



(GRANT AGREEMENT 689029)

European Climate Observations, Monitoring and Services initiative (2)

Deliverable D2.3

Final version of the empirical forecasting system

Deliverable Title	<i>Final version of the empirical forecasting system</i>	
Brief Description	<i>This report describes the final version of the empirical forecasting system, together with an evaluation of its skill.</i>	
WP number	2.4	
Lead Beneficiary	<i>[Esteban Rodríguez Guisado], AEMET</i>	
Contributors	<i>Esteban Rodríguez-Guisado, AEMET</i>	
Creation Date	<i>15/12/2020</i>	
Version Number	<i>1.0</i>	
Version Date	<i>15/12/2020</i>	
Deliverable Due Date	<i>30/09/2020</i>	
Actual Delivery Date	<i>25/12/2020</i>	
Nature of the Deliverable	<input checked="" type="checkbox"/>	<i>R – Report</i>
	<input type="checkbox"/>	<i>P - Prototype</i>
	<input type="checkbox"/>	<i>D - Demonstrator</i>
	<input type="checkbox"/>	<i>O - Other</i>
Dissemination Level/ Audience	<input checked="" type="checkbox"/>	<i>PU - Public</i>
	<input type="checkbox"/>	<i>PP - Restricted to other programme participants, including the Commission services</i>
	<input type="checkbox"/>	<i>RE - Restricted to a group specified by the consortium, including the Commission services</i>
	<input type="checkbox"/>	<i>CO - Confidential, only for members of the consortium, including the Commission services</i>

Version	Date	Modified by	Comments
1.0	17/12/2020	Esteban Rodríguez	

Table of contents

1. Executive Summary.....5

2. Detailed Report5

2.1.- Introduction..... 5

2.2. Methodology and data 7

2.2.1 Domain:..... 7

2.2.2 Predictands 7

2.2.4 Dynamical models used for skill analysis..... 9

2.2.5 Running the model 9

2.3.- Sources of predictability; choosing predictors 11

2.4.- Verification of the empirical model and comparison with dynamical models..... 15

2.5.- Discussion 22

2.6.- Conclusions..... 26

3. References.....28

1. EXECUTIVE SUMMARY

One of MEDSCOPE main goals aims at improving predictability on seasonal timescales over the Mediterranean. With this purpose, Work Package (WP) 2 is intended at gaining knowledge over possible sources of predictability and relevant mechanisms, through a series of targeted sensitivity experiments with dynamical models. Within WP2, and together with sensitivity experiments, task 2.4 is dedicated to the development of an empirical forecast system able to operationally produce seasonal forecasts, taking advantage of results from sensitivity experiments.

A preliminary version of the empirical system was developed during the first part of the project, based on multiple linear regression and global climate indices, with fixed predictors and configuration. This version showed the potential of this type of model. Details can be found at Report M2.3 and at Rodríguez-Guisado et al., 2019.

Here, a new and final version of the empirical model based on partial least squares regression is presented. Originally designed as a flexible tool able to automatically select predictors from an initial pool, the model can be run with many configurations including different predictands, resolutions, leads and aggregation times. We present here results from a configuration producing probabilistic forecasts of seasonal (3 month averages) temperature and precipitation, their verification and comparison against a selection of state-of-the-art seasonal forecast systems based on dynamical models in a hindcast period (1994-2015). The model is able to produce spatially coherent anomaly patterns, and reach levels of skill comparable to those based on dynamical models. Examples of the model usage for evaluating the impact on skill of certain predictor helping in the search and understanding of new sources of predictability are also shown.

2. DETAILED REPORT

2.1.- Introduction

Dynamical models for seasonal forecasting have noticeably improved during the last decades mainly due to the advance, both in the estimate of the atmospheric initial conditions as well as the model physics supported further by computing capabilities. However, they still show low skill over extratropical latitudes (Kim et al. 2012). As a result of this low skill over extratropical latitudes, seasonal forecasts tend to show lack of consistency among seasonal forecasting systems preventing the automatic usage of model outputs. Consequently, from an operational perspective, progress on seasonal forecasting has been historically constrained in many regions of the globe. The surroundings of the

Mediterranean Sea are specially affected by this low skill, due either to lack of predictability or to deficiencies in forecasting systems exacerbated by its complex orography and land-ocean distribution (Weisheimer et al. 2011; Doblas-Reyes et al. 2013).

The MEDSCOPE project has as one of its main objectives a better comprehension of the mechanisms driving the climate variability in the Mediterranean region. Among these mechanisms are those at the basis of the tropical and extratropical as well as polar and mid-latitudes teleconnections, and their impact on the predictability at different time scales, in particular at seasonal time scale. This better knowledge of the mechanisms behind climate predictability was achieved by a thorough evaluation of the state-of-the-art climate forecast systems and of the most recent and comprehensive observational databases and by conducting new sensitivity studies with improved or idealized representations of boundary components (land surface, tropical oceans) in participating climate models. A final by-product of the project was the development of an empirical forecast system incorporating new predictors based on the analyses of predictability sources for the Mediterranean region. This empirical forecasting system can be of practical use as a new source of seasonal forecasts for the Mediterranean region and as a benchmark for the evaluation of existing dynamical prediction systems. A preliminary version of the empirical system, based on multiple linear regression (MLR) was developed within the first part of the project. In this version, the Mediterranean domain was divided in a series of sub regions, and fixed predictors were assigned for every one of them (details can be found at Rodríguez-Gusido et al, 2019). Despite lacking skill over large areas of the domain, for certain areas and seasons the empirical system showed skill comparable or to that of the dynamical models, showing the potential for this type of model.

The final version of the empirical model has abandoned the sub regions scheme, and it has been developed independently from the original. The empirical model is programmed to serve as a flexible tool, able to automatically extract relevant information from a pool of predictor candidates. The model is prepared to perform dimensional reduction of predictors by calculation of main variability modes using Partial Least Squares (PLS) regression (Haenlein and Kaplan 2004; Wegelin 2000). The model can produce probabilistic forecasts for any variable with different lead or aggregation times. Temperature and precipitation probabilistic forecasts for three months averages and one month lead time will be shown as an example. As this is the second version of the empirical model presented, this final version has been named “AEMET-S2”, and it will be consequently named hereafter.

This text is structured as follows: in Section 2, the method and datasets are described; sources of predictability and selection of predictors are described in Section 3; verification of the empirical model and comparison with dynamical

models are presented in Section 4; a discussion of results is included in Section 5; and finally conclusions are summarized in Section 6.

2.2. Methodology and data

2.2.1 Domain:

As the MEDSCOPE project is focused on the Mediterranean area, all results shown here will be presented in a domain defined by 15W, 50E, 27.5N, 55N (Fig 1a). However, the forecast domain is defined as an input parameter in the code and can be easily changed. In order to support activities to other RCOFs, all calculations were conducted using an extended domain (20W, 60E, 12.5N, 55N) (Fig 1b).

a)

b)

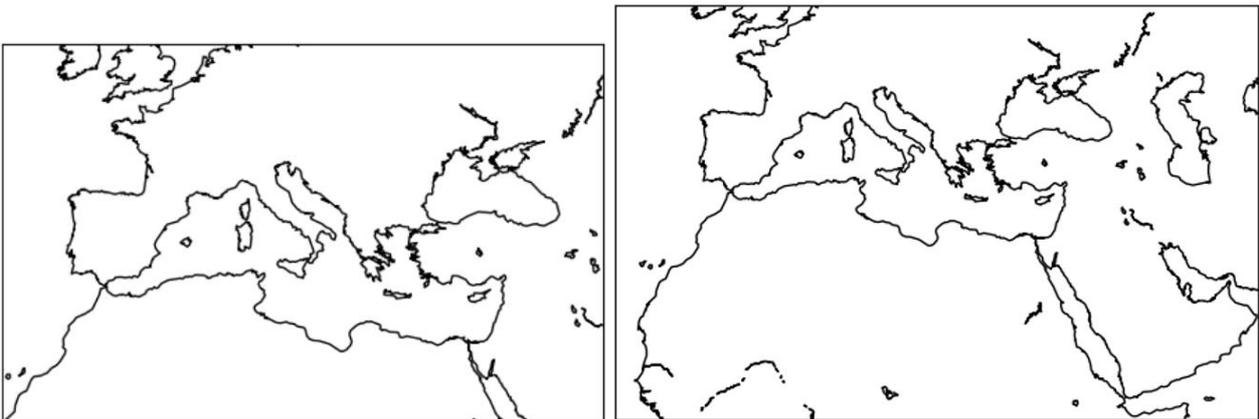


Fig 1. a) Domain selected; b) extended domain

2.2.2 Predictands

As most seasonal forecasts are based on limited to surface temperature and precipitation, and as these variables are mostly analysed in sensitivity experiments, the developed method and code will be focused on them. Anyway, predictands can also be selected as input parameters. Lead time of forecasts and time aggregation (e.g., 1 month, 3 months) of predictands can be selected too. Here, forecast for three months average of predictands and for one month lead time (i.e., last predictor data available is two months before forecast date. For example, if we are producing a forecast for February-April, predictors will be available until December) will be presented.

