MEDSCOPE
MEDiterranean Services
Chain based On climate PrEdictions

ERA4CS Joint Call on Researching and Advancing Climate Services Development – Topic B

Deliverable D4.1

Assessment of the quality of sectoral prediction-based indicators

MEDSCOPE

MEDiterranean Services Chain based On climate PrEdictions

Deliverable Title	*Assessment of the quality of sectoral prediction-based indicators*		
Brief Description	*D4.1 is focused on the identification and evaluation of variables and indicators relevant for the Mediterranean region and specific for each sector. It summarizes for each prototype developed during the project: i) variables and indicators identified as relevant; ii) evaluation in real time and/or hindcast period of seasonal forecasts for the selected variables and indicators; iii) results from probabilistic and deterministic verification scores applied to forecasted variables and indicators.*		
WP number	WP4		
Lead Beneficiary	*Ernesto Rodriguez-Camino (AEMET)*		
Contributors	*Silvia Terzago (CNR), Andre Chancy (INRAE), Beatriz Navascues (AEMET), Nieves Garrido (AEMET), Jean-Michel Soubeyroux (Météo-France), Llorenç Lledó (BSC), Valentina Bacciu (CMCC)*		
Creation Date Version Number Version Date	*15/05/2021* *1* *04/06/2021*		
Deliverable Due Date Actual Delivery Date	*15/06/2021* *15/06/2021*		
Nature of the Deliverable	X	*R – Report*	
		P - Prototype	
		D - Demonstrator	
		O - Other	
Dissemination Level/ Audience	X	*PU - Public*	
		PP - Restricted to other programme participants, including the Commission services	
		RE - Restricted to a group specified by the consortium, including the Commission services	
		CO - Confidential, only for members of the consortium, including the Commission services	

Version	**Date**	**Modified by**	**Comments**
1	04/06/21	E. Rodriguez-Camino	Preliminary version – structure of the document
2	15/06/21	E. Rodríguez-Camino	Final version

JPI Climate ERA4CS European Research Area for Climate Services

MEDSCOPE
MEDiterranean Services
Chain based On climate PrEdictions

Outline

JPI Climate ERA4CS European Research Area for Climate Services

Executive summary

The MEDSCOPE WP4 has aimed at providing specific sectoral seasonal predictions for the Mediterranean region. Its main goal was to demonstrate the feasibility of climate services and generate prototypes for three sectors: i) renewable energy, ii) hydrology (including water resources and flood risk assessment) and iii) agriculture and forestry.

Deliverable D4.1 is focused on the identification and evaluation of variables and indicators relevant for the Mediterranean region and specific for each sector. This deliverable D4.1 summarizes for each prototype developed during the project: i) variables and indicators identified as relevant; ii) evaluation in real time and/or hindcast period of seasonal forecasts for the selected variables and indicators; iii) results from probabilistic and deterministic verification scores applied to forecasted variables and indicators. Details of probabilistic and deterministic scores for each prototype are collected in the Annexes. The main text only presents a brief summary of evaluation results.

Application or impact models are typically used to translate climate variables into users' defined variables and indicators. Users' variables and indicators cover a wide typology and depend strongly on the specific sector under consideration. The evaluation -when possible- has been conducted for the users' defined variables and indicators against observational data frequently provided by users. Most of the assessments in WP4 have been based on their performance on past conditions. This performance was generally evaluated making use of retrospective forecasts computed for each seasonal forecasting system and the suite of steps leading to forecasted sectoral indicators. Sectoral indicators are mostly expressed in the form of probabilistic forecasts which are finally verified following the standard procedures for seasonal forecast verification including cost/benefit scores.

Specific issues, such as, the usage of verifying (actual *versus* synthetic) observations, the comparison against a reference forecast (usually based on climatology), the important role of memory (in agriculture/forestry prototypes), etc. are also discussed in the document and annexes.

