D3.7
Report on sources of predictability for flow features in subseasonal forecasts

Lead Beneficiary: ECMWF
Work Package: WP3
Dissemination level: PUBLIC
Type: Report
Due date: August 2022
Submission date: 09.09.2022
Authors: Nikolaos Mastrantonas, Linus Magnusson, Jörg Matschullat

1. Introduction

Extreme precipitation events (EPEs) cause devastating consequences to society, economy, and the environment (Jonkman, 2005). The Mediterranean frequently experiences such events, which are considered to be the most impactful meteorological hazard in the region (Llasat et al., 2010; Tramblay and Somot, 2018). Thus, better forecasting of EPEs at longer lead times can provide the time needed to implement mitigating actions that will reduce their negative impacts.

Previous work has already demonstrated that using the large-scale weather variability over the Mediterranean can improve the EPEs' forecasts for the medium- and extended-range forecasts, in particular for the end of week 1 and the week 2 of predictions (Mastrantonas et al., 2022).

The large-scale variability in the designed forecasting tools was introduced in the form of nine weather patterns, with distinct atmospheric characteristics over the domain. These patterns, known as "Mediterranean Patterns", are depicted in Figure 1. They were derived based on empirical orthogonal function and subsequent K-means clustering on sea-level pressure and geopotential height at 500 hPa (for more information see Mastrantonas et al., 2021). These patterns indicate negative anomalies over the western Mediterranean (Atlantic/Biscay/Iberian Low), central Mediterranean (Sicilian Low) and the eastern Mediterranean (Balkan/Black Sea Low), positive anomalies over the whole domain (Mediterranean High), or, finally, non-substantially anomalous conditions over the entire domain (Minor High, Minor Low).

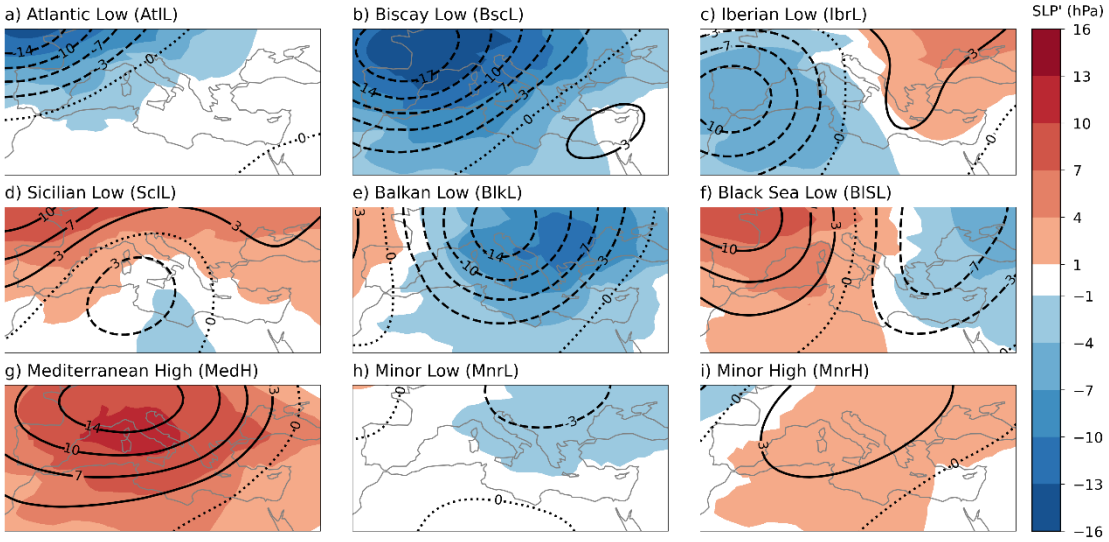

Figure 1. The 9 Mediterranean patterns (Mastrantonas et al. 2021). The figure presents the composites of clusters derived with K-means clustering on the principal components' projections of SLP and Z500 anomalies. Colour shading refers to SLP anomalies (hPa), and contours to Z500 anomalies (dam). In parentheses their abbreviations are given that are used in the other figures of this document.

CAFE project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813844.

Given the usefulness of the Mediterranean patterns for EPEs predictions at longer lead times, there is an interest in identifying other modes of atmospheric variability that are associated with these patterns and can influence their frequency or predictability. Especially for such modes with high forecasting skill, possible (tele)connections between them and the Mediterranean patterns can further support the predictability of the latter, and thus the predictability of EPEs.

This document provides insights into the above question and quantifies the connections between the Mediterranean patterns and other modes of atmospheric variability.

2. Data and Methods

For this analysis, we used daily data from ERA5, the latest reanalysis dataset of ECMWF. The selected atmospheric variables are sea-surface temperature (SST), the upper 7-centimetres of soil moisture layer (SML), the sea-ice area fraction (SIAF), precipitation (TP), geopotential height at 500 hPa (Z500), and sea-level pressure (SLP). These data were pre-processed to generate daily timeseries of atmospheric variability, as will be described later.