As verifying data, given the differences in horizontal variability of temperature and precipitation and also the different observational coverage and quality, we have decided to use reanalysis data for temperature and observational gridded datasets for precipitation. However, we are aware that observational gridded datasets lack in situ observations over many areas. Figure 2 shows data availability for GPCC (a) and for EOBS (b) datasets. As can be seen in the figure, there are large parts of the domain that are not covered by observations. Usually, gaps are filled by statistical procedures or by satellite data. As the PLS regression method here applied calculates predictors decomposition based on covariance between predictors and predictands, artificially filled data can affect the way this decomposition is calculated, and so, affect forecasts skill over regions actually with good observational coverage. For that reason, datasets using satellite data are preferred over the ones using statistical procedures to fill the gaps, and GPCPv2.3 have been chosen for precipitation (Adler et al. 2003. <https://psl.noaa.gov/data/gridded/data.gpcp.html>)

In a similar way, resolution of the predictands can affect the way in which predictors are decomposed: higher resolution predictands can capture variability modes of smaller scale. (WMO, 2020) recommends to perform an analysis of general circulation anomalies and main variability modes when producing a seasonal forecast, so low resolution predictands will be used for the version presented here, aiming at focusing on large scale variability. However, higher resolution predictands can be used when producing forecasts for reduced areas and there is interest on local features

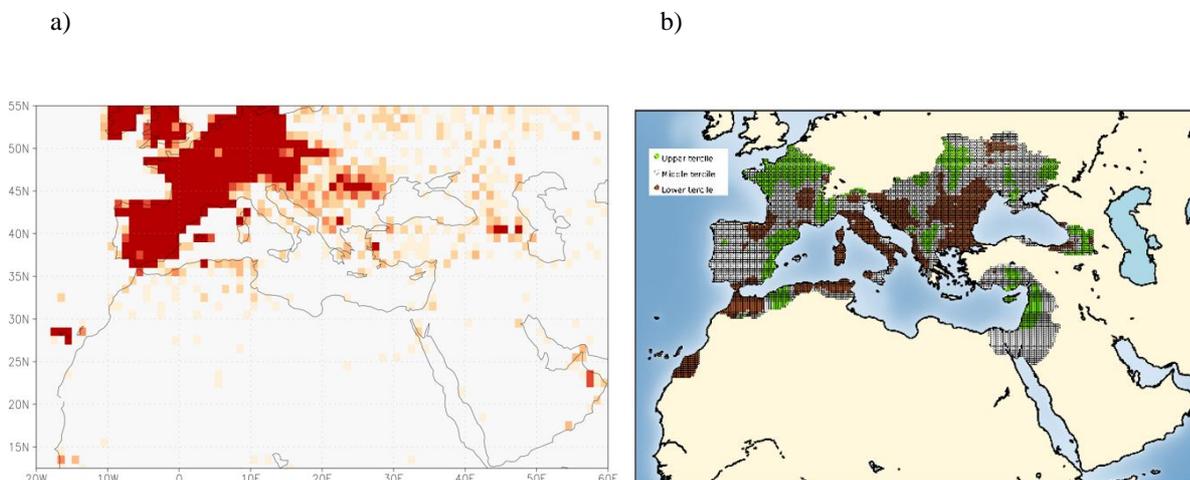


Fig 2. a) Availability of observations on every grid point on the Global Climatology Precipitation Center (GPCC) version 8 data set for December 2016; b) EOBS data available data for 2019-2020 winter (DJF).

Finally, the following data has been chosen for the predictands:

-Temperature: ERA5 2 metres temperature, monthly averaged, 2.5x2.5 degrees, from 1979 to 2019, C3S (2017).

<https://cds.climate.copernicus.eu/cdsapp#!/home>

-Precipitation: Global Precipitation Climatology Project (GPCP), monthly averaged, 2.5x2.5 degrees, from 1979 to 2019 (Adler et al, 2003, <https://psl.noaa.gov/data/gridded/data.gpcp.html>)

2.2.4 Dynamical models used for skill analysis.

In section 4, results will be presented and forecast maps produced will be analysed. Together with every forecast, score maps are generated to provide information about the skill of the empirical model. However, to put it in perspective, skill needs to be compared to that of dynamical models currently used for operational seasonal forecasting. For that purpose, scores are calculated too for Copernicus C3S models, in particular for versions on the following list:

-ECMWF-S5: <https://gmd.copernicus.org/articles/12/1087/2019/gmd-12-1087-2019.html>

-DWD-S2: https://www.dwd.de/EN/ourservices/seasonals_forecasts/project_description.html

-CMCC-S3: <https://sps.cmcc.it/documentation/>

-MF-S6: http://seasonal.meteo.fr/sites/data/Documentation/doc_modele/Model_MF-S6_C3S_technical_en.pdf

-UKMO-S13 (GloSea5): <https://www.metoffice.gov.uk/research/climate/seasonal-to-decadal/gpc-outlooks/user-guide/technical-glosea5>

2.2.5 Running the model

Given the potential high number of candidates in the initial pool of potential predictors, a reduction of dimensionality is frequently recommended before starting any computation (see, e.g., WMO 2020). Considering that predictands are 2 dimensional spatial fields, is also advisable to incorporate information both on predictands themselves and their covariance. In order to meet both requirements we have made use in the empirical model of the PLS-regression method (Haenlein and Kaplan 2004, Wegelin 2000) as it is available on python sklearn 0.23.2 module (https://scikit-learn.org/stable/modules/generated/sklearn.cross_decomposition.PLSRegression.html). This method performs a

decomposition of the predictor space in a similar way as EOFs. The EOF analysis allows a set of predictors to be rearranged into a new set of predictors, orthogonal with each other and that maximizes the amount of predictors variance that can be explained with the smallest number of EOF predictors. The PLS-regression method, instead of fixing the amount of variance explained makes use of an iterative method that fixes as an input parameter the number of factors (or independent vectors) used (Haenlein and Kaplan 2004), and maximizes the amount of predictand variance explained using those factors. Previously, and to reduce the initial number of predictors, only those surpassing a minimum correlation threshold with variability modes of the predictand are retained and incorporated into the algorithm system.

So, as we have two different parameters that can be set arbitrarily, forecast is performed for several configurations of both of them. A maximum number of components n (4 for the version presented here) is introduced as input parameter, and three values of correlation c (0.35, 0.42, 0.49). Information for the year “ i ” to be forecasted is removed from the datasets. Then, for the remaining $(n-1)$ years, a hindcast H is calculated using Leave-One-Out (LOO) technique, training the model with every possible combination of c and n . Every time the model is trained with a set of $n-2$ years and a (c,n) configuration, a forecast F for the year i is produced. Then, at every grid point, correlation between every H and the predictand is calculated, and those combinations below percentile 80 of correlation are discarded. Forecast ensemble for the year i will be composed of F values for the remaining (c,n) combinations. The process is repeated for every year i of the series, to generate a hindcast that allows skill evaluation.

Trends (e.g. global warming trend affecting temperature) are removed from both predictands and predictors previously to run the model and the predictand trend is added at the end of the computation.

Parameters like lead time of the forecast, aggregation time of the predictand, resolution...can be introduced as input parameters. To help interpret the forecast produced for a particular configuration, skill maps are calculated and displayed based on its hindcast.

2.3.- Sources of predictability; choosing predictors

The model automatically selects and combines predictors from an initial pool. Table 1 shows the predictors used for this version of the model. A comprehensive list of global climate indices has been selected, including main teleconnection

patterns from the Climate Prediction Center (CPC, <https://www.cpc.ncep.noaa.gov/>), SST-based indices, ENSO indices like SOI and MEI, and snow cover indices. Additionally, some specific predictors relevant for the Mediterranean region have been included. All SST based indices (with the exception of WHWP, which is downloaded from NOAA's CPC) are calculated from ERA5 SST data (downloaded from <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels-monthly-means?tab=form>), averaging over the areas detailed in Table 1. The same latitude-longitude areas are used to calculate averages of surface layer heat content. Heat content is calculated from MetOffice EN4.2.1 reanalysis, data (<https://www.metoffice.gov.uk/hadobs/en4/download-en4-2-1.html>, Levitus et al. 2009), averaging temperature data for a layer between surface and 200m depth.

