Impacts of Climatic Oscillations on Precipitation in an Extended Mediterranean Area

LAURA CROCETTI[1] & WOUTER DORIGO[1]

Abstract: Oceanic-atmospheric oscillation patterns, described by so-called climate modes, control a large portion of the variability in the water cycle. However, the relation between ocean-atmospheric oscillations and hydrology is still poorly understood. Better knowledge about these connections is needed to accurately predict the impact of climate change on hydrology. In this study, a supervised machine learning approach based on least absolute shrinkage and selection operator (LASSO) models is used to identify and disentangle the impact of 17 major climate modes on precipitation anomalies over an extended Mediterranean area. Analysing the results of the winter model and using precipitation from the Climate Research Unit (CRU) as a response variable, shows that up to 70% of the variability can be explained by ocean-atmospheric oscillations. These results help to improve the general understanding of how individual climate modes affect different parts of the Mediterranean area.

1 Introduction

In recent decades, global warming has led to significant changes in the processes of our climate system, including the cycles of energy, water, and biogeochemicals. A thorough understanding of these cycles is needed to assess the potential implications of global warming on our future climate. A large portion of the natural variability of climate is controlled by the coupled state of ocean and atmosphere, the so-called ocean-atmospheric oscillations. These climate modes characterized by a repeating time-space pattern are important drivers of the hydrology and can even influence regions that are far apart. Therefore, they are known as teleconnections (WALLACE & GUTZLER 1981; WANG & SCHIMEL 2003). Each climate mode is expressed through its corresponding Climate Oscillation Index (COI) that describes the state of the atmospheric-oceanic oscillations in numbers. One of the most relevant teleconnections affecting the Mediterranean hydrology is the well-known North Atlantic Oscillation (NAO) (BARNSTON & LIVEZEY 1987). Its index is defined through the difference between the normalized sea-surface pressure over Lisbon, Portugal and Reykjavik, Iceland. However, not only the NAO influences the Mediterranean climate but also other large-scale phenomena, like the East Atlantic Oscillation (EA), the East Atlantic West Russia Pattern (EAWR) or the Scandinavian Pattern (SCAND).

A lot of studies investigate the relationship between individual climate modes and their impact on the Mediterranean climate (MIKHAILOVA & YUROVSKY 2016; BLADÈ et al. 2012; CHRONIS et al. 2011). It is shown that teleconnection patterns exert a significant impact on climatic drivers (KRICHAK et al. 2014; MARTENS et al. 2018). Although several climate modes show co-varying behaviour, i.e. individual climate modes are highly correlated, most of the studies do not address

[1] TU Wien, Department of Geodesy and Geoinformation, Wiedner Hauptstraße 8, A-1040 Wien
E-Mail: [laura.crocetti, wouter.dorigo]@geo.tuwien.ac.at

the coupled interactions between individual climate modes. However, these interactions are important and their consideration improves the accuracy of the results compared to correlations. Therefore, further research is necessary to improve the general understanding of the connection between ocean-atmospheric oscillations and precipitation anomalies.
The objective of this study is to reveal the dominant climate modes controlling precipitation anomalies in an extended Mediterranean area. Therefore, a supervised machine learning approach, called least absolute shrinkage and selection operator (LASSO) regression is used. It is a data-driven multivariate method that analyses the joint behaviour of multiple variables concurrently. This means the co-varying behaviour of the individual climate modes is considered and the impact of many climate modes on precipitation anomalies is determined simultaneously.

2 Data

2.1 Climate modes

Climate modes describe the state of the ocean-atmospheric circulations and follow distinct repeating spatio-temporal patterns. Each climate mode has a corresponding Climate Oscillation Index (COI) that monitors the strength and the temporal pattern of the individual climate mode. In this study, 17 major climate modes, expressed through their associated COI, listed in Table 1 are used as predictive features in a LASSO regression framework. The COIs have a monthly temporal resolution. The raw time series of the COIs get pre-processed. First, the multi-year climatology is eliminated to remove the seasonal variations of the signal, then the COIs are standardized by months and the whole time series gets normalized by dividing it by the l2-norm.