1.- Introduction

Skillful climate forecasts on seasonal-to-decadal timescales can have large socio-economic benefits. Seasonal predictions have therefore been performed operationally at centers around the world for more than 40 years. The skill of these predictions derives primarily from tropical phenomena such as the El Niño Southern Oscillation (ENSO), whereas predictability and forecast quality is currently relatively lower over Europe and the Mediterranean region (e.g. Weisheimer et al., 2011; Doblas-Reyes et al. 2013). There is also growing evidence that skillful climate prediction on multiannual (decadal) timescales may be possible, but while recent studies provide encouraging results, much work remains to be done to understand the potential predictability on these longer timescales (Meehl et al., 2014). In general, our limited understanding of the mechanisms and processes responsible for predictability and model systematic errors limit our capability to simulate and predict seasonal-to-decadal climate variability, especially over the Euro–Mediterranean area. Improved global climate model calibration and regionalization techniques as well as better forecast verification methods need to be developed for this region to extract as much climate information as possible from existing forecast systems. This is the primary and underpinning goal of the MEDSCOPE project, from which other user-oriented goals are derived and enabled, such as authoritative tailored climate information, tools for the creation of such information for a wider audience, and examples of the application of the tools in selected sectoral case studies of service development and demonstration of value.

The MEDSCOPE project had as main objective to enhance the exploitation of climate predictions from seasonal to decadal timescales, maximizing the potential of their application in different economic sectors, public and private, of relevance for the Mediterranean region, here defined as the domain encompassing the Mediterranean basin and the surrounding areas, including North Africa and the Middle East. However the MEDSCOPE project has mainly focused on the seasonal timescale as a wealth of forecasts (including retrospective forecasts) is already available and, in general, the state-of-the-art of both scientific knowledge and applications are more mature for this time scale.

The overall outcome of MEDSCOPE has been the development of a set of tools and methods aimed at improving the production of climate services based on climate forecasts, enhancing the capability of public and private users and stakeholders to develop and implement strategies of adaptation to climate variability and climate change. Such tools are essentially based on further development and refining of existing techniques and procedures, such as model output calibration (to tackle problems of systematic model errors), regionalization (to enhance model resolution and therefore capability of representing small spatial scales) and mixed statistical-dynamical post-processing techniques (to address problems of specific sectoral applications intractable by standard prediction methods).

The main specific objectives of MEDSCOPE were:

- Improvement of our comprehension of the mechanisms driving the climate variability in the Mediterranean region, and especially 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 (seasonal-to-decadal) (WP2).
- Provision of a set of generalized methods and ready-to-use tools for forecast verification and comprehensive skill assessments – including those for user-oriented applications – for downscaling, calibration and bias adjustment of the forecasts, and develop methodologies of optimal forecast combination to provide a single source of information. The developed tools have been compiled in a toolbox that is freely available and addressed to practical users from both public and private sectors (WP3).
- Provision of prototypes of climate services products based on end-user tailored climate forecasts at seasonal and multi-annual timescales, in relevant economic sectors for the Mediterranean, such as wind energy, water management (hydrology), and agriculture and forestry (and fire risk) (WP4).

Focusing on this last objective, The MEDSCOPE project WP4 has provided forecasts of variables and indicators relevant to the considered sectors by applying the tools and the methods (downscaling, calibration, bias correction, forecast combination) developed in the project itself. During the MEDSCOPE project lifetime, users have played an important role distinctly for each particular prototype. Depending on prototypes, users have had an active role in the definition and/or development of indicators, tools, evaluation methods or presentation of the final products. The strong involvement of users (including MedCOF) have ensured that the developed prototypes meet the community needs, building upon the partners' experience on social science applied to climate services.

The WP4 has aimed at providing specific sectoral seasonal predictions for the Mediterranean region. The goal was to demonstrate the feasibility of climate services and generate prototypes for three sectors: i) renewable energy, ii) hydrology (including water resources and flood risk assessment) and iii) agriculture and forestry. Climate services design has also benefited from previous projects results such as EUPORIAS, MOSES, CLIMRUN, etc. Identification and evaluation of indicators relevant for the Mediterranean region and specific to sectors have been the main focus of WP4. The

WP4 was fed by WP2 on sources of predictability and by WP3 on correcting, combining and synthesizing different sources of climate information.