Snow based indices include snow cover extent (Robinson et al. 2012) over Northern Hemisphere, Eurasia and North America and averages of it over certain regions of Siberia, North America and Scandinavia that show good correlation with winter North Atlantic Oscillation (NAO). Averages have been produced and downloaded at https://climexp.knmi.nl/select.cgi?id=someone@somewhere&field=rutgers_nhsnow

Averages of zonal wind at 10 hPa both over polar areas and over the tropics (to represent polar vortex variability and Quasi Biennial Oscillation (QBO)) are incorporated into the pool of predictors, too (see Table 1). ERA5 monthly averaged zonal wind have been used to produce these predictors.

Sensitivity experiments in the frame of the MEDSCOPE project (Ardiluoze et al. 2020) have shown the impact of May soil moisture in the intensity and frequency of heat waves, so modes of variability (rotated main EOFs (Barnston and Livezey 1987) for every month) for soil moisture on a window defined between 30N-60N, 10W-40E have been included. Soil moisture data is derived from ERA5 volumetric soil water layer, as the average of the three first layers available.

As many predictors are climate indices, which are calculated from monthly anomalies, predictors will be also based on monthly averaged values. In addition to series based on their monthly values, an incremental series is constructed for every index, based on the difference between one month and the previous one values. This is done under the assumption that some particular climate system processes are able to provide predictability not only by the absolute value of a predictor, but by its rate of change as well. An example of this behaviour is the Snow Advance Index (Cohen and Jones 2011), where the predictor is based on the difference in snow extent at two different times.

	Acron ym	Index name	Source
Teleconnections (z500 patterns)	AAO	Antartic Oscillation	https://psl.noaa.gov/data/correlation/aao.data
	AO	Artic Oscillation	https://psl.noaa.gov/data/correlation/ao.data
	EA	East Atlantic	ftp://ftp.cpc.ncep.noaa.gov/wd52dg/data/indices/ea_index.tim
	EA/W R	East Asia/West Russia	https://psl.noaa.gov/data/correlation/ea.data
	EP/NP	East Pacific/No rth Pacific	https://psl.noaa.gov/data/correlation/epo.data
	NAO	North Atlantic Oscillation	https://psl.noaa.gov/data/correlation/nao.data
	NP	North Pacific	https://psl.noaa.gov/data/correlation/np.data
	PNA	Pacific North American	https://psl.noaa.gov/data/correlation/pna.data
	SCAN D	Scandina vian	ftp://ftp.cpc.ncep.noaa.gov/wd52dg/data/indices/scand_index.tim
	WP	West Pacific	https://psl.noaa.gov/data/correlation/wp.data
Stratosphere	QBO	Quasi Biennial Oscillation	Averaged u wind (10S-10N at 10hPa)
	u10n	Northern u at 10hPa	Averaged u wind (60N-85N at 10 hPa)
	u10s	Southern u at 10hPa	Averaged u wind (85S-60S at 10 hPa)
ENSO	MEIv2	Multivariat e ENSO index	https://psl.noaa.gov/enso/mei/data/meiv2.data

	SOI	Southern Oscillation index	https://psl.noaa.gov/data/correlation/soi.data
Indian Ocean	WTIO	Western Tropical Indian Ocean	Averaged (50E-70E,10S-10N)
	SWIO	South Western Indian Ocean	Averaged (31E-45E,32S-25S)
	SETIO	South Eastern Tropical Indian Ocean	Averaged (90E-110E,10S-0)
	DMI	Dipole Mode Index	Calculated (WTIO-SETIO)
Pacific Ocean	Nino1 2	Nino12	Averaged (90W-80W,10S-0)
	Nino3	Nino3	Averaged (150W-90W,5S-5N)
	Nino4	Nino4	Averaged (160E-150W,5S-5N)
	Nino3 4	Nino34	Averaged (170E-120W,5S-5N)
	Nino3 4- Nino1 2	Nino34 - Nino12	Calculated
Atlantic+Mediterranean Oceans	TNA	Tropical Northern Atlantic	Averaged (57.5W-15W, 5.5N-23.5N)
	TSA	Tropical Southern Atlantic	Averaged (30W-10E, 20S-0)
	NAT	North Atlantic Tropical	Averaged (40W-20W, 5N-20N)
	SAT	South Tropical Atlantic	Averaged (15W-5E, 20S-5N)

	TASI	NAT - SAT	Calculated
	Atl1	Locally derived	Averaged (40W-10W, 50N-60N)
	Med1	Locally derived	Averaged (0-8E, 33N-46N)
	Med2	Locally derived	Averaged (10E-22E, 33N-46N)
	Med3	Locally derived	Averaged (25E-40E, 30N-40N)
	WHWP	Western Hemisphere Warm Pool	https://psl.noaa.gov/data/correlation/whwp.data
Snow extent	SnNH	Snow extent (N. Hemisphere)	https://climexp.knmi.nl/getindices.cgi?WMO=RutgersData/nh_snow&STATION=NH_snow_cover&TYPE=i&id=someone@somewhere
	SnNA	Snow extent (North America)	https://climexp.knmi.nl/getindices.cgi?WMO=RutgersData/namerica_snow&STATION=North_America_snow_cover&TYPE=i&id=someone@somewhere
	SnEA	Snow extent (Eurasia)	https://climexp.knmi.nl/getindices.cgi?WMO=RutgersData/eurasia_snow&STATION=Eurasia_snow_cover&TYPE=i&id=someone@somewhere
	Snow 1	Locally derived	Averaged (10E-20E, 58N-67N)
	Snow 2	Locally derived	Averaged (110W-85W, 38N-50N)
	Snow 3	Locally derived	Averaged (38E-67E, 38N-54N)
	Snow 4	Locally derived	Averaged (75E-145E, 25N-55N)
	Snow 3-Snow 4	Locally derived	Calculated locally (Snow3-Snow4)
	Soil_wetness	Variability modes	Rotated PCA from monthly anomalies (calculated 10W-40E, 30N-60N box)

Table1. List of predictors used. All oceanic averaged indices are calculated both for Sea Surface Temperatures (SST from ERA5, downloaded from Climate Data Store). Snow averages have been calculated at https://climexp.knmi.nl/select.cgi?id=someone@somewhere&field=rutgers_nhsnow). Soil wetness variability modes have been calculated from monthly anomalies of ERA5 volumetric soil water layer (average of layers 1, 2 and 3)

2.4.- Verification of the empirical model and comparison with dynamical models

Figures 3 and 4 show some examples of the empirical model operational display. Figure 3 shows the probability of the most likely tercile for temperature forecast produced by the model and plotted together with a skill map (Random Probability Skill Score, RPSS) based on the hindcast. Figure 4 shows forecasted precipitation probability for upper and lower terciles again plotted together with skill maps (ROC area). Besides, a dry mask has been applied for those grid points recording less than 2 mm over the season during at least 10 years during the hindcast period.