We used precomputed modes of daily variability that are freely available online. These modes are the North Atlantic Oscillation (NAO) and Atlantic Oscillation (AO), provided by NOAA (https://ftp.cpc.ncep.noaa.gov/cwlinks/), the Mediterranean Oscillation (MO; normalized pressure difference between Algiers and Cairo), provided by the University of East Anglia (https://crudata.uea.ac.uk/cru/data/moi/), and the real-time multivariate Madden-Julian Oscillation (MJO-RMM) index, provided by the ECMWF (https://acquisition.ecmwf.int/ecpds/data/list/RMMS/ea/).

Coming back to the data from ERA5, for SST and SML, we used the area over the Mediterranean (26/50N, -11/41E) and implemented empirical orthogonal function analysis with varimax rotation to obtain the dominant modes of variability over the domain. For this, we used the xMCA package (https://github.com/nicrie/xmca). We then used the projections to the initial six principal components of each variable (Figure 2) for further analysis. As can be seen in the relevant figure, these modes relate to centres of variability in various parts of the domain, from the very west (e.g., SST-3, SST-4, SML-1), up to the very east (e.g., SST-6, SML-5, SML-6).

Additionally, we derived the Western Mediterranean Oscillation (WeMO), based on the normalized pressure difference between Cadiz and Padua, as in Martin-Vide and Lopez-Bustins (2006).

CAFE project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813844.

Finally, we derived two normalized modes of variability for distant domains: the precipitation over Sahel (10/20N, -10/20E), and the SIAF over Europe (ArcticEurope; 70/90N, -30/60E).

All of the above indices (20 timeseries in total) were selected because previous works have identified links between (some of) them and the atmospheric variability over extended Europe or parts of the Mediterranean in particular (Kretschmer et al., 2016; Lee et al., 2019; Raicich et al., 2003). Thus, with this analysis we assess and quantify their connections specifically for the Mediterranean patterns.

We used the SLP and Z500 fields for getting projections of all daily fields to the nine Mediterranean patterns. As there are two variables that are projected, we initially normalized the derived timeseries and then averaged them for extracting the final nine indices.

To remove possible implications due to existing trends (e.g., climate change), and intraseasonal variability, we detrended the data and removed their seasonal average for each year, as suggested in a relevant work of Di Capua et al (2020). Finally, to remove possible noise in the data, and since the interest is on medium- and extended-range links, we used 5-day aggregations (mean) of all timeseries.

After the above-described data pre-processing, we performed the core part of the analysis, which is the identification of links between the Mediterranean patterns and other modes of variability. For this, we used (multi)linear correlation/regression; either in their simple form, as well as when considering causality between the timeseries, with the use of the Tigramite package (https://jakobrunge.github.io/tigramite/; Runge, 2020; Runge et al., 2019;). For the multilinear regression, we initially implemented a feature selection, so that we extracted the 10 indicators of highest connection to each of the Mediterranean patterns (based on correlation), thereafter, we implemented the regression for all possible combinations of the selected indicators.

The analysis was conducted for the years 1980–2020 that were common for all used indices. We additionally focused on winter-half (Sep–Feb), the time when most EPEs occur in the Mediterranean (Mastrantonas et al., 2021).

CAFE project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813844.