Tab. 1: Climate modes used in this study with their corresponding COI

Climate mode	Climate Oscillation Index (COI)
Atlantic Meridional Mode (AMM)	AMM sea-surface temperature (AMMSST) index
Atlantic Multidecadal Oscillation (AMO)	AMO index
Indian Ocean Dipole (IOD)	Dipole Mode Index (DMI)
East Atlantic Pattern (EA)	EA index
East Atlantic West Russia Pattern (EAWR)	EAWR index
East Pacific North Pacific Pattern (EPNP)	EPNP index
Northern Annular Mode (NAM)	NAM index
North Atlantic Oscillation (NAO)	NAO index
Pacific Decadal Oscillation (PDO)	PDO index
Polar Eurasia Pattern (PEA)	PEA index
Pacific-North American Pattern (PNA)	PNA index
Southern Annular Mode (SAM)	SAM index
Scandinavian Pattern	SCAND index
El Niño-Southern Oscillation (ENSO)	Southern Oscillation Index (SOI)
Tropical Northern Atlantic Dipole (TNA)	TNA index
Tropical Southern Atlantic Dipole (TSA)	TSA index
West Pacific Pattern (WP)	WP index

2.2 Precipitation data

The target variable in the LASSO framework are precipitation anomalies. In this study, precipitation provided by the Climatic Research Unit (CRU) TS v.3.23 (HARRIS et al. 2013) is used. It is a gridded high-resolution climate dataset from monthly observations at meteorological

stations across the world's land areas. It is provided on a regular latitude-longitude grid with a common spatial resolution of 0.5°. The LASSO regression is trained by the anomalies of the precipitation data instead of the raw time series to remove seasonal signals. First, the linear trend is removed, then the anomalies are calculated based on the multi-year climatology.

3 Method – LASSO regression

For this study, a supervised machine learning algorithm, called least absolute shrinkage and selection operator (LASSO) regression (TIBSHIRANI 1996) is used to determine and disentangle the impact of climate modes on precipitation anomalies. In this framework, 17 COIs (see Table 1) are used as predictive features, and anomalies of precipitation of CRU are used as the target variable. One benefit of the LASSO regression is, that it takes the cross-correlations of the features into account. This is important since some climate modes are highly correlated, as shown in Figure 1. Furthermore, the method uses feature selection and regularization to quantify the impact of each feature on the target variable. The regularization helps to prevent overfitting by setting the coefficients of weak features to zero. Thus, only a subset of features remain, making the final predictive model simpler and improving the prediction accuracy of the model. Compared to other machine-learning algorithms like random forest, results from the LASSO regression are easier to understand and evaluate.