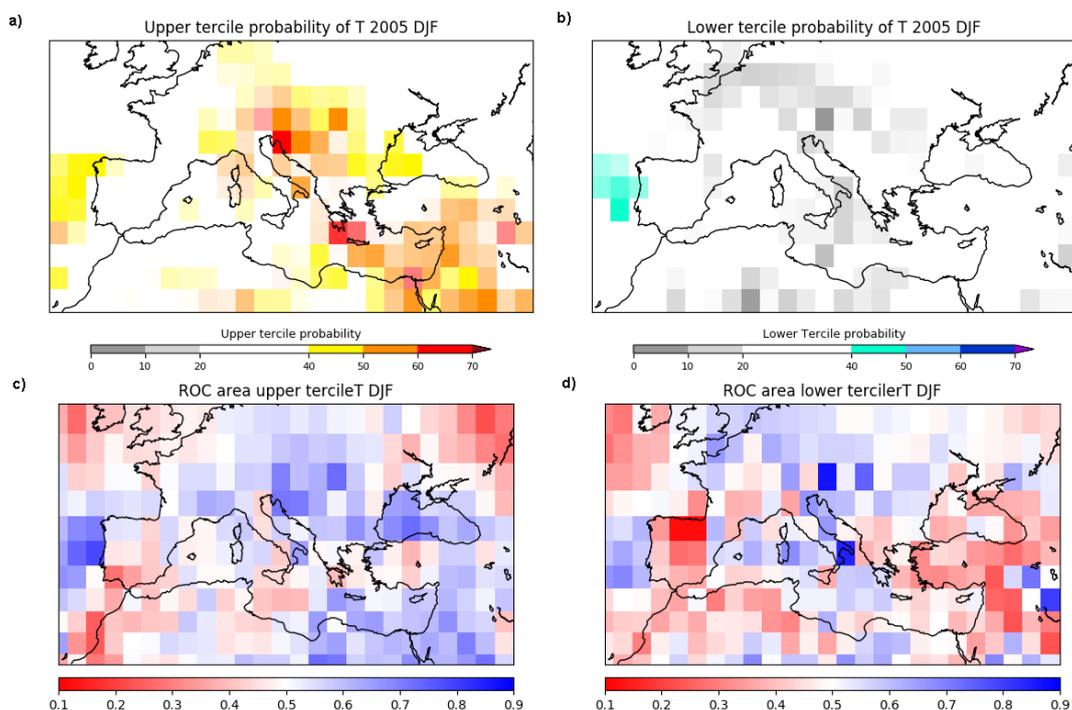


Fig 3. Example of forecasted temperature. Probability of upper (a) /lower (b) tercile, with transparency layer based on ROC area score. ROC area score for upper (c) /lower (d) tercile (hindcast period 1993-2019)

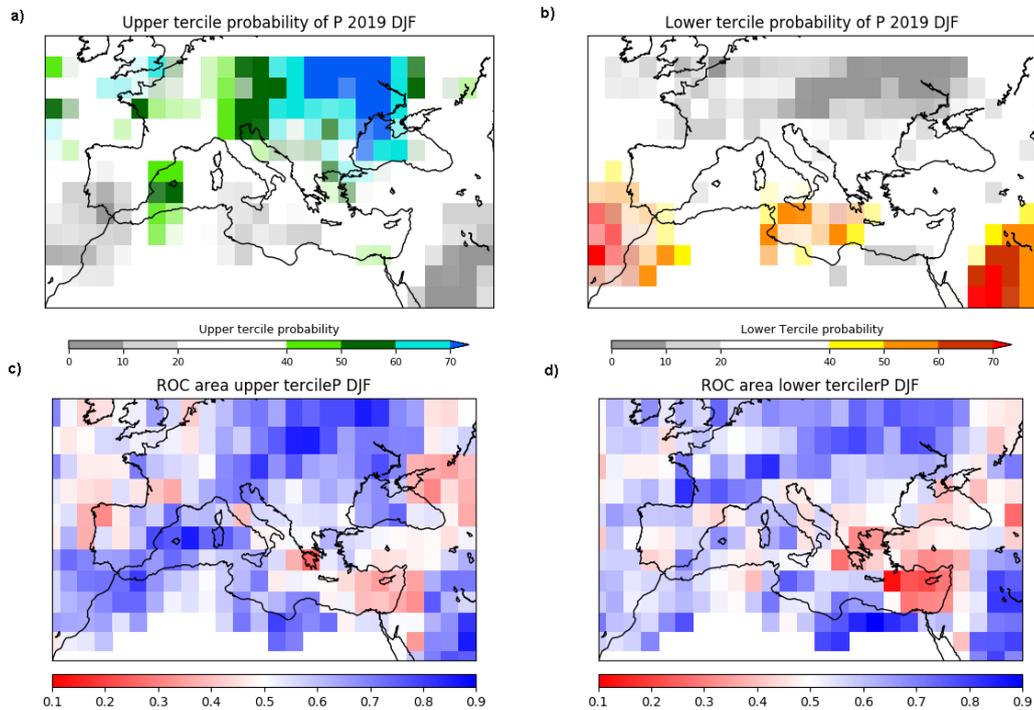


Fig 4. The same as Figure 3 but for precipitation.

As explained in section 2.2, initially low resolution predictands have been chosen aiming at producing forecasts focused on large scale anomaly patterns. Watching at Figures 3 and 4, relatively smooth maps can be appreciated, showing synoptic scale patterns. Skill maps suggest, as happens with dynamical models, that skill is better for temperature than for precipitation, and that it varies from one region to another. However, areas with high/low skill seem to be distributed following synoptic scale patterns too.

Higher temperature skill could be related as well to trend, which is removed both from predictands and from predictors before any calculation. After removing it, the model is run, and then predictand trend is added to the results, so a significant part of temperature may come from it.

Although the model version presented here makes use of large scale predictands, if higher resolution data are available over smaller domains, the system can also be run using them, as shown in Fig. 5. That may allow the model to capture locally relevant modes of variability, improving skill over the large scale version.

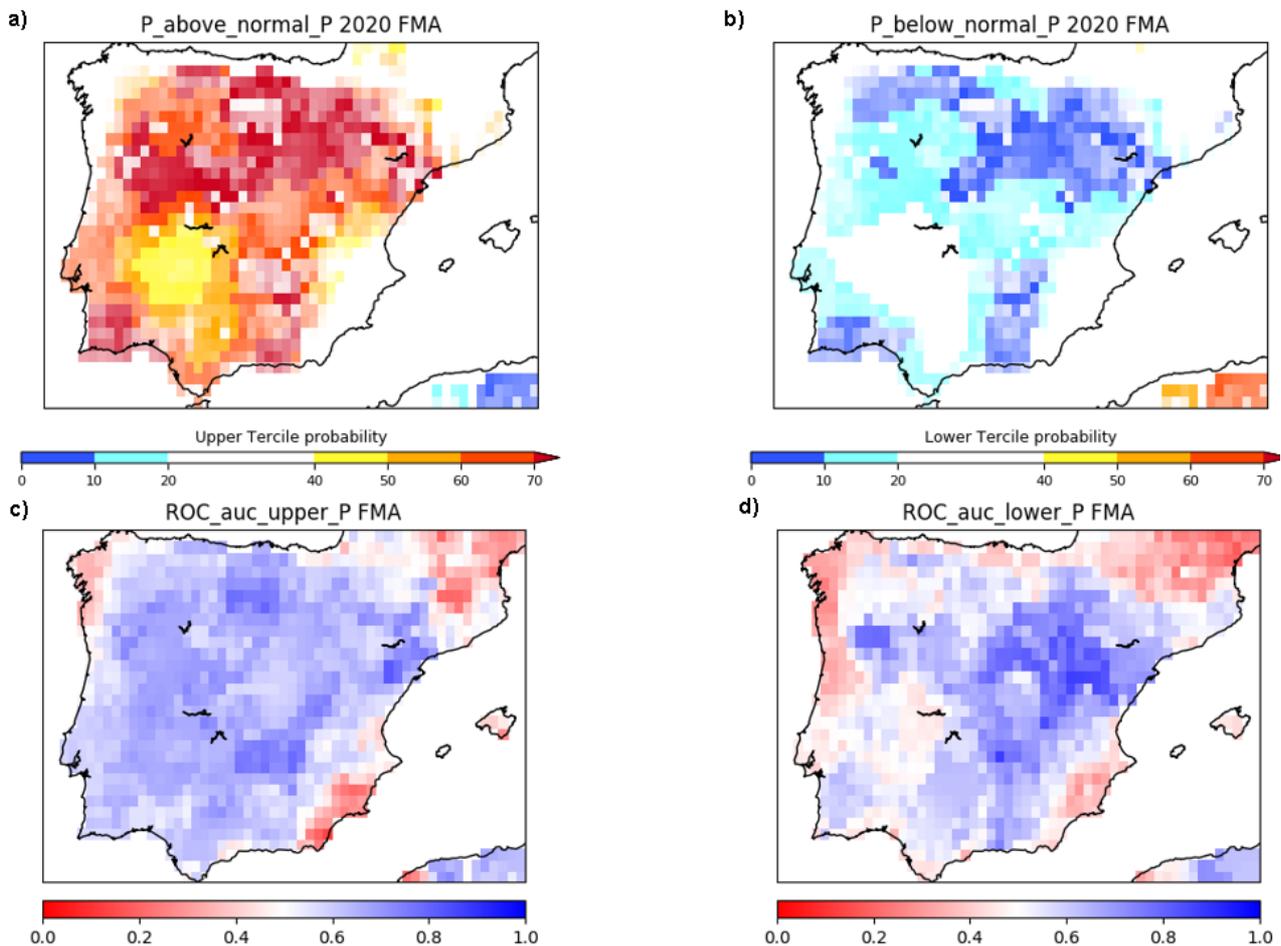
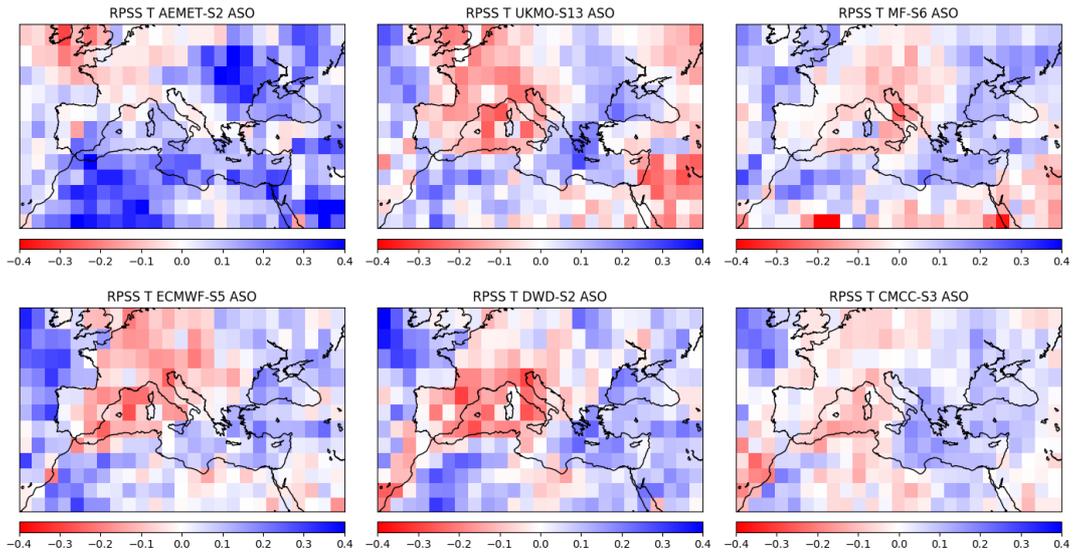