The main objectives of WP4 were:

- identification of specific indicators known to affect wind energy and hydropower, hydrological applications (e.g. inflow to water dam, river discharge, soil moisture, agriculture (yield, irrigation needs, pest risk) and forest (vulnerability to water stress, fire risk);
- generation of seasonal forecasts of the different sectoral indicators evaluation of seasonal capabilities making use of a set of indicators selected for the Mediterranean region;
- development and assessment of sectoral indicators in decadal time scale simulations, when relevant, and evaluation in selected case studies;
- delivery of an effective and sound communication of the main project results towards identified target groups of relevant stakeholders, particularly through MedCOF.

This deliverable D4.1 summarizes, for each prototype developed during the project and having completed a comprehensive evaluation: i) variables and indicators identified as relevant for each prototype; ii) evaluation in real time and/or hindcast period of seasonal forecasts for the selected variables and indicators; iii) results from probabilistic and deterministic verification scores applied to forecasted variables and indicators. Details of probabilistic and deterministic scores for each prototype are collected in the Annexes. The main text only presents a brief summary of evaluation results.

2.- Climate services based on seasonal predictions for the Mediterranean

The climate services developed and evaluated in WP4 are grouped into three categories corresponding to three relevant economic sectors for the Mediterranean: renewable energy, water management (hydrology), and agriculture and forestry (and fire risk).

Three prototypes were developed within the MEDSCOPE project for the renewable energy sector: (1) capacity factor of wind power to assess the relative performance of any generating power plant -with particular focus on extreme events relevant for the industry- evaluated in Spanish sites (Lledo et al. 2019); (2) hydropower potential over specific pilot areas in the Alps and Pyrenees in summer, making use of the amount of accumulated snow mountain in late spring (Viel et al. 2016; Habets et al.

2008); (3) hydropower potential making use of an upgraded version of the S-ClimWaRe system (Voces et al. 2019) evaluated in a selected pilot hydropower dam in Spain.

Hydrology and water management sector was covered with the following four prototypes: (1) extension of RIFF climate service (Viel et al. 2016)) developed in the EUPORIAS project using SURFEX/TRIP model for land surface and river routing, with atmospheric forcing inputs from ensemble seasonal forecasts and evaluation in three selected river basins (Durance, Ebro and Po); (2) extension of S-ClimWaRe service developed in the EUPORIAS project to generate seasonal forecasts of water inflow to reservoirs in Spain (Sanchez et al. 2021); (3) service for estimating hydro-power and irrigation water availability using SCHEME evaluated over a river basin in Greece (Roulin and Vannitsem, 2015); (4) service to forecast snowpack conditions and evolution and the response of selected glaciers evaluated in the Alpine region (Bartelt and Lehning 2002, Oerlemans et al. 2011)

The following services were developed within the project for the agriculture and forestry sector: (1) estimation of winter cereal yield using the Aquacrop crop growth model forced with seasonal forecast evaluated over a region in Spain (Garrido et al. 2021); (2) selection of agriculture, forestry and forest fire risk seasonal forecasted indicators (Costa Saura et al., 2020) for the whole Mediterranean region; (3) estimation of seasonal soil wetness forecasts; (4) estimation of water requirement for irrigation; (5) estimation of a set of impact indicators describing the influence of climatic factors on plant growth, development and vulnerability, and computation of seasonal forecasts of these indicators, taking into account the current vegetation status. This last service includes the following four groups of indicators: (i) agro/eco-climatic metrics (Caubel et al., 2015); (ii) soil water balance, crop phenology and yield production using Crop and Soil Vegetation Atmosphere Transfer models to identify potential yields and water shortage risk; (iii) forest productivity and mortality indicators derived from eco-physiological modelling (Dufrêne et al., 2005); (iv) forest fire risk using both consolidated and new indicators that account for added fire drivers (fuel moisture, biomass, e.g. Ruffault et al., 2012). Two types of outcomes were produced for each of the four considered groups: one group (i) over the whole Mediterranean region on a ~25 km resolution grid. For the other groups (ii - iv), the indicators were obtained from biophysical modelling at high-resolution (~5 km or less) for a few selected pilot areas (in Italy, France and Spain), where the necessary additional local information regarding soil-plant initialization and parameterization can be obtained.