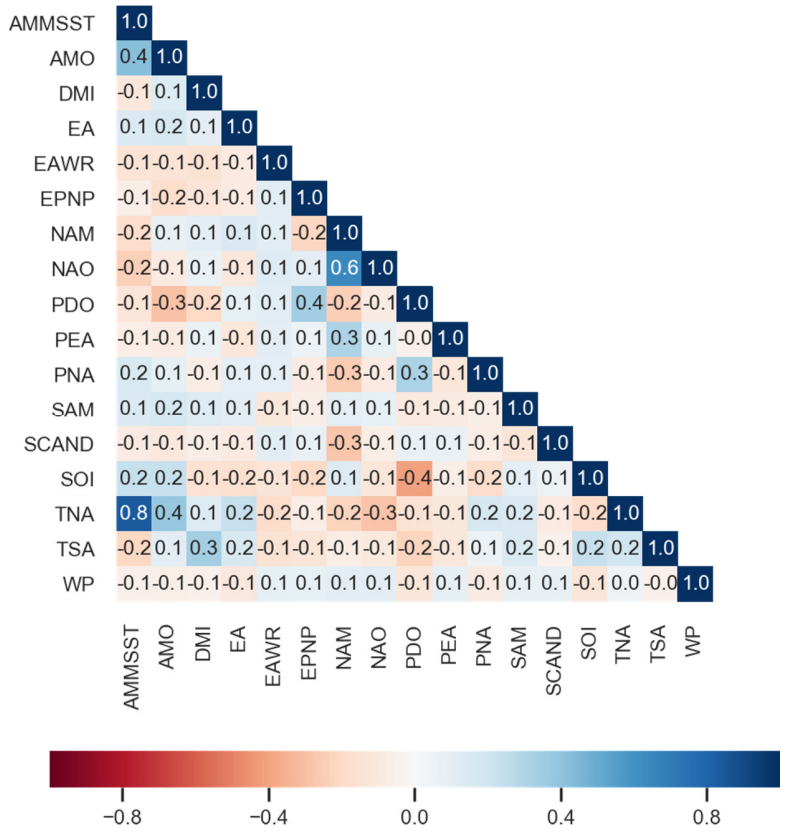

Fig. 1: Cross-correlation of COI anomalies

The LASSO regression determines the estimated regression coefficients $\hat{\beta}$ by minimizing a penalized residual sum of squares (TIBSHIRANI 1996).

$$\hat{\beta} = argmin\left\{\sum_{i=0}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \alpha\sum_{j=1}^{p}|\beta_j|\right\} \quad (1)$$

In equation (1), $\hat{\beta}$ is the p-dimensional vector with the estimated regression coefficients, n the number of training samples in the dataset, y_i the value of the target variable in sample i, p the number of features, x_{ij} the value of features j in sample i, and α a hyper-parameter controlling the amount of shrinkage (MARTENS et al. 2018).

The estimated regression coefficients $\hat{\beta}$ are used to predict the target variable y_{pred},

$$y_{pred} = \sum_{feature=1}^{17}\left\{\sum_{lag=0}^{5}\left(x_{feature_{lag}}\cdot\hat{\beta}_{feature_{lag}}\right) + bias\right\} \quad (2)$$

where x is the feature vector for a given time lag and $\hat{\beta}$ is the vector with the corresponding estimated regression coefficients.
To validate the predictive performance of the model, the coefficient of determination R^2 is calculated. It is defined as the squared correlation between the target variable y and the predicted target variable y_{pred}.

$$R^2 = corr\left(y, y_{pred}\right)^2 \quad (3)$$

4 Experimental set up

In this study, analysis over an extended Mediterranean area spanning from 28.5°N to 56.5°N and from 10°W to 46°E are carried out. A LASSO regression model is fitted individually for every grid point using 17 major COIs as features and precipitation anomalies as the target variable.
Figure 2 shows the concept of the LASSO regression. The left block is the target variable vector y and the right block represents the feature matrix X. Six time lags ranging between zero and five months are introduced for every feature to take a potential delayed response of precipitation anomalies to ocean-atmospheric oscillations into account. This results in a total of 102 features (17 COIs times 6 time lags). The overlapping period for the COIs and the anomalies of the CRU precipitation dataset is from 1957 to 2014. This period is used for the analysis.