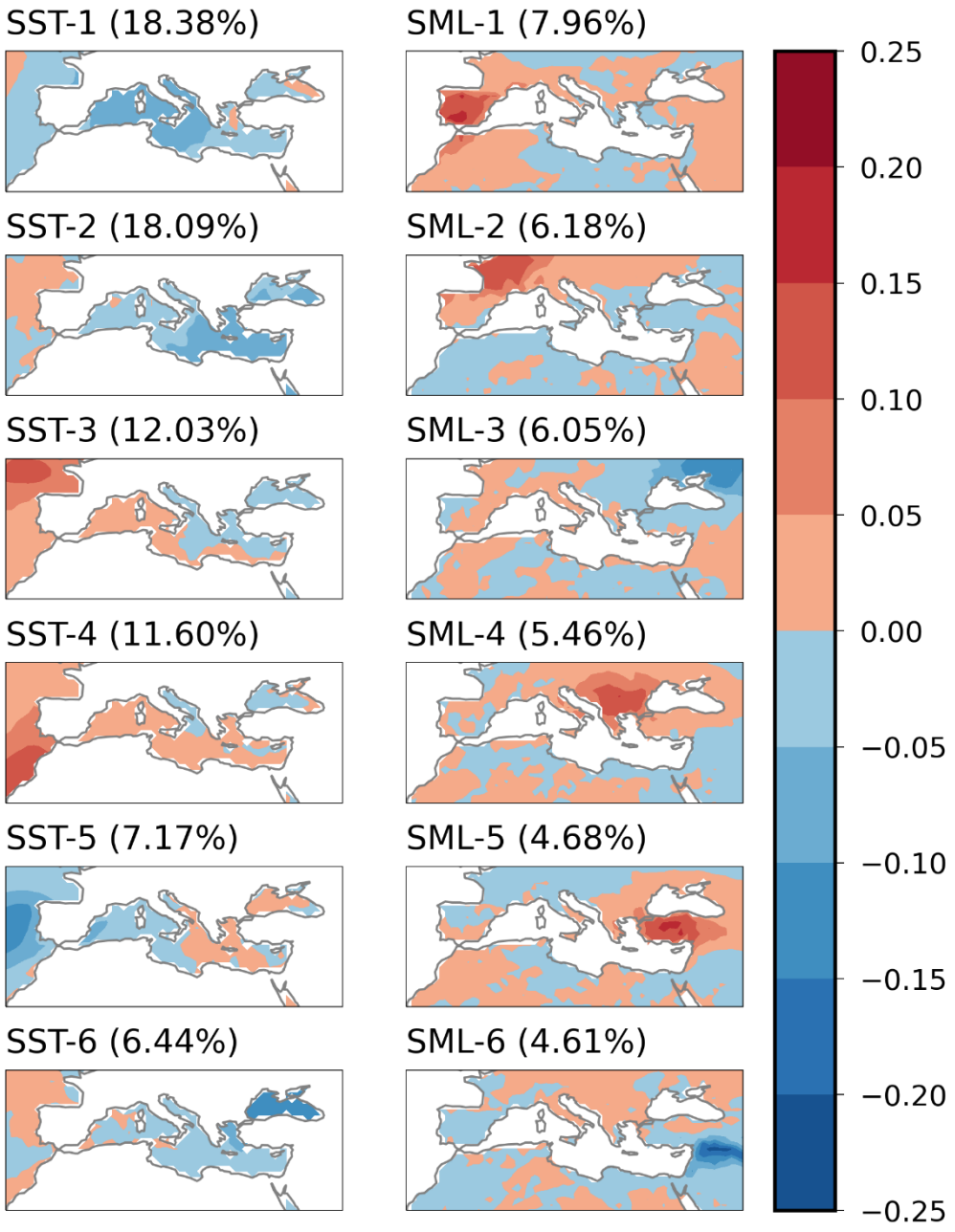

Figure 2. The initial six components of SST and SML over the Mediterranean domain based on empirical orthogonal function analysis with varimax rotation.

3. Results

Before presenting the core results of the analysis, it is worth commenting on the connections among the Mediterranean patterns, based on the cross-correlation of the relevant indices, as depicted in Figure 3. From this figure it can be noticed that there is a high degree of connectivity for most of the patterns, with either strong positive links

CAFE project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813844.

(e.g., Minor High – Mediterranean High; 0.94), or strong negative links (e.g., Black Sea Low – Iberian Low; -0.91). This is an outcome of the spatial characteristics of the patterns: Many of them have essentially a latitudinal/longitudinal shift of their low-/high-pressure centre compared to other patterns (e.g., Biscay Low has an easterly shift compared to Atlantic Low), or are a "mirror image" (e.g., Minor Low and Minor High). These strong links will be also demonstrated in the connections between the Mediterranean patterns and the remaining atmospheric indices used in this work, based on the results in the following figures.

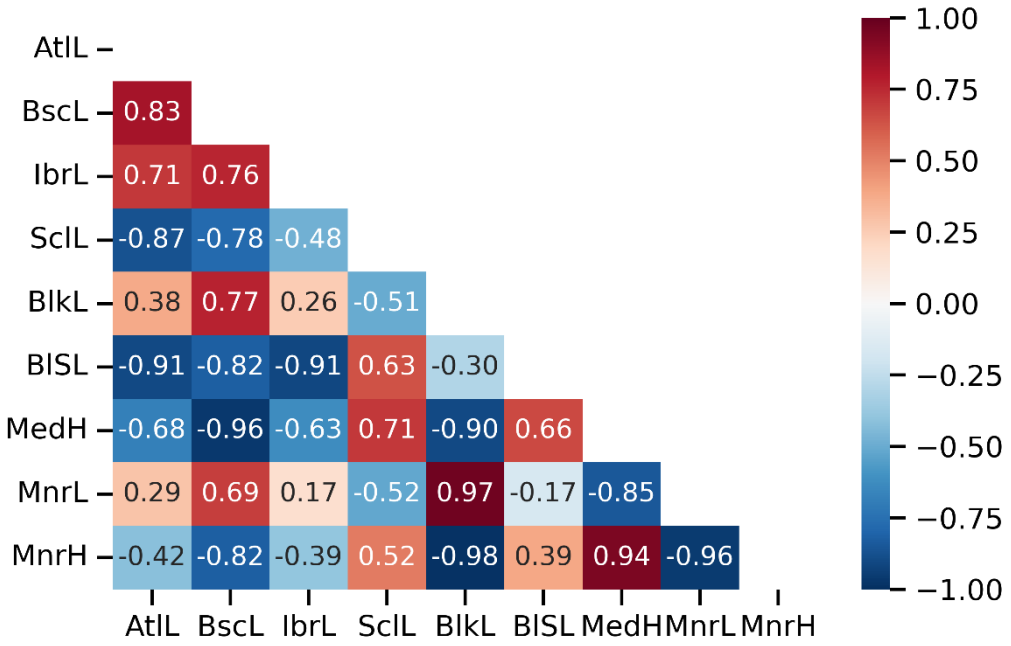

Figure 3. Cross-correlation of the 5-day aggregated indices of the Mediterranean patterns.

