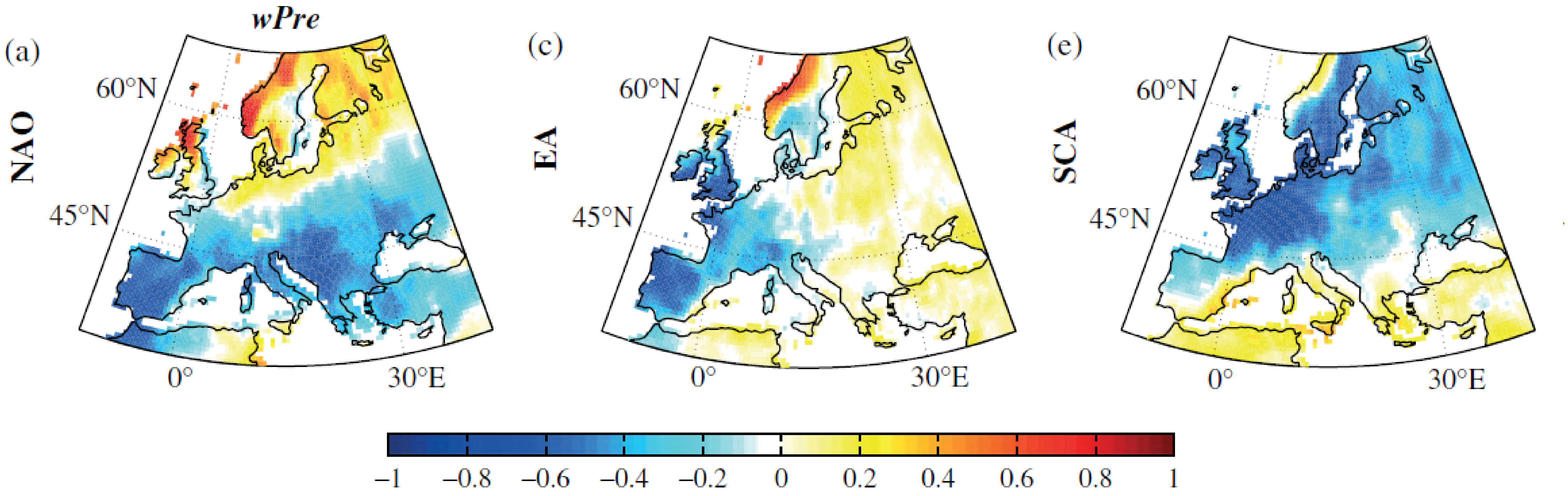
NAO Negative Mode

laura.crocetti@geo.tuwien.ac.at

EGU cc BY

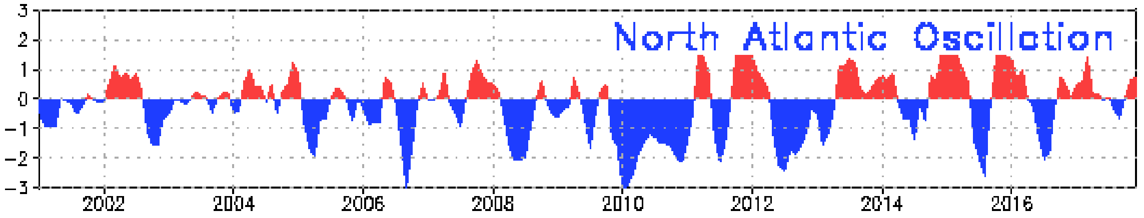
Previous studies

- Correlations between winter precipitation and climate modes **North Atlantic Oscillation, East Atlantic Pattern, Scandinavian Pattern** *[Comas-Bru and McDermott, 2013]*