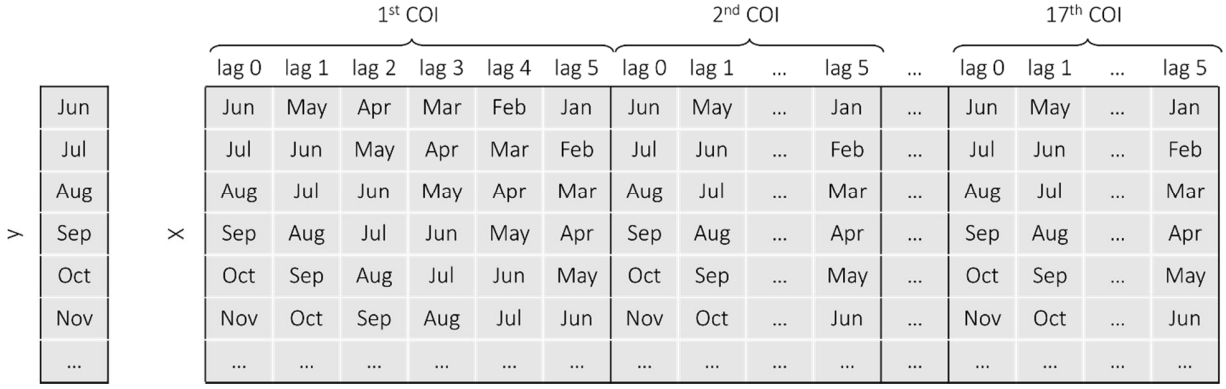

Fig. 2: Basic framework of the LASSO regression

Besides developing models based on the entire available time series, additional models for each season are calculated. The winter model includes values from December, January and February, the spring model includes March, April and May, the summer model includes June, July and August and the autumn model includes September, October, and November.
To optimize and validate the predictive LASSO regression models, two five-fold cross-validations are applied. One cross-validation is used to determine the regularization parameter α and the second determines the coefficient of determination R^2. To ensure that the results are robust and trustful, a significance test using the Benjamini-Hochberg procedure (BENJAMINI & HOCHBERG 1995) at a significance level of 95% is applied and areas which do not fulfil this test are masked.
Furthermore, to increase the predictive performance of the model and make the results more robust, the 3×3 neighbourhood is included. The framework is extended by the data of the eight neighbouring grid points. It is expected that by including the neighbourhood information and, therefore, extending the time series by a factor of nine, more information can be extracted from the time series. Thus, if one individual grid point does not provide enough information to create a good model, at least the neighbouring grid points can help to improve the prediction.

5 Results and Discussion

Figure 3 depicts the explained variance R^2 of the precipitation anomalies obtained by the LASSO regression for the winter model from 1957 to 2014. In the left plot, the LASSO regression is trained for each pixel individually, while in the right plot the 3×3 neighbourhood is taken into account. By comparing the two maps it can be seen that the explained variance R^2 gets significantly higher when the 3×3 neighbourhood is included and variabilities up to 70% can be estimated. The spatial pattern detected look very similar, however, the explained variances R^2 are significantly higher for the model including the 3×3 neighbourhood. The reason for the higher coefficient of determination R^2 is, that the impact of the noise in the data is reduced due to the nine times longer time series. Thus, the time series gets more robust.

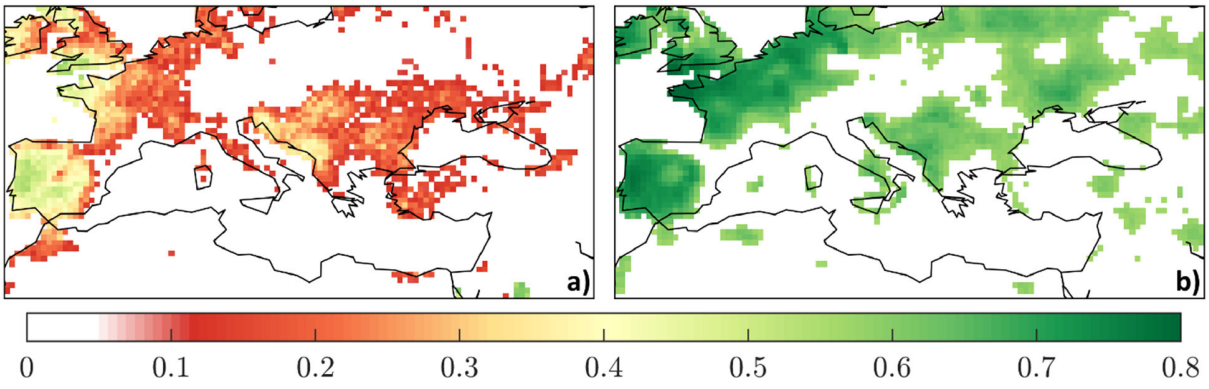

Fig. 3: Explained variance R^2 of the LASSO regression for the winter model. Areas that are classified as non-significant are marked white. a) the model is trained for each pixel individually, b) additional 3x3 neighbourhood is taken into account