Figure 4 presents the lagged correlation between the nine Mediterranean indices and the remaining 20 atmospheric indices. Negative lags essentially inform about connections between the Mediterranean indices and the past state of the selected indicator, while positive lags inform about the reverse; how the past state of the Mediterranean indices affects the studied indicator. For most indicators, the peak occurs at 0 lag, meaning that most information is exchanged at a synchronous state. As previously explained, the results show high resemblance/opposite behaviour for different Mediterranean patterns, especially for the large-scale atmospheric modes, as the NAO, and ArcticEurope (SIAF). Not surprisingly, the MO has the highest correlation, as it is derived based on SLP over the Mediterranean. This correlation for lag 0 though, drops remarkably for the rest lead times. AO is stronger connected compared to NAO, despite that NAO is closer

CAFE project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813844.

related to the atmospheric variability over the (western) Mediterranean. One reason can be that AO is more "large-scale" compared to NAO. For the SST and SML indicators, the stronger links occur mainly from a previous state of the patterns to a future state of the indicators, thus it is mainly the patterns that affect these indicators and not the opposite. Precipitation over Sahel and sea ice area fraction over Europe (ArcticEurope) show relatively weak connections with the patterns. Finally, the results suggest that the past state of the MJO has an influence on the Mediterranean indices, yet of low magnitude.

Figure 4. Lagged correlation between the Mediterranean indices and the rest atmospheric indicators.

CAFE project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813844.

The above results refer to bivariate cross-correlation analysis. Yet, the atmosphere is a chaotic and highly complex system. Thus, a single indicator is not expected to provide a large amount of information. Multilinear regression analysis is the next step of this analysis. Figure 5 presents the correlation of the multilinear regression for the nine Mediterranean indices and the most connected combination of the indicators, when selecting 1 up to 10 indicators for the linear model. For the linear models of each Mediterranean index, we used the 20 atmospheric indicators and the studied Mediterranean index due to their high autocorrelation, thus obtaining a pool of 21 predictors. Here, we only focus on the negative lags, as information about the future state of the Mediterranean patterns is the main interest of this work. Initially, the results do not vary notably for the different patterns in terms of the magnitude of correlation for the different lags and number of indicators (predictors) used. The correlation improves for an increasing number of indicators, but the results start to saturate after using 4–5 indicators in the model. Moreover, there is a high drop of the connections from lag -1 (meaning 5-day window back in time) to lag -2 (meaning two 5-day windows back in time), with the decrease being less sharp for the more distant lags.