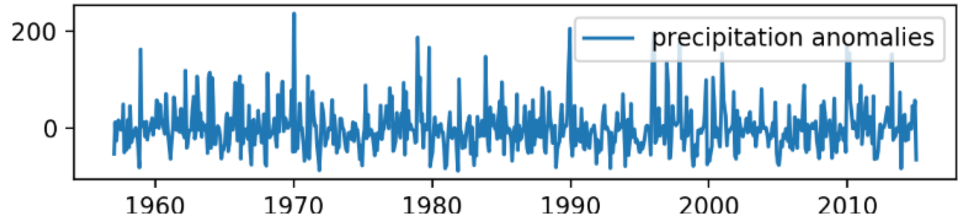


- Reveal dominant modes of climate variability controlling hydrology

Climate modes



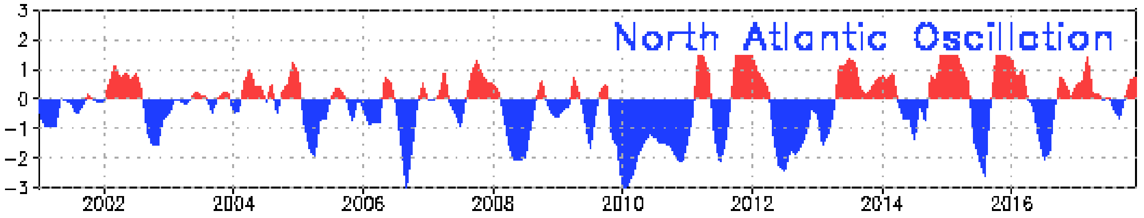
Precipitation anomalies



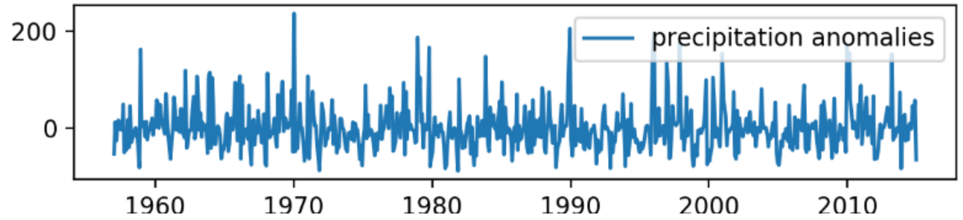
laura.crocetti@geo.tuwien.ac.at

- Reveal dominant modes of climate variability controlling hydrology

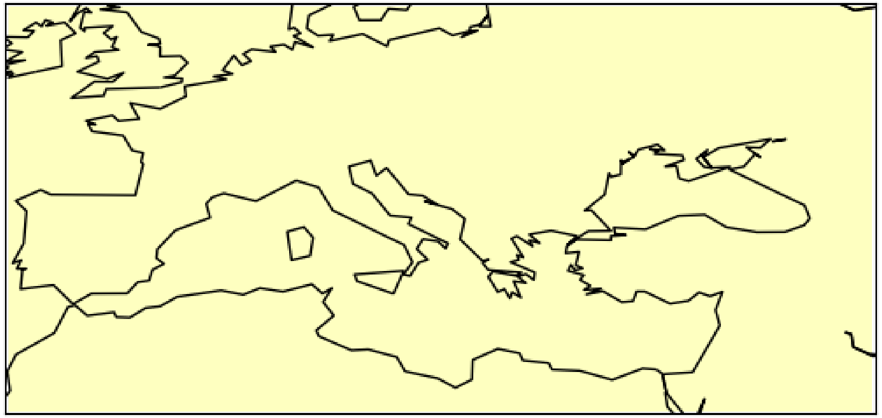
Climate modes



Precipitation anomalies

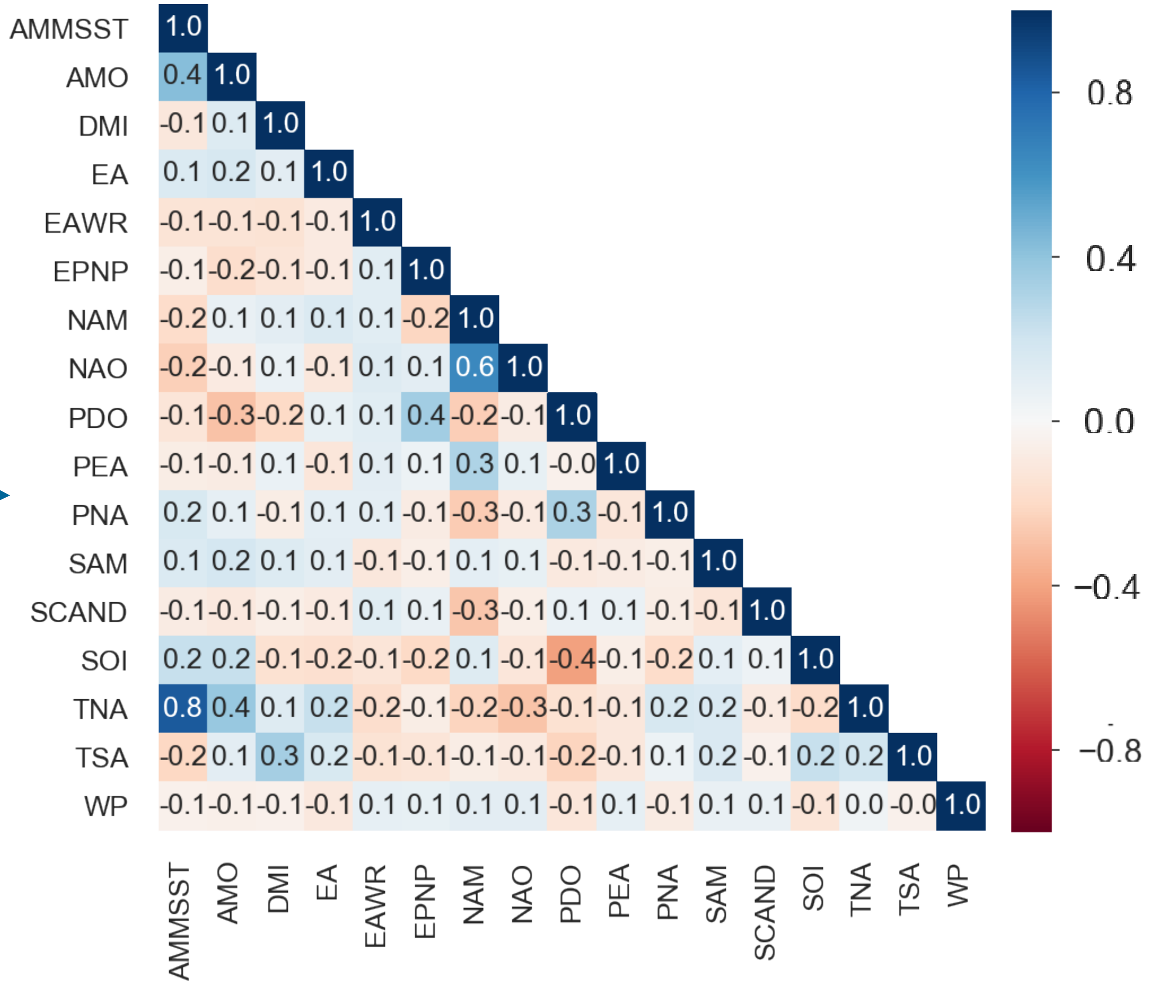


- Extended Mediterranean area
  Latitude: 28.5°N – 56.5°N
  Longitude: 10°W – 46°E