3.- Steps in sectoral climate services based on seasonal predictions

The transformation of climate data into products that can be easily integrated into decision-making can be described through a basic climate services development chain based on the GFCS components (Lucio and Grasso, 2016). The climate variability at seasonal timescales is particularly one of the factors playing an important role in many climate-sensitive sectors (Doblas-Reyes et al. 2013b). One straightforward method to estimate future climate variability is to assume the same variability derived from the past conditions or climatology (Goddard et al. 2010). As at seasonal time scale future climate is not simply a reproduction of climatology, this approach could lead to incorrect decisions. In these cases, the use of seasonal forecasts could be more helpful by providing a probabilistic estimate of how essential climate variables such as temperature or precipitation may develop in the coming months and seasons and, thus, can help to inform, focus and improve decision making (Soares and Dessai 2016; Torralba 2019). However, the large amount of probabilistic information arising from seasonal forecast systems, which is usually un-tailored and difficult to understand by the non-expert public, makes the integration of these forecasts into decision-making processes. Hence, the reduction of society's vulnerability to seasonal climate related-risks requires the contextualisation and interpretation of the seasonal forecast data, as well as the development of tailored products and tools based on this source of climate information (Hewitt et al. 2013).

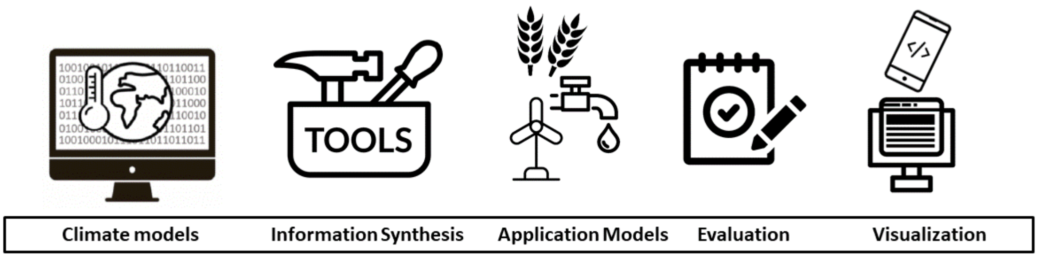


Figure 3.1. – Diagram with a simplified climate services chain based on seasonal forecasts

Most sectoral seasonal prediction-based climate services share a common suite of steps starting from global seasonal model outputs and ending up with probabilistic forecasts for user's defined indicators. The scheme in Figure 3.1 shows the main steps in a typical sectoral prediction based suite. The Copernicus Climate Change Service (C3S) Climate Data Store (CDS) is the most used source of seasonal forecasts in the MEDSCOPE project and many prototypes described and evaluated in this deliverable make extensive use of such data. This clearly means that the final quality and skill of sectoral prediction-based indicators will have a strong dependency on the skill of the selected seasonal forecasting system(s) (SFSs). Once selected the SFS data the next step consist in the application of a set of tools and techniques aiming at synthetizing and correcting seasonal forecasts, such as model output calibration (to tackle problems of systematic model errors), regionalization (to enhance model resolution and therefore capability of representing small spatial scales), selection, combination and weighting of ensemble members (to deal with the different quality of systems and ensemble members) and mixed statistical-dynamical post-processing techniques (to address problems of specific sectoral applications intractable by standard prediction methods). All such tools are collected in a toolbox developed in WP3 that is publicly available (https://cran.r-project.org/web/packages/CSTools/). The next step is the use of application or impact models translating climate variables (e.g., precipitation, temperature, wind) into users' defined indicators (e.g., crop yield, river discharge, dam inflow, wind energy capacity factor, etc.). Finally, evaluation is conducted making use of objective verification skill scores also included in the developed toolbox. This evaluation is typically carried out for a number of seasonal reforecasts (between 20 and 30 years) of users' indices computed from hindcasts provided by different SFSs.

4.- Sectoral indicators

As explained in the previous section, application or impact models are typically used to translate climate variables into users' defined indicators. Ideally evaluation should be conducted for the users' defined indicators against observational data frequently provided by users.