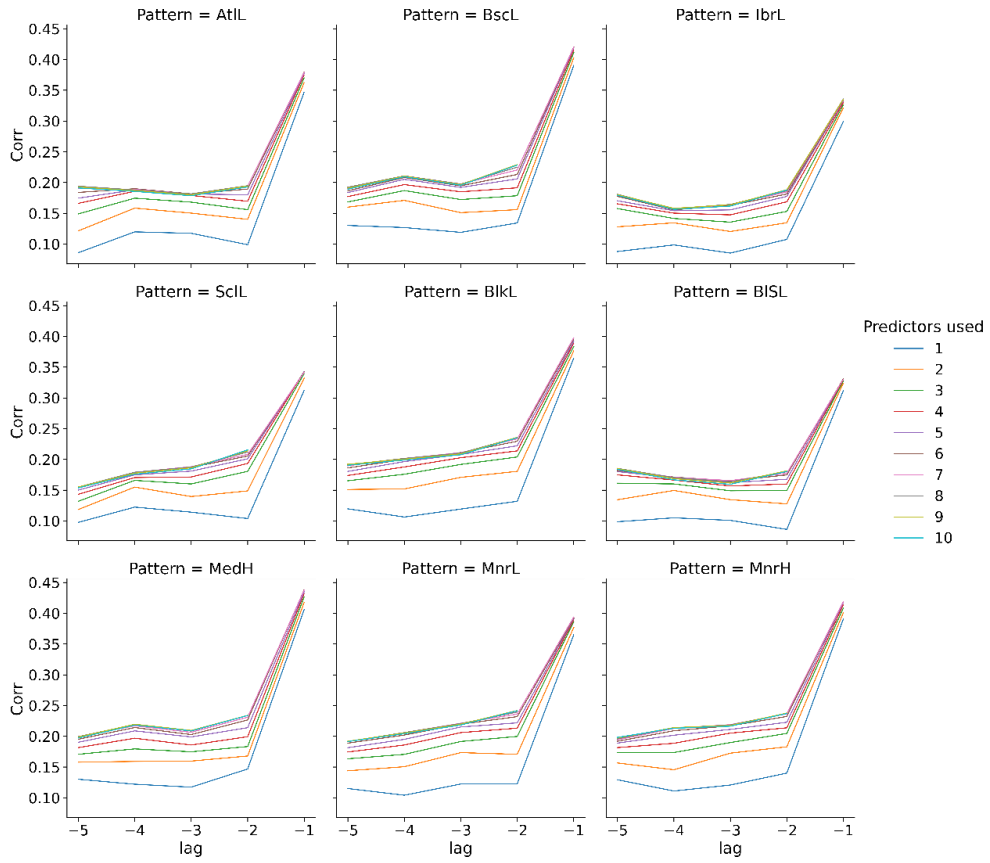

Figure 5. Lagged correlation between the Mediterranean indices and combined atmospheric indicators, based on multilinear regression model.

CAFE project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813844.

The above figure 5 informs about the magnitude of the correlation for the best linear model considering different number of indicators. However, it does not tell which indicators are used in each model. A summary of this information is provided in Figure 6. The figure shows, for the different pattern and lag, which atmospheric indicators are used for the best multilinear models. Since there are many models studied (from 1 up to 10 predictors), the results refer to the times that each indicator is used for the best models using from 1 up to 6 predictors. This information is provided in percentages, with the values varying from 0 (the indicator is not used at all), up to 1 (the indicator is used in all 6 best models for the selected Mediterranean pattern and lag).

Initially we can notice that SML_6 is not in the figure, as it is never used in any of the models. Also, there is a clear differentiation between the lag -1, and all subsequent lags. For lag -1, the predictor "Pat", meaning the lagged timeseries of the studied Mediterranean index, is used for all best models considering from 1 up to 6 predictors. This is clearly the reason of the high autocorrelation, that is key for informing about the future state of the index. Continuing for lag -1, we see that AO and MO are the other two indicators that are frequently used for the different models and for most Mediterranean indices. The state of MJO is finally another important predictor for many of the models.

For longer lead times, the first observation is that the lagged state of the Mediterranean index is not useful anymore for the vast majority of the selected multilinear regression models. Different modes of the SST variability start picking up for most of the Mediterranean patterns, while the importance of MJO is further increasing. SML1 is also used in a large portion of models, especially as we go further in time.

Despite its strong connection with atmospheric variability over the Mediterranean, the WeMO is used in only a handful of models. Moreover, as with the bivariate correlation in Figure 4, it can be noticed that NAO is used in substantially less models compared to AO, with the reason being that NAO and AO are highly correlated (corr=0.54) and that AO has stronger connections with the Mediterranean patterns, thus NAO does not provide independently any additional information on their state for most cases. Precipitation over Sahel is used only for lag -1 for some patterns, and, finally, ArcticEurope provides some useful information for the longer lead times.

Summarizing, the multilinear regression suggests that the previous state of the Mediterranean indices is the most important predictor for lags of one timestep, with the state of the atmosphere in the northern hemisphere (AO), and the tropics (MJO) being also useful predictors. For longer lead times, local information about the state of the moisture over the land and the temperature over the sea, in combination with the state of the MJO can bring some useful insights about the future state of the Mediterranean atmospheric variability.

CAFE project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813844.