Objective of this study

- Some climate modes are highly correlated



- Supervised learning approach: **LASSO regression**
  = **l**east **a**bsolute **s**hrinkage and **s**election **o**perator

- LASSO regression
  = **l**east **a**bsolute **s**hrinkage and **s**election **o**perator

$$\hat{\beta} = argmin\left\{\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \alpha\sum_{j=1}^{p}\left|\beta_j\right|\right\}$$

$\hat{\beta}$… p-dimensional vector with the estimated regression coefficients

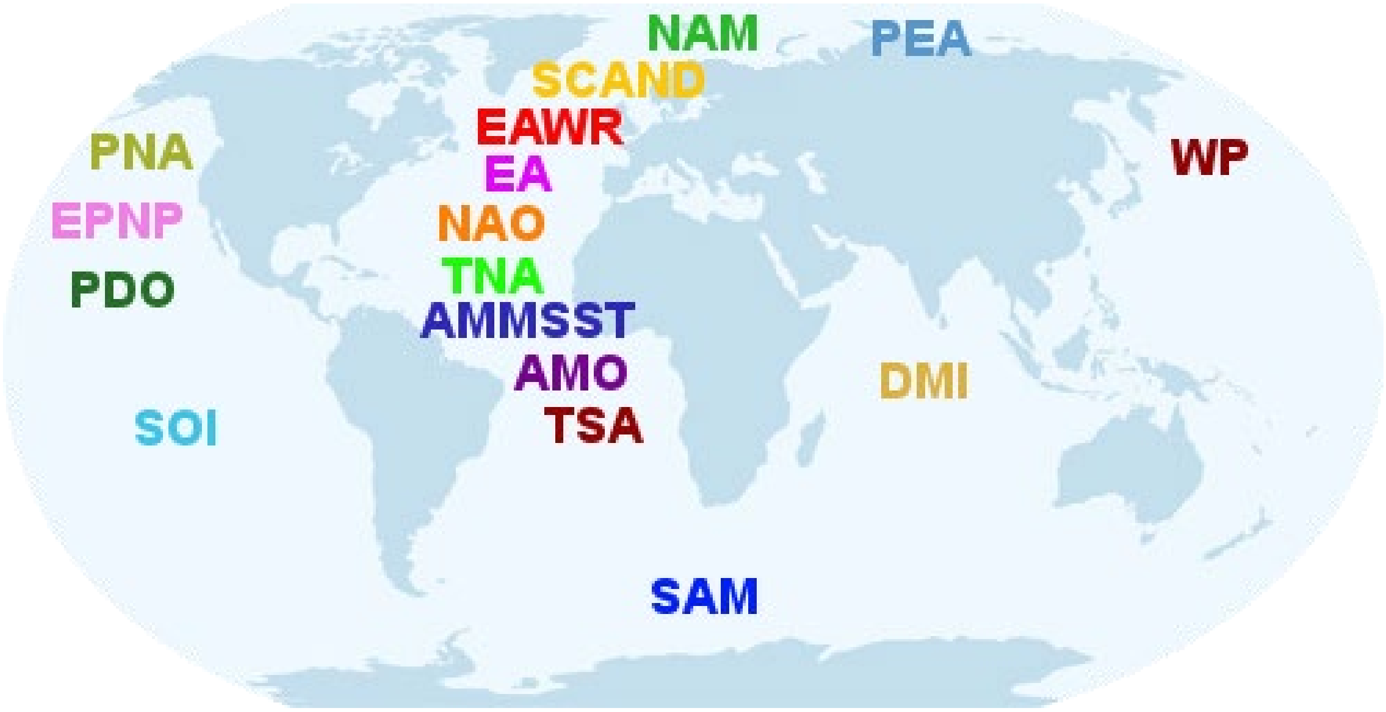
n … number of training samples in the dataset
y_i … value of the target variable in sample i
p … number of features
x_{ij} … value of feature j in sample i
$\alpha\sum_{j=1}^{p}\left|\beta_j\right|$… regularization → prevents overfitting by restricting the model, typically to reduce its complexity

→ α controls the amount of regularization

EGU CC BY

Climate modes

- 17 major climate modes = **predictive features x_i**
- 6 time lags between zero and five months → **102 features**
- Temporal resolution: monthly