Fig 5. Same as figure 4, but for a smaller domain (using EOBS-21e (0.25), Cornes et al. (2018) as predictand)

Consequently, and looking at results, the model seems to produce coherent anomaly patterns, and to have skill over certain areas and seasons. To analyse how it performs over the domain, skill scores have been calculated for both the empirical system and some selected state-of-the-art dynamical seasonal forecasting systems. Figure 6 shows RPSS maps for June-July-August (JJA) for temperature and precipitation. For all models, temperature forecasts show more skill,

and seem to have less skill over Atlantic shores. For precipitation, skill varies from one area to another, but the empirical model seems to perform at a level comparable to that of dynamical models.

a)



b)

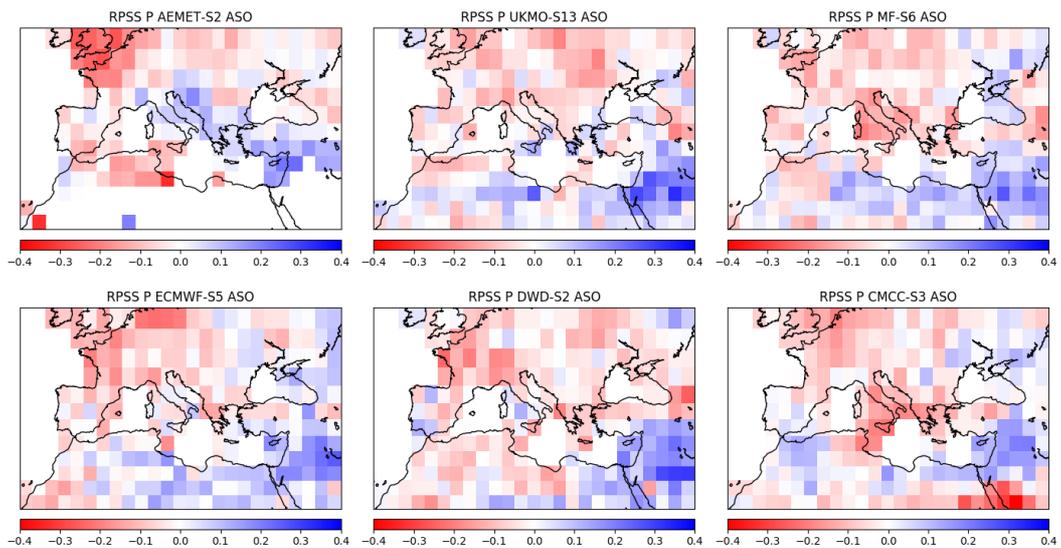


Fig 6: August-October RPSS for AEMET-S2 and Copernicus models for temperature (a) and precipitation (b)

However, skill changes drastically among seasons. To have a clearer picture of how models perform over the year, scores have been averaged for the domain and compiled in Table 1 representing RPSS area values for temperature and precipitation averaged over the domain. For temperature, the model seems to have more skill in summer months. For those months, skill is comparable to dynamical systems. Furthermore, model tends to show more/less skill for those months where dynamical models do, so it seems to be able to capture predictability when available

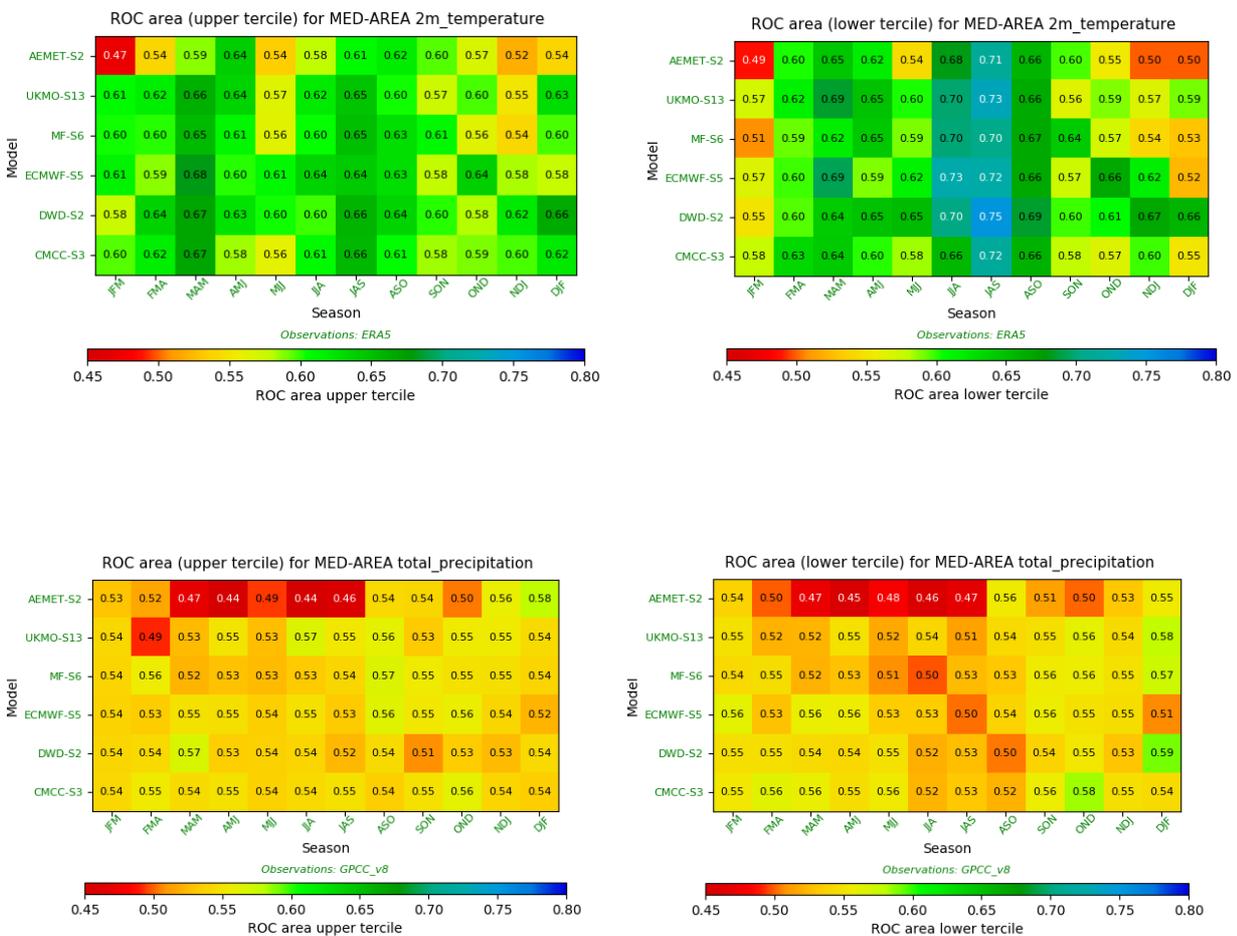


Figure 7. Skill comparison with C3S models averaged for the domain (15W-52.5E, 25.0-50N). Every column represents a season, and every row a model. Period is 1994-2015. Scores have been calculated using non-parametric bootstrapping (1000 samples)

As seen in Fig 6, skill varies among areas, so it can be interesting to do the same exercise but averaging over more reduced areas. Fig 8 shows a couple of examples, for the Balkans and for Iberian Peninsula. Although behaviour for precipitation is less consistent than for temperature, over certain areas and seasons, skill for precipitation can be at the level or even slightly higher as for dynamical models.