Precipitation anomalies in the winter model show a high sensitivity to the climate modes. In particular, the Iberian Peninsula, Ireland, the United Kingdom, France, Belgium, the Netherlands, as well as some parts of Italy, and the eastern coast of the Adriatic Sea are influenced by the modes of climate variability. The white areas in the maps in Figure 3 have been declared as non-significant by the permutation test using the Benjamini-Hochberg procedure at a significance level of 95%.

The impact of each COI is investigated by calculating the explained variance for each COI R^2_{COI}. This value is calculated by taking the six regression coefficients $\hat{\beta}$ of one individual COI from the LASSO regression and based on these coefficients the predicted target variable $y_{pred}{}_{COI}$ and finally the explained variance R^2_{COI} is calculated.

Figure 4 visualizes the explained variance for the winter model for the five COIs R^2_{COI} detected having a significant impact on precipitation anomalies in this area, i.e. the East Atlantic Pattern (EA), the East Atlantic West Russia Pattern (EAWR), the Northern Annular Mode (NAM), the North Atlantic Oscillation (NAO), and the Scandinavian Pattern (SCAND).

The explained variance of each COI R^2_{COI} is considerably lower than the explained variance R^2 of the complete model. This can be explained by the now missing interaction of the individual COIs.

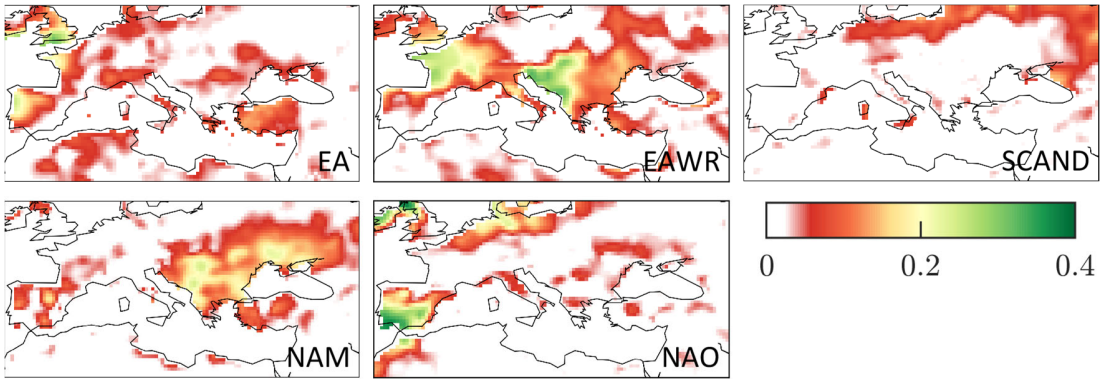

Fig. 4: Impact of the five climate modes identified having the highest impact on precipitation anomalies R^2_{COI} for the winter model with additional neighbourhood information. The respective abbreviations of each climate mode are written in the bottom right corner.

To better understand how the LASSO model works and what the coefficient of determination R^2 shows, the results of one randomly chosen grid point are discussed in more detail. The grid point is located at a latitude of 38.75°N and 7.25°W. It lies in the south-western region of the Iberian Peninsula.

Figure 5 shows the results of a step-wise addition of COIs to the LASSO regression model for that grid point. The LASSO model is trained by using all 17 COIs. In case of the no neighbourhood model, for this specific grid point only three features, namely EA, NAO, and NAM are selected by the automatic feature selection of the LASSO regression to build the final model.