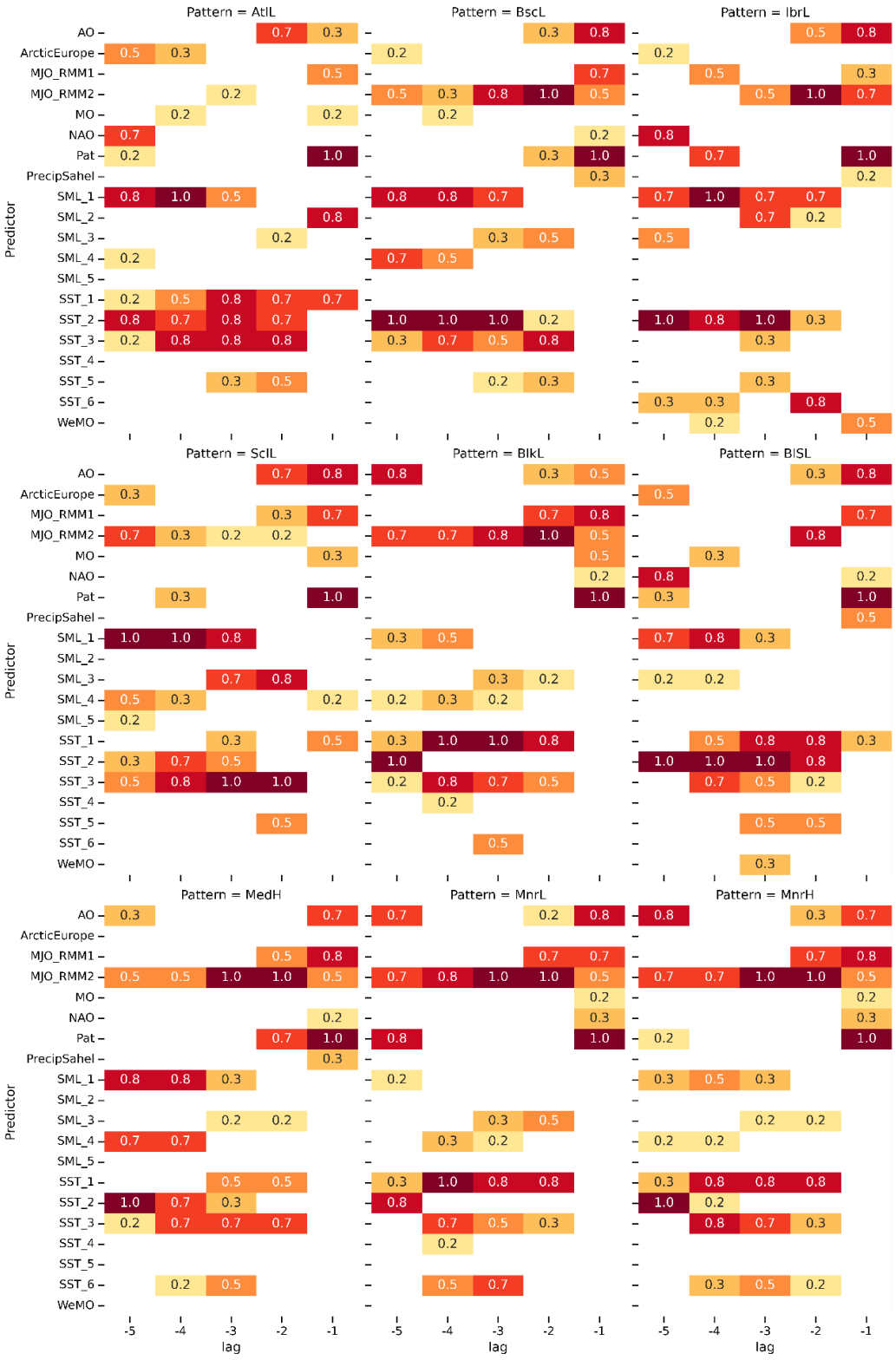

Figure 6. Percentage (0–1) of usage of each atmospheric indicator to the best multilinear regression models when using from one up to six predictors. Results for all Mediterranean patterns and negative lags.

CAFE project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813844.

Bivariate correlation and multilinear regression consider correlations between the variables. This does not necessarily infer causality. To address causal relationships, we used the PCMCI (Runge et al., 2019b) causal discovery algorithm with the help of the Tigramite package. This algorithm uses PC_1^i and subsequently momentary conditional independence (MCI) removing thus "false" connections due to common drivers, autocorrelations, and other indirect effects, depicting therefore only the causal connections in a time series dataset. The results of this analysis are shown in Figure 7. Note that for simplifying the graph, we only show the connections from the atmospheric indicators to the studied Mediterranean patterns, and all other connections are masked. We developed 9 different networks, considering the 20 indicators and each Mediterranean pattern for each network, so that connections within Mediterranean are excluded. The analysis was conducted for lags -1 up to -5.

Initially, each Mediterranean pattern is causally affected by three to six indicators. This is a very low proportion of the total 20 indicators analysed. From the colour of the nodes, it can be noticed that the autocorrelation of the atmospheric indicators is stronger compared to the one of the Mediterranean indices. This can be because atmospheric variability over the Mediterranean (Mediterranean patterns) evolves faster as compared to the remaing indicators that are of larger extend (e.g., AO, NAO) or of slower varying processes (e.g., SST, SML).

Where available, AO has a causal effect for the lag -1, while NAO, has effects only at lag -5, meaning 25 days in advance. Such distant links though should be further analysed for their robustness.

MJO has a causal effect at lag -2 for most patterns, and additionally at lag -4 for the Iberian Low. This lag -2 (10 days lead time) agrees with the study of Cassou (2008) concluding that successful forecasts of the NAO state can be derived when using the previous ~12-day MJO phase as predictor. The state of sea-surface temperature and soil moisture over the Mediterranean region has causal connections to the patterns mainly at short lags. This contradicts results of the multilinear regression that showed that the importance of these variables increases with longer lead time.

Finally, it is worth noting that MO and WeMO do not have any causal effect to the Mediterranean patterns, even though both indicators inform about the atmospheric state over the region based on SLP.

CAFE project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813844.