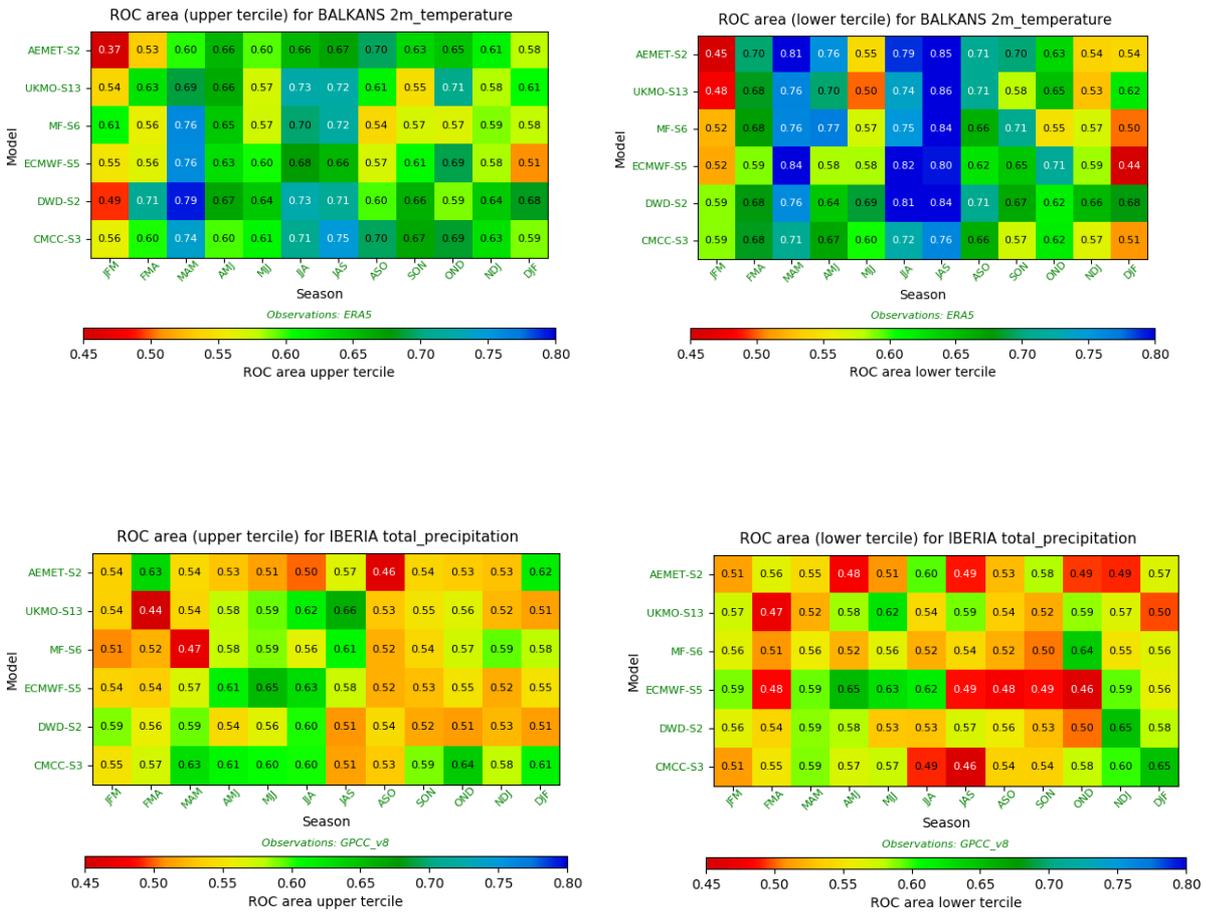


Figure 8. Same as figure 6, but for selected domains: Iberia (36-44-N, 10W-4E, (a)) and the Balkans (35-50N, 15-30E, (b))

2.5.- Discussion

Although the system is prepared to add predictors as new sources of predictability emerge, the version of the empirical model here evaluated represents an additional source of information to be integrated together with systems based on dynamical models when producing subjective or objective forecasting. As it has been shown in Sec 4 this empirical model is able to reach similar levels of skill as dynamical models. As can be appreciated from figure 8 (c and d)), in many occasions, the empirical model tends to show higher skill for those seasons and regions where dynamical models do, and vice versa. It seems that, in those cases, dynamical models are able to capture, at least partially, the same sources of predictability contemplated by the empirical system. Moreover, the empirical model shows for certain regions and seasons some amount of skill where dynamical models fail. Future work should include a more detailed analysis of cases where discrepancies appear in order to identify new predictors that may contribute to the improvement of forecasts and gain understanding on processes related to sources of predictability.

In fact, a better understanding of predictability sources was one of the goals of the MEDSCOPE project and different sensitivity experiments were conducted to analyse the role of several sources of predictability (see other contributions in this special issue). The empirical model was conceived not only as an operational system capable of producing seasonal forecasts but also as a tool -because of the automatic selection of predictors from a pool- to analyse the impact on skill of one particular predictor (or a group of them) by just adding or removing it from the pool, and comparing the skill.

One of the sources of predictability analysed in the MEDSCOPE project is related with soil moisture in spring. Ardiluoze et al. (2020) have shown that May soil moisture can influence the frequency and intensity of summer heat waves. Following these results, soil moisture data from ERA5 reanalysis were included in the initial predictors pool. As an example of usage of this empirical model as a diagnostic tool in the search of predictability sources, we have included an analysis of soil moisture influence in temperature forecasts. Figure 9 shows the different skill of the forecasted temperature upper tercile, including (a) vs not including (b) soil moisture information on the initial pool. An increment can be seen over large parts of Continental Europe. The impact of including soil moisture can be seen more clearly on Fig 10, where differences of two maps in Fig 9 are plotted. Positive values indicate areas where skill is higher when including soil moisture information. Most continental Europe shows a clear improvement of the warmer tercile skill when soil moisture information is added. Iberian Peninsula, Western France and Turkey areas constitute an

exception, where the influence of the SST and Atlantic variability (in the case of Iberian Peninsula and France) can be dominating.

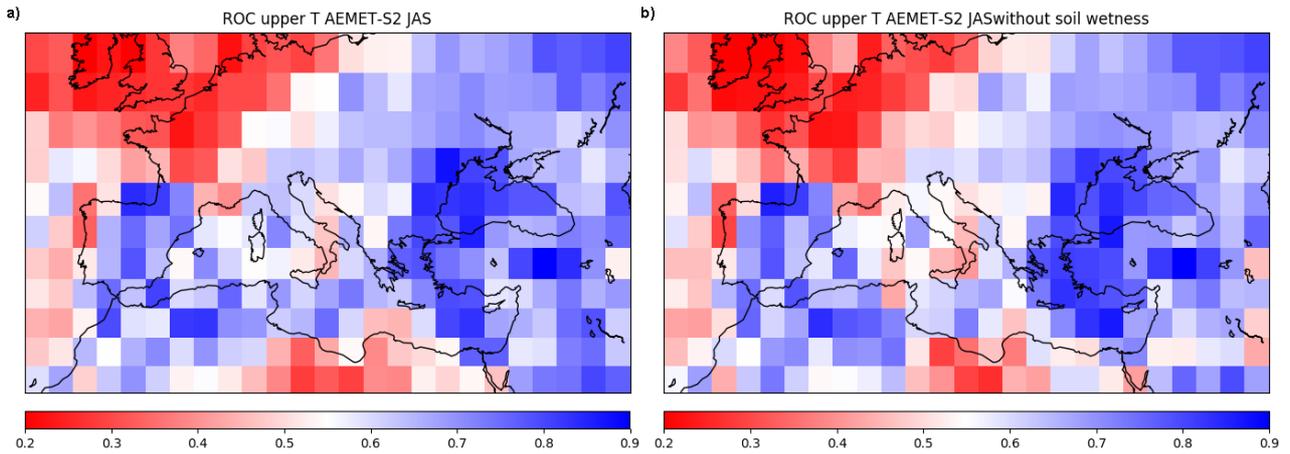


Figure 9: Comparison of temperature upper tercile skill (JAS) with (a) and without (b) soil wetness as predictor