The first plot in Figure 5 shows the result, if only the regression coefficients $\hat{\beta}$ from the full LASSO regression model of one feature, namely EA, would be used to predict the precipitation anomalies. The corresponding coefficient of determination would result in R^2=0.12. By adding the regression coefficients $\hat{\beta}$ of the second feature, i.e. NAO, the coefficient of determination would increase

considerably to R^2=0.54. Adding the third feature, i.e. NAM, shows the final LASSO result with a R^2 of 0.55. The pattern of the predictions y_{pred} of the final LASSO model follow the temporal patterns of the target variable y well but the amplitude is lower.

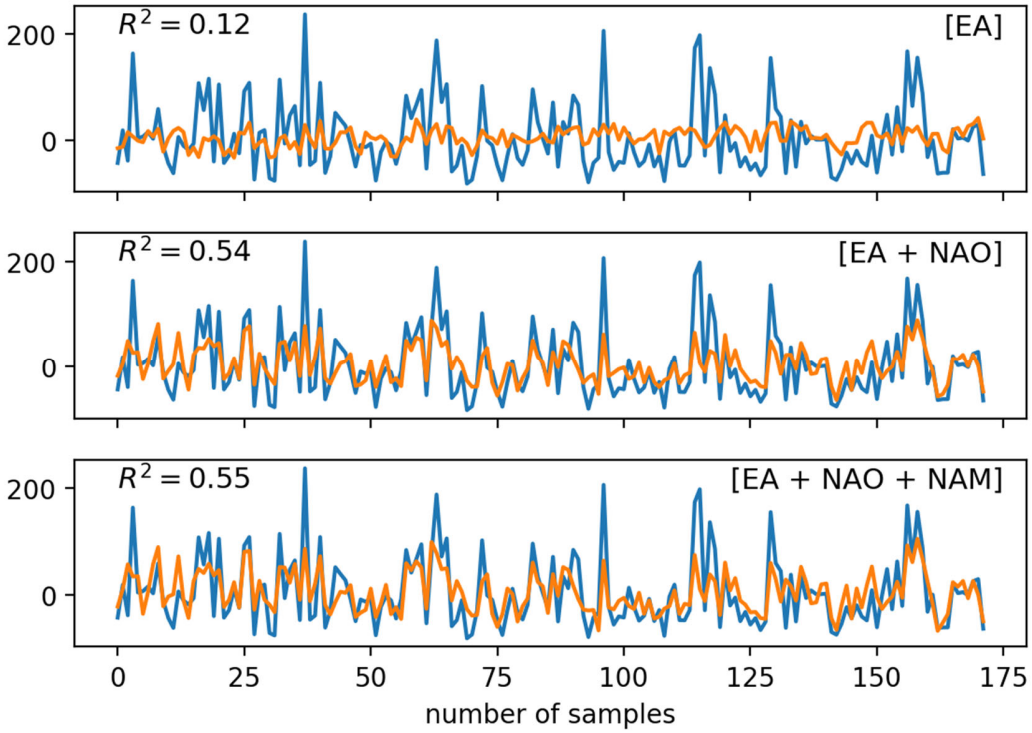

Fig. 5: Step-wise addition of COIs building the LASSO winter model for precipitation anomalies at 38.75°N and 7.25°W without including the neighbourhood. Blue: target variable y, orange: predicted target variable y_{pred}. Only three COIs are selected by the LASSO regression model.

Figure 6 shows the same step-wise addition of individual features as Figure 5 but includes the 3x3 neighbourhood. By adding the neighbourhood information the LASSO regression uses a lot more features to build its prediction model. Therefore, three to four features are added simultaneously in Figure 6 to reduce the number of necessary plots. The final LASSO regression model is the last plot in Figure 6. It is clearly visible, that the final result including the neighbourhood corresponds better to the target variable y as it is the case in Figure 5. This time the temporal patterns and amplitudes are represented better. This can also be verified by the higher explained variance R^2 of 0.77. However, the contribution per COI is lower.