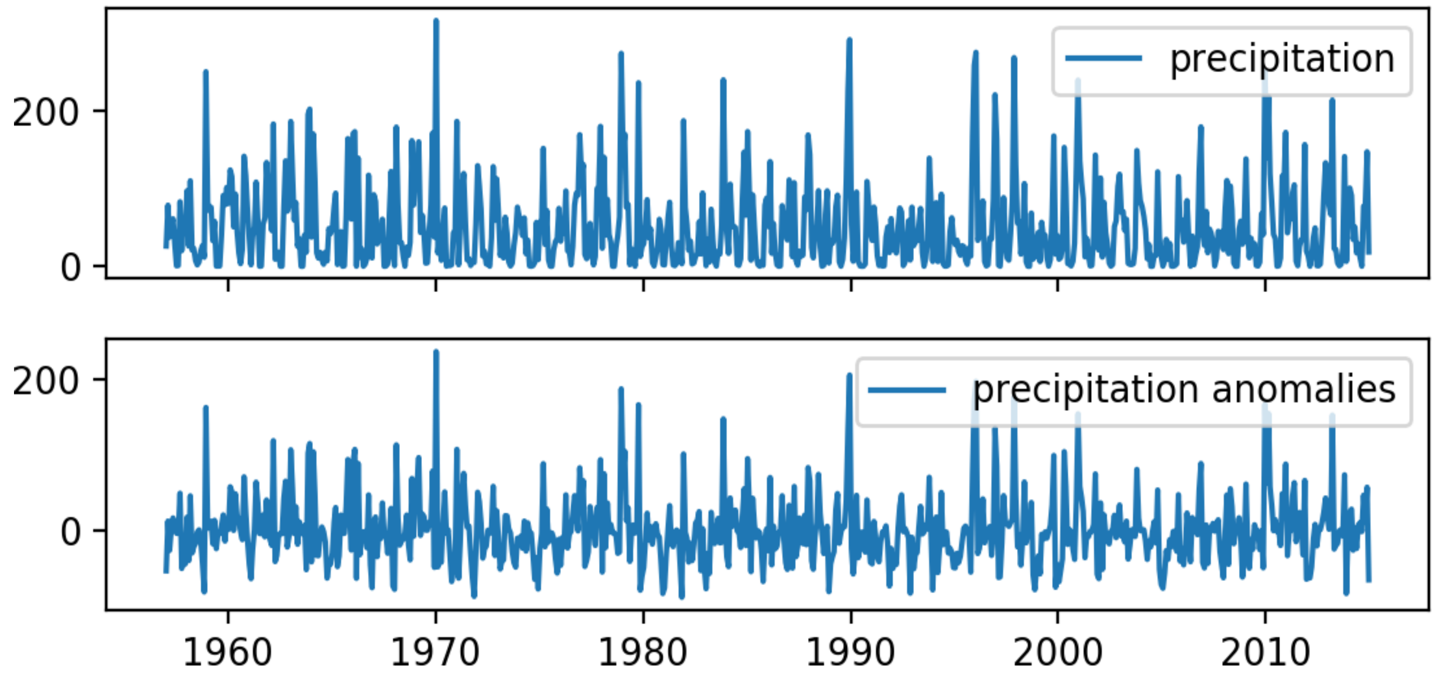
CLIMATE OSCILLATION INDICES
Atlantic Meridional Mode (**AMMSST**) index
Atlantic Multidecadal Oscillation (**AMO**) index
Dipole Mode Index (**DMI**)
East Atlantic (**EA**) index
East Atlantic / West Russia (**EAWR**) index
East Pacific-North Pacific (**EPNP**) index
Northern Annular Mode (**NAM**) index
North Atlantic Oscillation (**NAO**) index
Pacific Decadal Oscillation (**PDO**) index
Polar / Eurasia (**PEA**) index
Pacific / North American Index (**PNA**) index
Southern Annular Mode (**SAM**) index
Scandinavian Pattern (**SCAND**) index
Southern Oscillation Index (**SOI**)
Tropical Northern Atlantic (**TNA**) index
Tropical Southern Atlantic (**TSA**) index
Western Pacific (**WP**) index

laura.crocetti@geo.tuwien.ac.at

Precipitation

- Precipitation anomalies = **target variable y**
- Provided by Climate Research Unit (**CRU**) TS v.3.23
- Temporal resolution: monthly
- Spatial resolution: 0.5°

CRU precipitation and precipitation anomalies



laura.crocetti@geo.tuwien.ac.at

Experimental set up I

- LASSO regression model
- **Target y:** CRU precipitation anomalies
- **Features X:** 17 COIs · 6 time lags = 102 features
- Overlapping time period: 1957-2014

			1st COI						2nd COI					17th COI			
			lag 0	lag 1	lag 2	lag 3	lag 4	lag 5	lag 0	lag 1	…	lag 5	…	lag 0	lag 1	…	lag 5
	Jun		Jun	May	Apr	Mar	Feb	Jan	Jun	May	…	Jan	…	Jun	May	…	Jan
	Jul		Jul	Jun	May	Apr	Mar	Feb	Jul	Jun	…	Feb	…	Jul	Jun	…	Feb
	Aug		Aug	Jul	Jun	May	Apr	Mar	Aug	Jul	…	Mar	…	Aug	Jul	…	Mar
y	Sep	X	Sep	Aug	Jul	Jun	May	Apr	Sep	Aug	…	Apr	…	Sep	Aug	…	Apr
	Oct		Oct	Sep	Aug	Jul	Jun	May	Oct	Sep	…	May	…	Oct	Sep	…	May
	Nov		Nov	Oct	Sep	Aug	Jul	Jun	Nov	Oct	…	Jun	…	Nov	Oct	…	Jun
	…		…	…	…	…	…	…	…	…	…	…	…	…	…	…	…

EGU CC BY

Experimental set up II

TU WIEN GEO

- Two five-fold cross-validations (α ,R^2)
- Significance test using Benjamini-Hochberg procedure

- Seasonal models
  - Winter model: December, January, February
  - Spring model: March, April, May
  - Summer model: June, July, August
  - Autumn model: September, October, November

- Including 3x3 neighbourhood

Coefficient of determination R^2

$$R^2 = corr(y, y_{pred})^2$$

→ coefficient of determination R^2

TARGET VARIABLE = PRECIPITATION ANOMALIES
CRU TS v.3.23
Spatial resolution: 0.5°
Temporal resolution: monthly