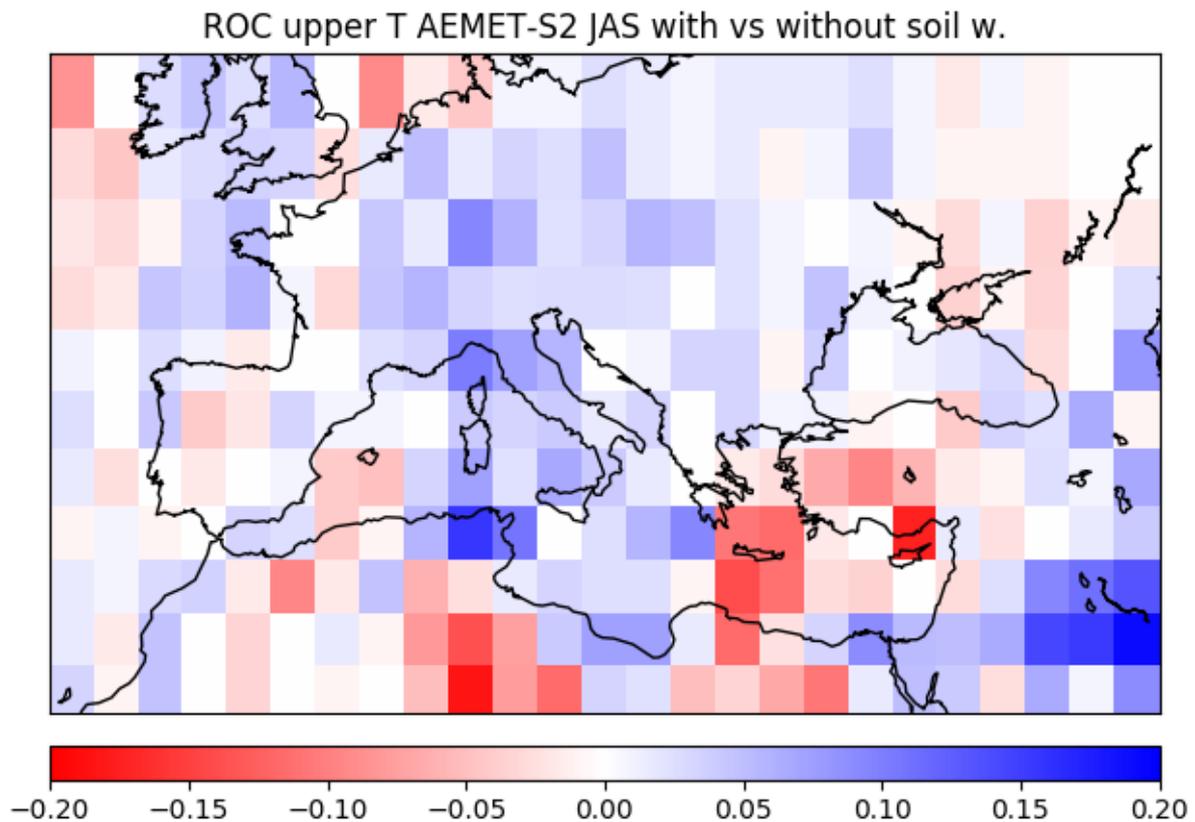


Figure 10: Difference between both figure 9 maps (a-b). Positive values represent improvement adding soil moisture information.

This analysis can be easily repeated for other predictors. Of course, there are many other factors that can influence skill, but we think it can be a useful tool for analysing a potential source of predictability.

However, for some seasons and areas, the model still lacks skill. Besides the insufficient identification of predictors for such areas and seasons, another possibility explaining this behaviour may lay on the linear relationships assumed by the model. Furthermore, we are assuming that teleconnections are stationary, and this could not be the case for some predictors. For example, winter has been recently considered as a “window of opportunity” for predictability in the Western European facade manifested by a better forecasting of the winter North Atlantic Oscillation (NAO) by dynamical models and the identified correlation between winter NAO and predictors based on autumn snow cover advance and sea ice. However, the nature of the associated mechanisms seems to change in function of the period

studied (Douville et al, 2017, Wegmann et al, 2020). One possibility to explain that variation could be just that long term variability may affect these teleconnections (Wegmann et al, 2020). To better illustrate this change of behaviour in snow teleconnections, Fig 11 shows correlation between winter (DJF) NAO and autumn snow cover over North America. Differences between 1980-2000 and 2005-2019 periods are evident, consequently a linear model based on snow cover over this region trained during the first period will consistently fail on its forecast during the second. Additionally, it could be hypothesized that rapid changes experienced in the Arctic regions during recent years may have driven the climate system to a new state with interactions and processes different to the ones observed in the past, used in the training of the model. Rapid changes, on the other hand, tend to be associated with non-linear processes, which are impossible to mimic using a linear model.

As expected -especially in a climate change context- results seem to be sensitive to the training period. A difficult balance between a period long enough for reliable fitting and at the same time representative of the forecasted conditions period has to be achieved. A future version of the model could incorporate cross validation tests to identify the best fitting period for every season.

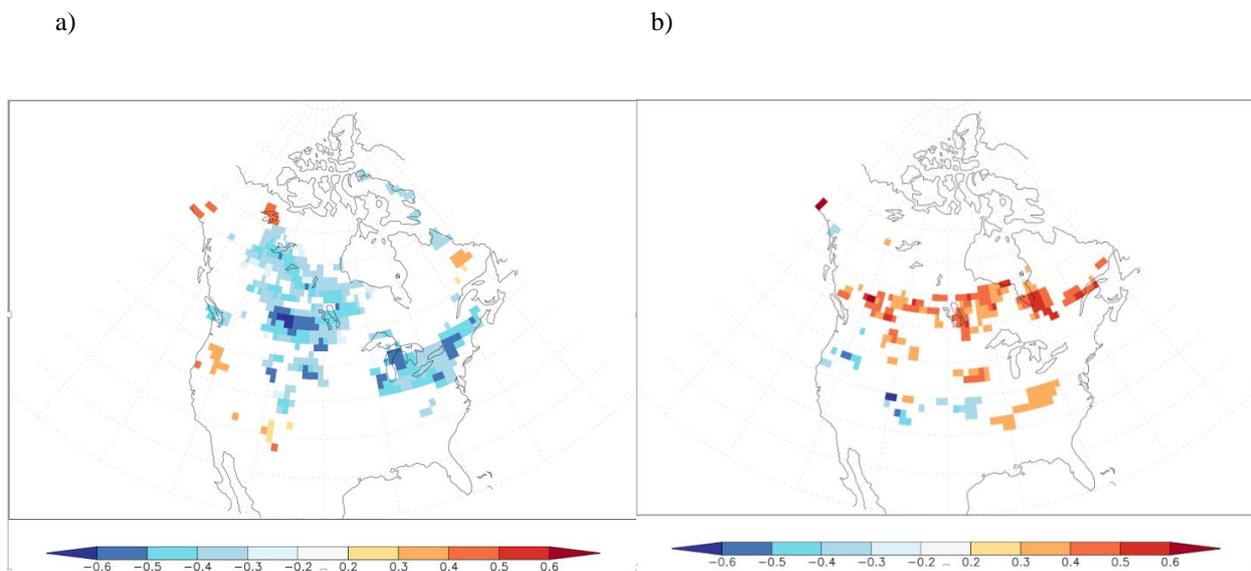


Figure 11. Correlation between winter (DJF) NAO and Snow Cover Extent (Rutgers University Snow Lab (<https://climate.rutgers.edu/snowcover/>) over North America, from 1980 to 2000 (a) and from 2005 to 2019 (b).