Although this is only the result of one grid point, the same can be seen at nearly every location. As shown in Figure 3 the explained variance R^2 is almost always higher when adding the information of the neighbourhood.

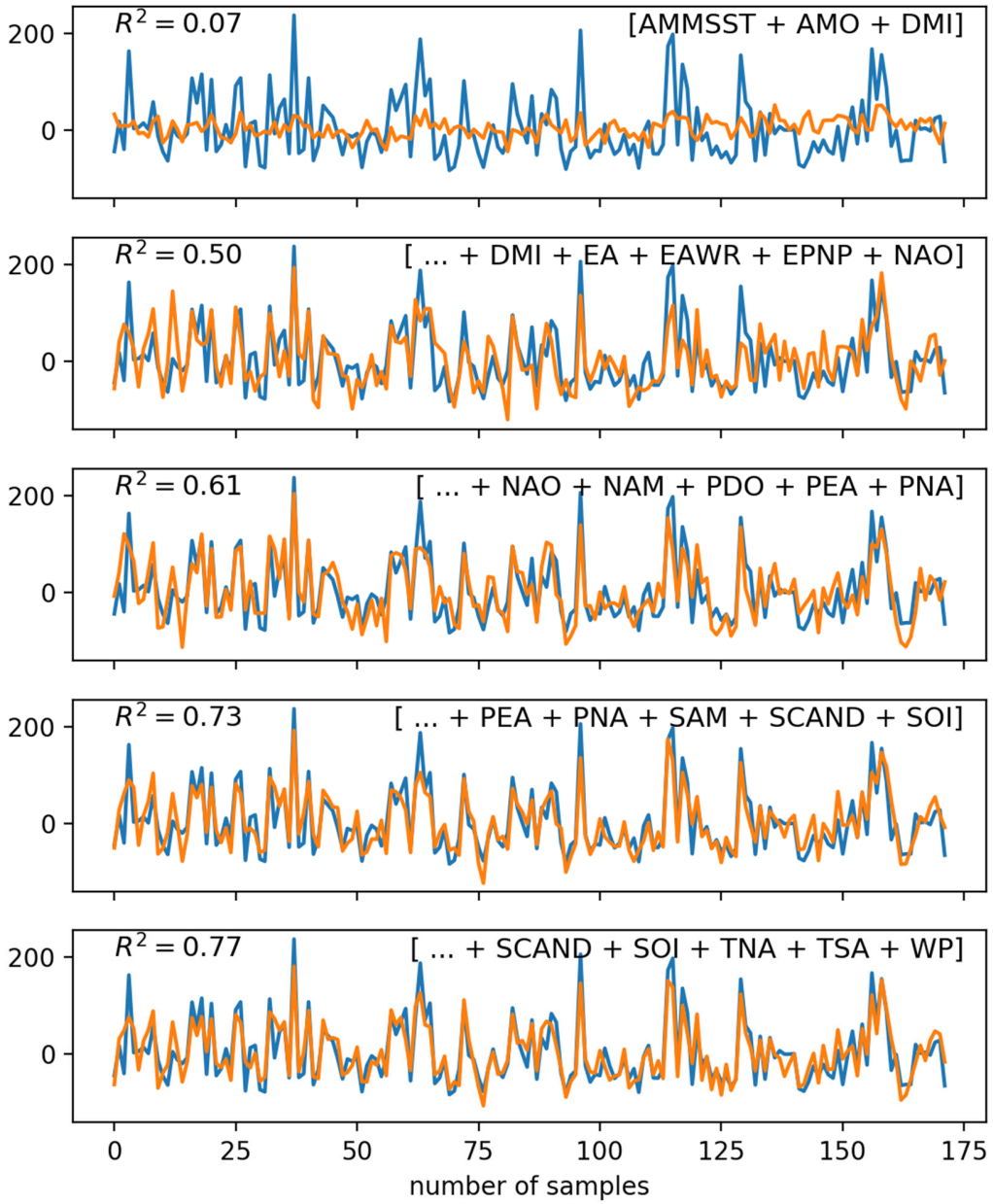

Fig. 6: Step-wise addition of COIs building the LASSO winter model for precipitation anomalies at 38.75°N and 7.25°W including the neighbourhood. Blue: target variable y, orange: predicted target variable y_{pred}