PREDICTIVE FEATURES = CLIMATE OSCILLATION INDICES
Atlantic Meridional Mode (**AMMSST**) index
Atlantic Multidecadal Oscillation (**AMO**) index
Dipole Mode Index (**DMI**)
East Atlantic (**EA**) index
East Atlantic / West Russia (**EAWR**) index
East Pacific-North Pacific (**EPNP**) index
Northern Annular Mode (**NAM**) index
North Atlantic Oscillation (**NAO**) index
Pacific Decadal Oscillation (**PDO**) index
Polar / Eurasia (**PEA**) index
Pacific / North American Index (**PNA**) index
Southern Annular Mode (**SAM**) index
Scandinavian Pattern (**SCAND**) index
Southern Oscillation Index (**SOI**)
Tropical Northern Atlantic (**TNA**) index
Tropical Southern Atlantic (**TSA**) index
Western Pacific (**WP**) index

17 Climate Oscillation Indices

$$y_{pred} = \sum_{feature=1}^{17} \left\{ \sum_{lag=0}^{5} \left(x_{feature_{lag}} \cdot \hat{\beta}_{feature_{lag}} \right) + bias \right\}$$

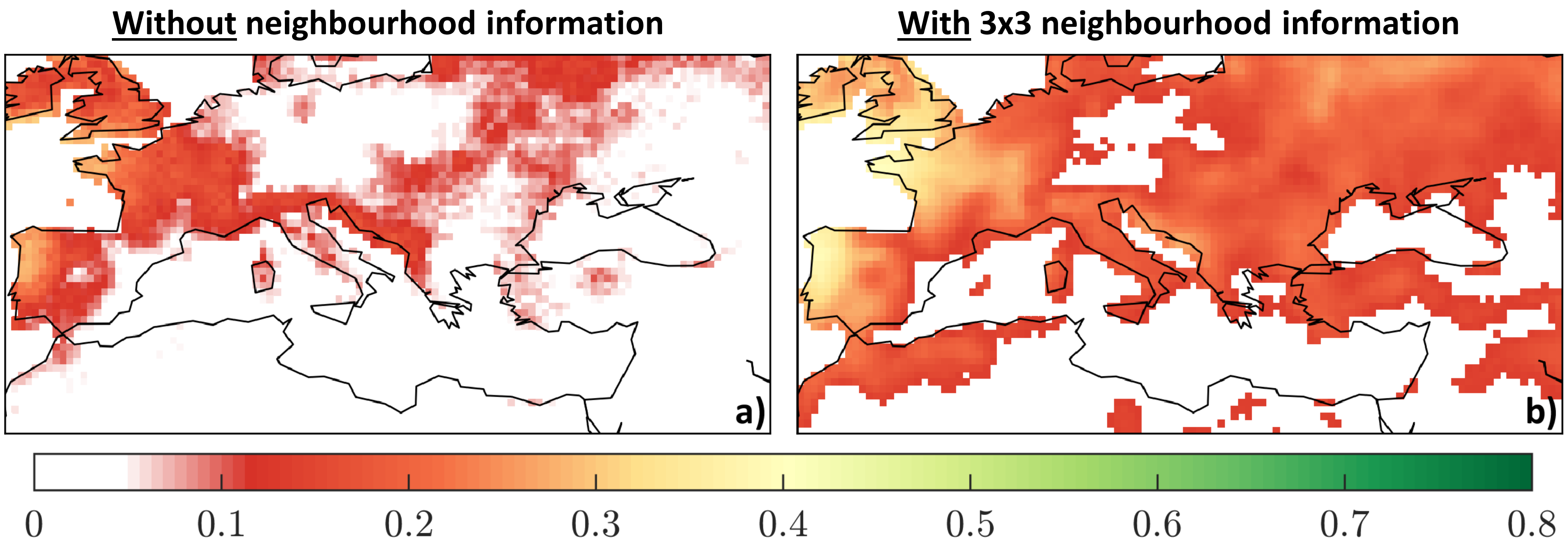
Precipitation from CRU

Regression coefficients

laura.crocetti@geo.tuwien.ac.at

Full-year model

- R^2 for full-year model



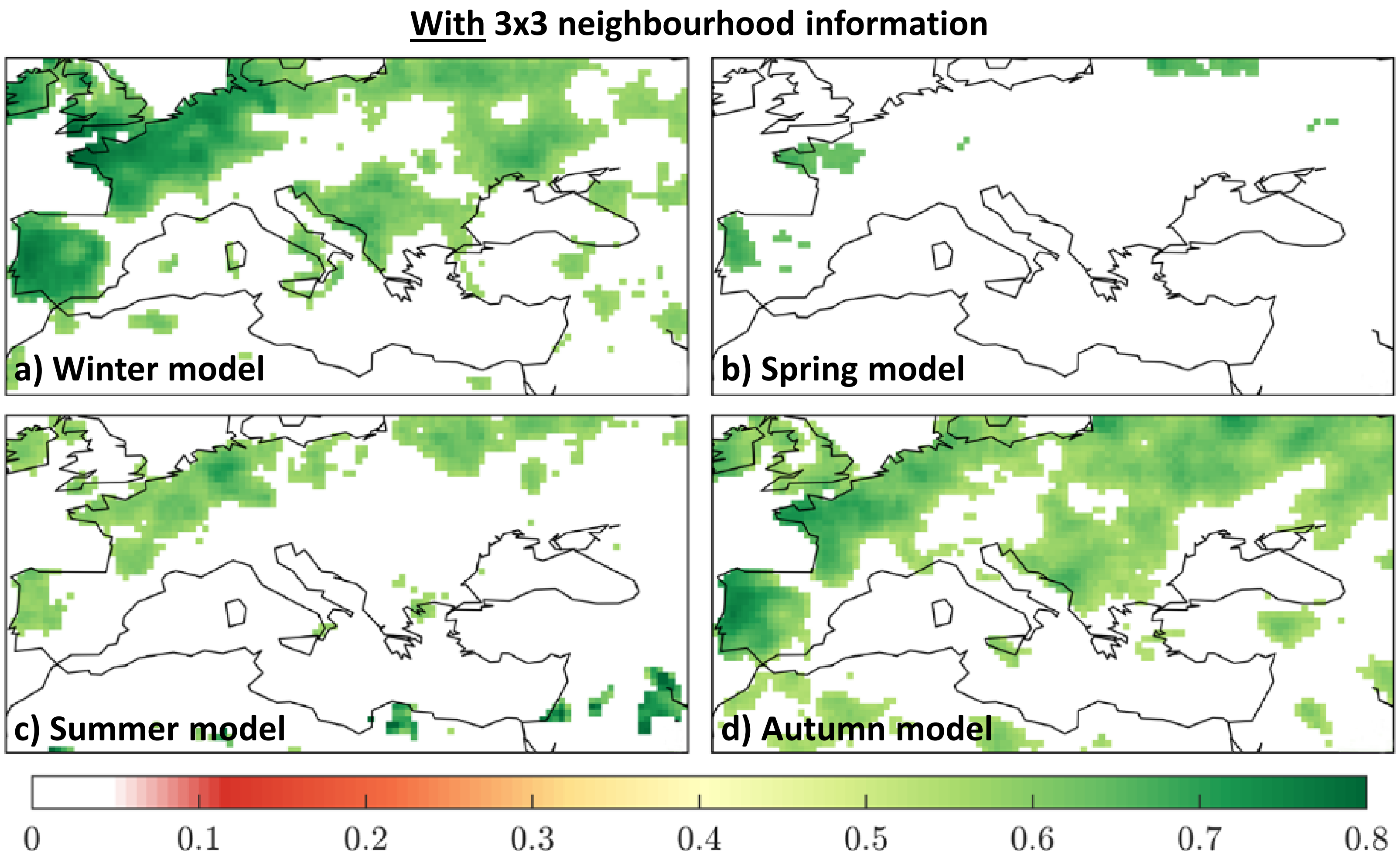
→ R^2 gets significantly higher when 3x3 neighbourhood is used

laura.crocetti@geo.tuwien.ac.at

Seasonal models

- R^2 for seasonal models

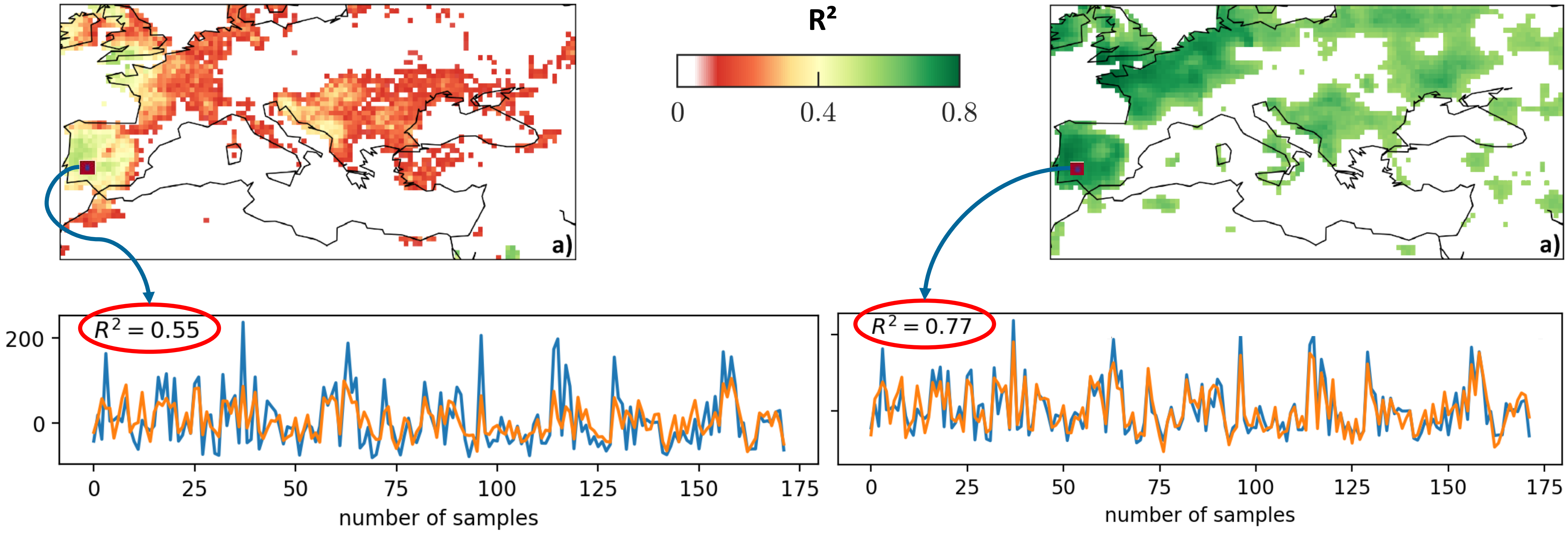
→ considerably higher R^2



Example: analysis over single location

Without neighbourhood information

With neighbourhood information



- target variable $\boldsymbol{y}$, predicted target variable $\boldsymbol{y_{pred}}$