2.6.- Conclusions

The final version of an empirical model developed within the framework of the MEDSCOPE project has been presented and verified for the Mediterranean domain. The model is able to produce spatially coherent anomaly patterns, and reach levels of skill comparable to those of seasonal forecasting systems based on dynamical models. The comparable skill of the empirical model described here and state-of-the-art dynamical systems permits its usage as an additional source of information when producing operational seasonal forecasts combining different models either subjectively or making use of some combination algorithm (Sánchez García et al. 2019). Pros and cons of empirical and dynamical seasonal forecasting systems have been analysed and discussed for many authors (Mason and Baddour 2008; WMO 2020) and many of them insist in the complementarity of both approaches. In favour of the empirical systems in general - and of this system in particular- we can mention that: first, they only need modest computational resources; second, the identification and selection of predictors allows a straightforward way for the predictability and skill attribution of forecasts at seasonal scale; and third, unveiling of new sources of predictability over the region can be very easily incorporated to the system simply by adding new predictors to the initial pool. Among the disadvantages of the empirical systems, it is widely acknowledged that: first, their quality heavily depends on the identification, selection and quality of predictors; and second, they are usually developed and tuned only for specific regions.

It has been shown that skill varies among variables, areas and seasons, as also happens with dynamical systems. In order to circumvent these differences in skill and to facilitate the operational use of model outputs, the system always incorporates verification maps attached to each seasonal forecast. Also forecast maps make use of a transparency layer based on skill over a hindcast period to highlight/fade out grid points showing some/no skill.

Additionally, the model has flexibility to be used with different resolution predictands, lead and aggregation times. Predictors are selected automatically from an initial pool, so the model can also be used to evaluate the impact on skill of a certain predictor, and help in the search and understanding of new sources of predictability. Tests done analysing the impact of soil moisture over summer temperature support results obtained in sensitivity experiments done with complex dynamical models.

Future work should incorporate an analysis of the seasons with low skill, and the possible impact of changes on the Arctic region and the length of the training period.

We think there is still room for improvement focusing on the search of new predictors. On this direction, a tool for spatial exploration is currently being developed, intending to identify areas that show stable correlation between predictor and predictand on a particular period, and automatically generate additional predictors whose impact can be tested using the model. This tool would have the additional advantage of helping to make the system less region-dependant, as automatically could search for relevant predictors for the region of study, mitigating one of the typical disadvantages of empirical models mentioned earlier.

A manuscript describing the empirical model has been submitted to Climate Dynamics.

3. REFERENCES

Adler, R. F., G. J. Huffman, A. Chang, R. Ferraro, P. Xie, J. Janowiak, B. Rudolf, U. Schneider, S. Curtis, D. Bolvin, A. Gruber, J. Susskind, P. Arkin, and E. Nelkin, 2003: The version 2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979-present). *J. Hydrometeorol*, 4(6), 1147-1167. Land + Ocean

Ardilouze, C., Materia, S., Batté, L. et al., 2020. Precipitation response to extreme soil moisture conditions over the Mediterranean. *Clim Dyn*, <https://doi.org/10.1007/s00382-020-05519-5>

Barnston, A. G., and R. E. Livezey, 1987: Classification, Seasonality and Persistence of Low-Frequency Atmospheric Circulation Patterns. *Mon. Wea. Rev.*, 115, 1083–1126, [https://doi.org/10.1175/1520-0493\(1987\)115<1083:CSAPOL>2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115<1083:CSAPOL>2.0.CO;2).

Bauer, P., Thorpe, A. & Brunet, G., 2015. The quiet revolution of numerical weather prediction. *Nature* 525, 47–55. <https://doi.org/10.1038/nature14956>

Cohen J, Jones J. A new index for more accurate winter predictions, 2012. *Geophys Res Lett*, 38:L21701.

doi:10.1029/2011GL049626

Cornes, R., G. van der Schrier, E.J.M. van den Besselaar, and P.D. Jones. 2018: An Ensemble Version of the E-OBS Temperature and Precipitation Datasets, *J. Geophys. Res. Atmos.*, 123. doi:10.1029/2017JD028200

Copernicus Climate Change Service (C3S) (2017): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate . Copernicus Climate Change Service Climate Data Store (CDS), date of access.

<https://cds.climate.copernicus.eu/cdsapp#!/home>

Doblas-Reyes, F. J., García-Serrano, J., Lienert, F., Biescas, A. P., and Rodrigues, L. R., However, areas with high/low skill seem to be distributed following synoptic scale patterns too, 2013: Seasonal climate predictability and forecasting: status and prospects, *WIREs Climate Change*, 4, 245–268, <https://doi.org/10.1002/WCC.217>.

Douville, H., Peings, Y., and Saint-Martin, D., 2017. Snow-(N)AO relationship revisited over the whole twentieth century, *Geophys. Res. Lett.*, 44, 569– 577, doi:10.1002/2016GL071584.

Medscope (689029) Deliverable D2.3

Haenlein, Michael & Kaplan, Andreas, 2004. A Beginner's Guide to Partial Least Squares Analysis. Understanding Statistics. 3. 10.1207/s15328031us0304_4.

Hoskins, B., 2013: The potential for skill across the range of the seamless weather-climate prediction problem: a stimulus for our science. Q. J. R. Meteorol. Soc., 139:573–584.

Jacob A. Wegelin, 2000. A survey of Partial Least Squares (PLS) methods, with emphasis on the two-block case. Technical Report 371, Department of Statistics, University of Washington, Seattle.

Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., Tietsche, S., Decremmer, D., Weisheimer, A., Balsamo, G., Keeley, S. P. E., Mogensen, K., Zuo, H., and Monge-Sanz, B. M.: SEAS5: the new ECMWF seasonal forecast system, Geosci. Model Dev., 12, 1087–1117, <https://doi.org/10.5194/gmd-12-1087-2019>, 2019

Kim, H.M., Webster, P.J. & Curry, J.A., 2012. Seasonal prediction skill of ECMWF System 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere Winter. Clim Dyn 39, 2957–2973. <https://doi.org/10.1007/s00382-012-1364-6>

Levitus, S., J.I. Antonov, T.P. Boyer, R.A. Locarnini, H.E. Garcia, and A.V. Mishonov (2009). Global ocean heat content 1955-2008 in light of recently revealed instrumentation problems. Geophys. Res. Lett., 36, L07608, doi:10.1029/2008GL037155

Lorenz, E.N., 1969: The predictability of a flow which possesses many scales of motion. Tellus A, 21:289–307.

Mason, S.J. and O. Baddour, 2008: Statistical Modelling. in Troccoli, A. et al., (eds.) Seasonal Climate: Forecasting and Managing Risk. Dordrecht: Springer Netherlands (Nato Science Series: IV: Earth and Environmental Sciences). doi: 10.1007/978-1-4020-6992-5.

Robinson, David A.; Estilow, Thomas W.; and NOAA CDR Program (2012): NOAA Climate Data Record (CDR) of Northern Hemisphere (NH) Snow Cover Extent (SCE), Version 1. [indicate subset used]. NOAA National Centers for Environmental Information. doi:10.7289/V5N014G9

Rodríguez-Guisado, E., Serrano-de la Torre, A. Á., Sánchez-García, E., Domínguez-Alonso, M., and Rodríguez-Camino, E., 2019: Development of an empirical model for seasonal forecasting over the Mediterranean, *Adv. Sci. Res.*, 16, 191–199, <https://doi.org/10.5194/asr-16-191-2019>, 2019.

Sánchez-García, Eroteida & Voces-Aboy, José & Navascués, Beatriz & Rodríguez-Camino, Ernesto, 2019. Regionally improved seasonal forecast of precipitation through Best estimation of winter NAO. *Advances in Science and Research*. 16. 165-174. [10.5194/asr-16-165-2019](https://doi.org/10.5194/asr-16-165-2019).

Wegmann, M., Rohrer, M., Santolaria-Otín, M., and Lohmann, G, 2020.: Eurasian autumn snow link to winter North Atlantic Oscillation is strongest for Arctic warming periods, *Earth Syst. Dynam.*, 11, 509–524, <https://doi.org/10.5194/esd-11-509-2020>.

Weisheimer, A., Palmer, T. N., and Doblas-Reyes, F. J., 2011: Assessment of representations of model uncertainty in monthly and seasonal forecast ensembles, *Geophys. Res. Lett.*, 38, L16703, <https://doi.org/10.1029/2011GL048123>.

World Meteorological Organization, 2018: *Guidance on Verification of Operational Seasonal Climate Forecasts* (WMO-No. 1220). Geneva.

World Meteorological Organization, 2020: *Guidance on Operational Practices for Objective Seasonal Forecasting* (WMO-No. 1246). Geneva.