6 Conclusion

In this study, the impact of 17 climate modes on precipitation anomalies over an extended Mediterranean area is analysed. The results show that the variability in precipitation anomalies are strongly affected by ocean-atmospheric oscillations.

A supervised learning approach, called LASSO regression is used to determine the impact of individual climate modes. The method uses feature selection and regularization and can help

revealing the dominant modes of climate variability by disentangling their complex impacts and interactions. The analysis is performed in two ways, once by analysing the time series of each grid point individually and, furthermore, by adding the information of the 3x3 neighbourhood. If the LASSO framework is extended by the 3x3 neighbourhood, the model improves significantly, resulting in a high explained variance R^2. Additionally, seasonal models are calculated which lead to a significantly higher explained variance R^2 compared to the full-year model that uses all months of a year. Especially, the winter model shows a high coefficient of determination R^2. In some regions up to 70% of the precipitation anomalies can be explained by the combination of specific climate modes. The LASSO regression detected that EA, EAWR, NAM, NAO, and SCAND have the biggest impact, meaning that these climate modes are mostly responsible for inter-annual precipitation anomalies during winter.
Based on these results, further research can be carried out to improve the understanding of the relation between climatic oscillations and hydrology.

7 References

BARNSTON, A. G. & LIVEZEY, R. E., 1987: Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. Monthly Weather Review, **115**(6), 1083-1126.

BENJAMINI, Y. & HOCHBERG, Y., 1995: Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B (Methodological), **57**(1), 289-300.

BLADÈ, I., LIEBMANN, B., FORTUNY, D. & OLDENBORGH, G. J., 2012: Observed and simulated impacts of the summer NAO in Europe: Implications for projected drying in the Mediterranean region. Climate Dynamics, **39**(3), 709-727.

CHRONIS, T., RAITSOS, D. E., KASSIS, D. & SARANTOPOULOS, A., 2011: The summer North Atlantic Oscillation influences on the eastern Mediterranean. Journal of Climate, **24**(21), 5584-5596.

HARRIS, I., JONES, P., OSBORN, T. & LISTER, D., 2013: Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 dataset. International Journal of Climatology, **34**(3), 623-642.

KRICHAK, S. O., BREITGAND, J. S., GUALDI, S. & FELDSTEIN, S. B., 2014: Teleconnection-extreme precipitation relationships over the Mediterranean region. Theoretical and Applied Climatology, **117**(3), 679-692.

MARTENS, B., WAEGEMAN, W., DORIGO, W.A., VERHOEST, N. E. C. & MIRALLES, D. G., 2018: Terrestrial evaporation response to modes of climate variability. NPJ Climate and Atmospheric Science, **1**(1), 1-7.

MIKHAILOVA, N. & YUROVSKY, A., 2016: The East Atlantic Oscillation: Mechanism and impact on the European climate in winter. Physical Oceanography, (4), 25-33.

TIBSHIRANI, R., 1996: Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B, (Methodological), **58**(1), 267-288.

WALLACE, J. M. & GUTZLER, D. S., 1981: Teleconnections in the geopotential height field during the northern hemisphere winter. Monthly Weather Review, **109**(4), 784-812.

WANG, G. & SCHIMEL, D., 2003: Climate Change, Climate Modes, and Climate Impacts. Annual Review of Environment and Resources, **28**(1), 1-28.