Users' indicators cover a wide typology and depend strongly on the specific sector under consideration. The following specific indicators -among others- have been used and evaluated in different prototypes (see Annexes for further description):

i) For the renewable energy sector:
   a. The **capacity factor (CF)** is a widespread performance indicator in the whole wind energy sector that allows fair comparisons between power plants of different sizes and types.
   b. **Inflow to reservoirs** is commonly used in the hydroelectric sector
   c. **River discharge** is also used in the hydroelectric sector

ii) For the water management (hydrology) sector:
   a. Monthly average **snow depth and snow water equivalent**
   b. **Inflow** to reservoirs
   c. River discharge
   d. Soil Wetness Index

iii) For the agriculture and forestry sector:
   a. T/Ha yield
   b. Soil Wetness Index

c. **Canadian Fire Weather Index**
d. **Potential Evapotranspiration**
e. **Potential Soil Moisture Deficit**

5.- Methodology for evaluation of the quality of sectoral prediction-based indicators

Sectoral prediction-based indicators are the basis for decision making on sectors affected by climate variability on seasonal time scales. Most of the assessments of prototypes developed in WP4 have been based on their performance on past conditions. This performance has been evaluated making use of the available retrospective forecasts computed for each seasonal forecasting system. Starting from retrospective forecasts the suite of steps leading to forecasted sectoral indicators is conducted. In most cases sectoral indicators are expressed in the form of probabilistic forecasts which are finally verified following the standard procedures for seasonal forecast verification (WMO 2018).

5.1.- Verifying observational data

One important issue common to all prototypes is the **selection of observational data to verify and evaluate the forecasted users variables or indicators**. Frequently, such data are not accessible or do not allow an easy comparison with forecasted indicators. Even the frequent lack of standardization of some users' variables is a serious hindrance for the direct comparison of forecasts and observations. In those cases some synthetic data mimicking observations can be generated for evaluation purposes. Of course**, if synthetic data are used for verification purposes, some estimation of their quality** should be conducted previously to their use.

This procedure has been followed by the prototype providing seasonal forecasts of cereal yield in Castilla y León (Spain). In this case, the available real cereal yield data was not directly comparable with seasonal forecast simulations –based on downscaled ECMWF S5 output followed by the Aquacrop growth model- as there exists different sowing and cultivation practices among farmers and also from those assumed by Aquacrop. Also, not all the entire surface of Castilla y León is cultivated with winter cereal. Therefore, an *a posteriori* Aquacrop run driven by meteorological observations (reanalysis) is routinely performed to obtain the reference cereal yield mimicking actual observations for evaluation purposes (see Annex 7).

The same procedure was applied by the prototype producing seasonal forecasts over Europe of wind capacity factor. As in the previous case, the combination of an optimal climatic forcing (re-analysis), a set of tools synthetizing available information and finally a theoretically perfect application model translating in this case winds into capacity factor, allows the generation of a synthetic observational data to compare against the probabilistic seasonal forecasts (see Annex 1).

Meteo-France has also developed two hydrological suites (called SIM2 and MSC) driven by SAFRAN and MESCAN analysis data, respectively, serving as synthetic observational data to compare the seasonal forecasts of hydrological variables (see Annexes 4, 5 and 7). The first suite, SIM2, is based on the hydrological seasonal application developed in the frame of the EUPORIAS project (Viel et al, 2016). The second suite, MSC, -developed during the MEDSCOPE project is based on the coupled hydrological model SURFEX-CTRIP (Decharme et al., 2019). SURFEX runs at a resolution of 0.05° and the routing model CTRIP at 0.5°. It covers the western part of the Mediterranean basin and uses the MESCAN analysis (Bazile et al., 2017) as an atmospheric reference. Results from both suites have been extensively compared and both against observations (Dayon 2019).

Fondazione CMCC used 6-hour retrospective seasonal forecasts (CMCC SPSv3 and ECMWF system5) and ERA5 climate dataset for hindcast verification for the period 1993-2015. The ERA5 climate dataset, referred to also as "observations" hereafter, was scaled to 1° resolution to compare against CMCC SPSv3 and ECMWF system5 predictions.