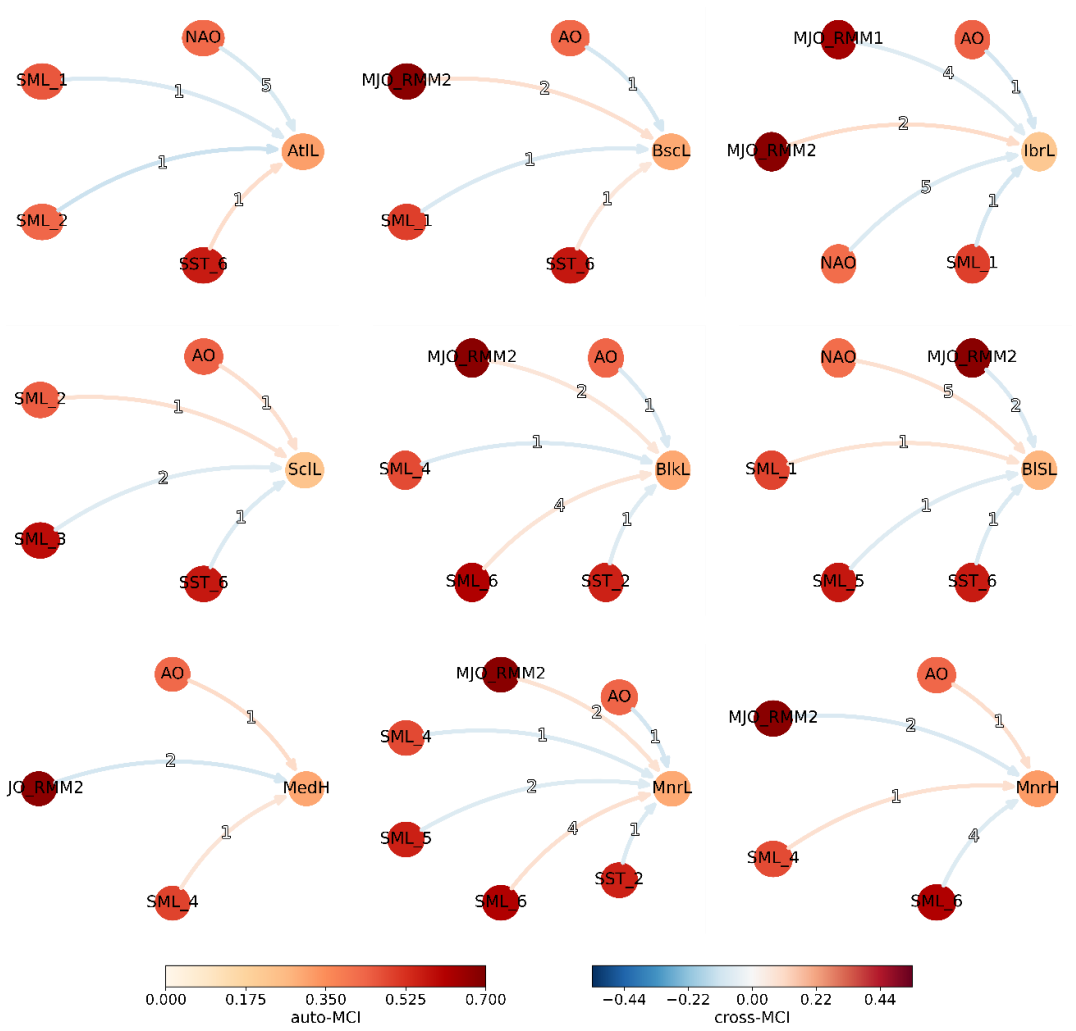

Figure 7. Causal network indicating the links from the atmospheric indicators to the Mediterranean patterns, and the autocorrelation of all studied indices. The numbers in the arrows indicate the lag of the relevant causal connection. Note that all other causal links between the indicators are masked.

4. Conclusions

This work analysed possible teleconnections and sources of predictability for the previously developed nine Mediterranean patterns. The analysis used a multitude of atmospheric indicators, either regional, as sea surface temperature and soil moisture, or more distant, as the state of Atlantic Oscillation and the Madden-Julian Oscillation. To assess possible links, we used bivariate correlation analysis, multilinear regression models, and causal networks accounting for common drivers, autocorrelation and indirect effects.

From the results, it can be concluded that the Mediterranean is considerably affected by the atmospheric state in the northern hemisphere, as depicted by the Atlantic

CAFE project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813844.

Oscillation. There is also a high degree of connection with the Madden-Julian Oscillation, a link that increases for longer lead times. Finally, the local state of the surface conditions in the sea (temperature) and over land (moisture) is also connected with the atmospheric variability over the Mediterranean and seems to be a useful predictor for longer lead times.

These findings are quite promising for providing information about the atmospheric state in the region at longer lead times. Given the fact that the Mediterranean patterns are strongly connected with extreme precipitation over the region, this information can further support actions targeting mitigation of negative consequences of such hazards.

To further advance on this topic, there are additional steps needed to strengthen the findings of this work. Initially this work used linear models to analyse connections. As the climate system is complex and non-linear, the use of non-linear tools, like mutual information, should be considered. Moreover, to develop a forecasting tool, the analysis should consider splitting the data to training/validation/testing, so that the quantification of the connections is robust and can be applied to unseen data. Of course, the latter becomes more complicated with ongoing climate change that substantially alters the behaviour of various atmospheric modes (e.g., Herrera-Lormendez et al., 2021), making the use of simple statistical tools questionable.