→ model including the neighbourhood corresponds better to the target variable y

laura.crocetti@geo.tuwien.ac.at

EGU CC BY

Impact of individual COIs

TU WIEN GEO

- R^2_{COI} for <u>winter</u> model

$$y_{pred_{COI}} = \sum_{lag=0}^{5} x_{COI_{lag}} \cdot \hat{\beta}_{COI_{lag}}$$

$$R^2_{COI} = corr\left(y, y_{pred_{COI}}\right)^2$$

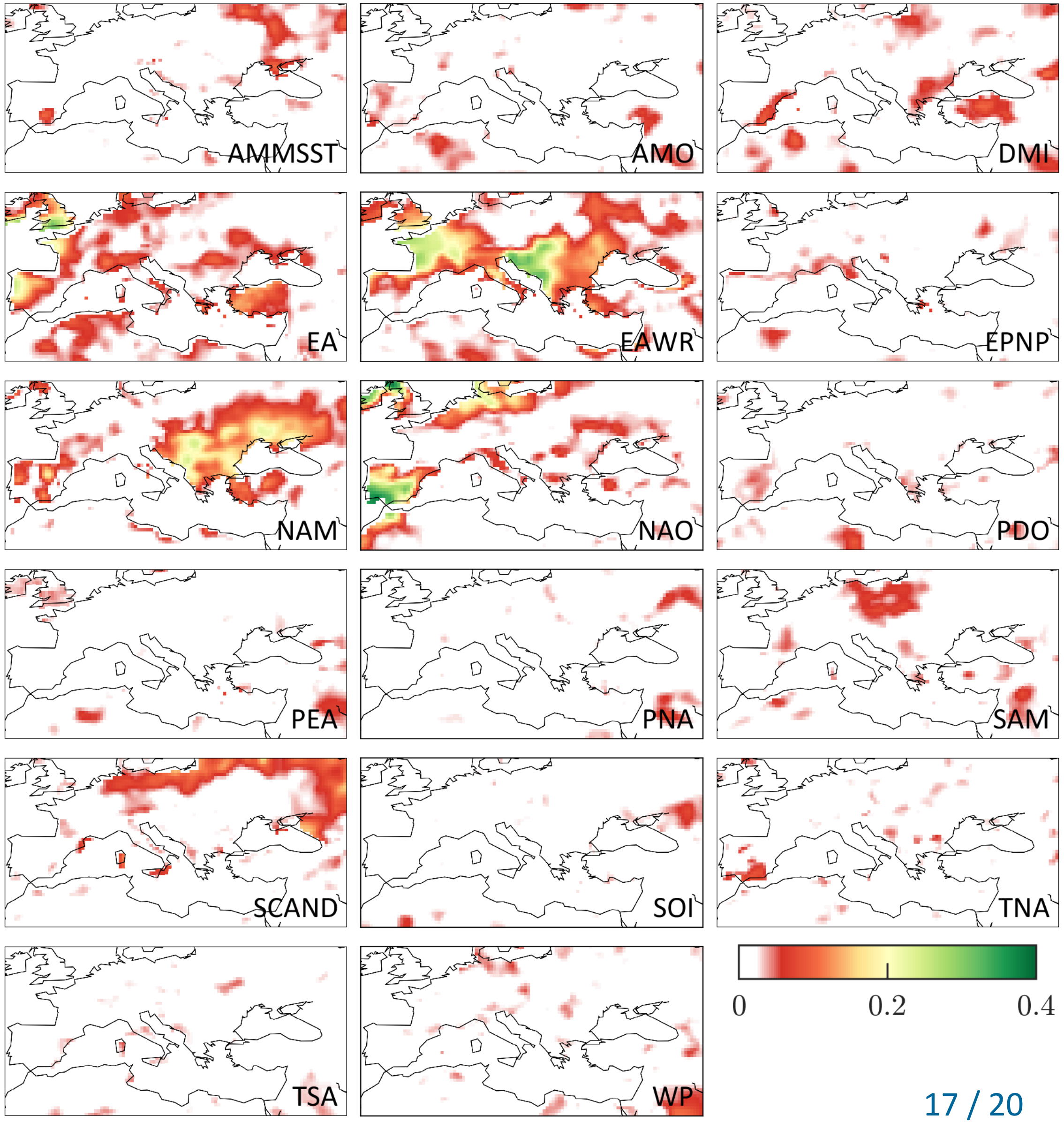
→ East Atlantic Pattern **EA**

→ East Atlantic West Russia Pattern **EAWR**

→ Northern Annular Mode **NAM**

→ North Atlantic Oscillation **NAO**

→ Scandinavian Pattern **SCAND**



EGU cc BY

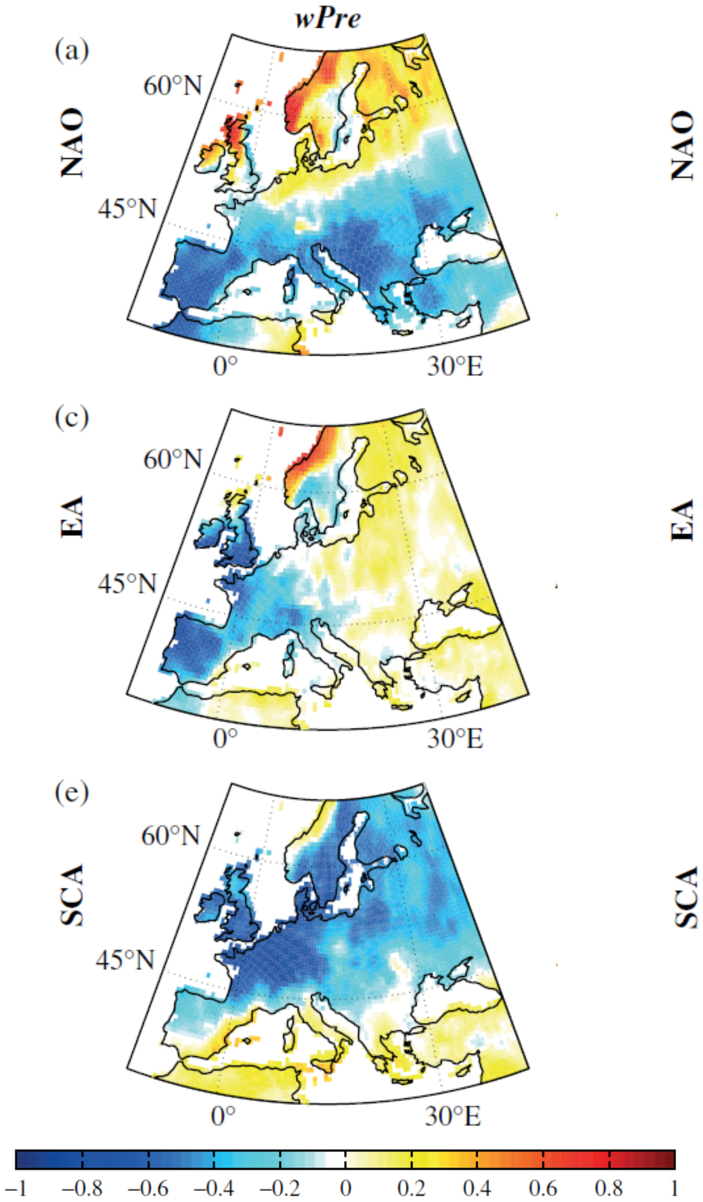
Conclusion

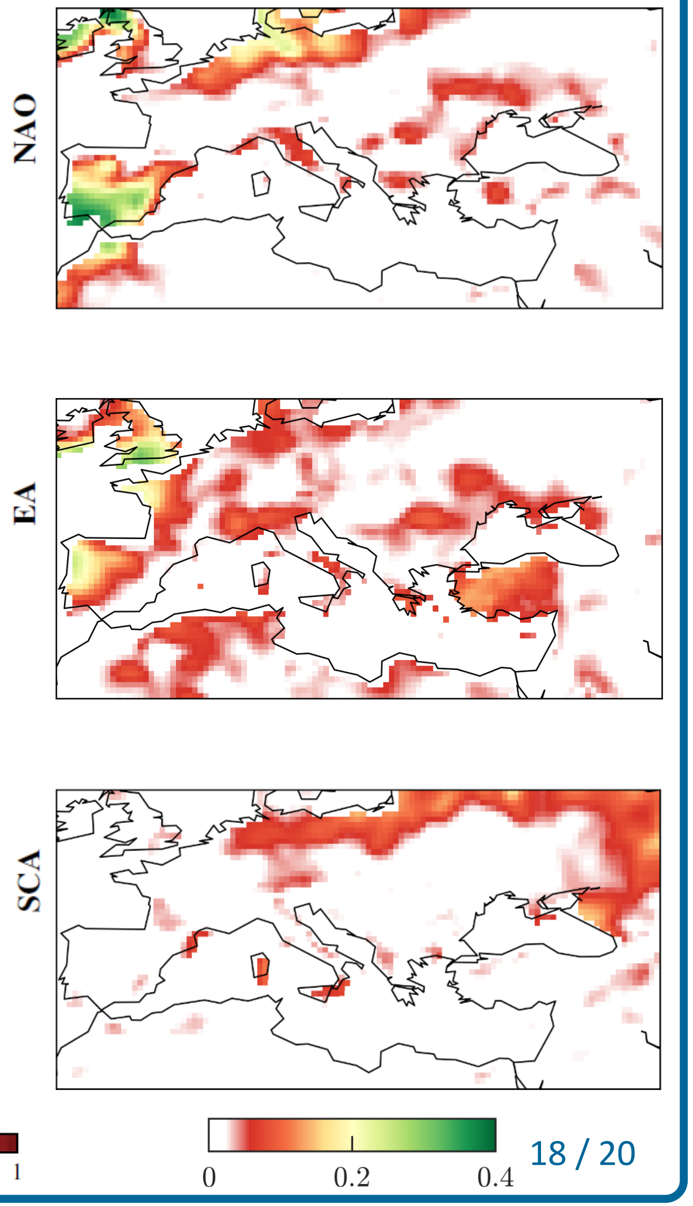
TU WIEN GEO

Comparison:

- Correlations between winter precipitation and climate modes **NAO, EA, SCAND** *[Comas-Bru and McDermott, 2013]*

- R^2_{COI} for winter model gained by LASSO regression





EGU CC BY

Conclusion

- Variability of **precipitation anomalies** is strongly affected by **climatic oscillations**

- LASSO regression
- Cross-correlations between individual climate modes are considered
- Especially winter model shows a high R^2
  - East Atlantic Pattern, East Atlantic West Russia Pattern, Northern Annular Mode, North Atlantic Oscillation, Scandinavian Pattern

- The model improves significantly when including the neighbourhood
  - Data quality

laura.crocetti@geo.tuwien.ac.at

Thank you for your attention! ☺

Further investigations:

- Limits of the method
- Pre-processing of the data
- Separate the investigated area into groups with similar hydro-climatic properties and train the model based on these areas
- Global scale
- Different response variables

- Improve understanding of the relation between **climatic oscillations** and **precipitation anomalies**

laura.crocetti@geo.tuwien.ac.at