5. References

Cassou, C., 2008. Intraseasonal interaction between the Madden–Julian Oscillation and the North Atlantic Oscillation. Nature 455, 523–527. https://doi.org/10.1038/nature07286

Di Capua, G., Kretschmer, M., Donner, R.V., van den Hurk, B., Vellore, R., Krishnan, R., Coumou, D., 2020. Tropical and mid-latitude teleconnections interacting with the Indian summer monsoon rainfall: a theory-guided causal effect network approach. Earth Syst. Dyn. 11, 17–34. https://doi.org/10.5194/esd-11-17-2020

Herrera-Lormendez, P., Mastrantonas, N., Douville, H., Hoy, A., Matschullat, J., 2021. Synoptic circulation changes over Central Europe from 1900 to 2100: Reanalyses and Coupled Model Intercomparison Project phase 6. Int. J. Climatol. n/a. https://doi.org/10.1002/joc.7481

Jonkman, S.N., 2005. Global Perspectives on Loss of Human Life Caused by Floods. Nat. Hazards 34, 151–175. https://doi.org/10.1007/s11069-004-8891-3

Kretschmer, M., Coumou, D., Donges, J.F., Runge, J., 2016. Using Causal Effect Networks to Analyze Different Arctic Drivers of Midlatitude Winter Circulation. J. Clim. 29, 4069–4081. https://doi.org/10.1175/JCLI-D-15-0654.1

Lee, R.W., Woolnough, S.J., Charlton-Perez, A.J., Vitart, F., 2019. ENSO Modulation of MJO Teleconnections to the North Atlantic and Europe. Geophys. Res. Lett. 46, 13535–13545. https://doi.org/10.1029/2019GL084683

CAFE project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813844.

Llasat, M.C., Llasat-Botija, M., Prat, M.A., Porcú, F., Price, C., Mugnai, A., Lagouvardos, K., Kotroni, V., Katsanos, D., Michaelides, S., Yair, Y., Savvidou, K., Nicolaides, K., 2010. High-impact floods and flash floods in Mediterranean countries: the FLASH preliminary database, in: Advances in Geosciences. Presented at the 10th EGU Plinius Conference on Mediterranean Storms (2008) - 10th Plinius Conference on Mediterranean Storms, Nicosia, Cyprus, 22–24 September 2008, Copernicus GmbH, pp. 47–55. https://doi.org/10.5194/adgeo-23-47-2010

Martin-Vide, J., Lopez-Bustins, J.-A., 2006. The Western Mediterranean Oscillation and rainfall in the Iberian Peninsula. Int. J. Climatol. 26, 1455–1475. https://doi.org/10.1002/joc.1388

Mastrantonas, N., Herrera-Lormendez, P., Magnusson, L., Pappenberger, F., Matschullat, J., 2021. Extreme precipitation events in the Mediterranean: Spatiotemporal characteristics and connection to large-scale atmospheric flow patterns. Int. J. Climatol. 41, 2710–2728. https://doi.org/10.1002/joc.6985

Mastrantonas, N., Magnusson, L., Pappenberger, F., Matschullat, J., 2022. What do large-scale patterns teach us about extreme precipitation over the Mediterranean at medium- and extended-range forecasts? Q. J. R. Meteorol. Soc. 148, 875–890. https://doi.org/10.1002/qj.4236

Raicich, F., Pinardi, N., Navarra, A., 2003. Teleconnections between Indian monsoon and Sahel rainfall and the Mediterranean. Int. J. Climatol. 23, 173–186. https://doi.org/10.1002/joc.862

Runge, J., 2020. Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets, in: Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI). Presented at the Conference on Uncertainty in Artificial Intelligence, PMLR, pp. 1388–1397.

Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., Glymour, C., Kretschmer, M., Mahecha, M.D., Muñoz-Marí, J., van Nes, E.H., Peters, J., Quax, R., Reichstein, M., Scheffer, M., Schölkopf, B., Spirtes, P., Sugihara, G., Sun, J., Zhang, K., Zscheischler, J., 2019a. Inferring causation from time series in Earth system sciences. Nat. Commun. 10, 2553. https://doi.org/10.1038/s41467-019-10105-3

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., Sejdinovic, D., 2019b. Detecting and quantifying causal associations in large nonlinear time series datasets. Sci. Adv. 5, eaau4996. https://doi.org/10.1126/sciadv.aau4996

Tramblay, Y., Somot, S., 2018. Future evolution of extreme precipitation in the Mediterranean. Clim. Change 151, 289–302. https://doi.org/10.1007/s10584-018-2300-5

[i] PC acronym comes from **P**eter Spirtes and **C**lark Glymour, the names of the two researchers who invented the method.

CAFE project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813844.