5.2.- Comparison against a reference forecast

The typical reference forecast is the one based on climatology. In fact, many applications for different sectors have so far made their seasonal estimates assuming that future evolution would have a climatological behavior. Therefore, any climate service based on seasonal forecasts would have to be compared with climatology based forecasts.

Climatological forecasts (up to 6 months) have been run for all initialization months over the hindcast period (1993-2016) for the prototypes on seasonal forecasts of Durance, Ebro and Po river discharge (see Annexes 4, 5, 6). Climatological forecasts were compared with forecasts from the Meteo-France System 6 on the period 1993-2016. The evaluation is done by comparing the forecasted streamflow with climatological forecast and with pseudo-observations coming from the MESCAN-SURFEX-CTRIP reanalysis.

This comparison was also conducted e.g. by the prototype providing seasonal forecasts of cereal yield in Castilla y León (Spain) (see Annex 7). In this case, the Aquacrop growth model was driven in two different experiments with seasonal forecasts based on the ensemble from the ECMWF SFS S5 and on an ensemble –called CLIM- reproducing the climatology. CLIM experiment makes use -for the forecasting part of the yield simulation- of a probabilistic SFs based on a 24 members ensemble being each member obtained from the meteorological forcing corresponding to each year of the reference experiment over 1995-2018. In this way a probabilistic SF is generated representing the climatology, as each member corresponds to actual observations in the 24 years period.

5.3.- Objective deterministic and probabilistic verification scores

One fundamental component of a climate service based on seasonal forecasts is the quality assessment of real-time forecasts. As important as producing forecasts is, the task of evaluating how good the real-time forecasts have been is of equal importance. Verification assessment of real-time forecasts should be performed in addition to establishing the average predictive skill of the seasonal forecasting system (based on the accompanying hindcasts). Verification should follow the recommended procedures for the verification of seasonal forecasts based on the WMO Guidance on Verification of Operational Seasonal Climate Forecasts (WMO 2018). The procedures have been selected to inform end-users of the forecasts. While the focus of the recommended procedures is on how well the forecasts correspond with the observations (forecast quality), care is taken to measure those attributes that can make forecasts potentially useful (forecast value). The WMO Guidance emphasizes the importance of using various complementary skill measures for a complete examination of outlook forecast quality. The most fundamental attributes recommended to be assessed are **discrimination, reliability, resolution, sharpness and skill.**

Ideally the **multiple attributes of forecast quality should be measured individually**, but some commonly used procedures measure more than one attribute at once. These procedures can lead to results that are difficult to interpret, and may lead to misleading conclusions. Alternative procedures that measure individual attributes are suggested in preference because of their simpler interpretation and more informative results. The **most important attributes are resolution or discrimination**. Resolution measures whether the outcome differs given different forecasts, while discrimination measures whether the forecasts differ given different outcomes. As long as there is some resolution or discrimination the forecasts contain potentially useful information, regardless of how poor the reliability is. It is generally easier to measure discrimination than it is to measure resolution because discrimination can be measured more accurately with smaller samples than can resolution. The generalized discrimination score is therefore recommended as an initial score for assessing forecast quality. This score measures the ability of the forecasts to discriminate between the wetter, or warmer, of two observations. Since forecast quality can be conditional upon the outcome, it is also recommended that the score be calculated for the individual categories. The score then generalizes to the **area beneath the relative operating characteristics (ROC) curve**. Construction of the ROC curve is recommended for more detailed diagnostics of the ability of the forecasts to discriminate observations in each category.

The **reliability** (do the forecast probabilities give an accurate indication of the uncertainty in the outcome?) is unquestionably an important attribute. Unfortunately, measuring reliability requires large samples, and so it is only viable to measure it by pooling the forecasts from different locations. **For scores at individual locations it is recommended that summary scores** (those that measure more than one attribute) be used instead. For detailed diagnostics of forecast quality, **reliability diagrams** are recommended; these diagrams measure **reliability and resolution**.

The most frequent **deterministic skill score** is the **correlation** between the predicted and the observed mean value of anomalies over the different land domains.

The following **probabilistic skill scores** have been also computed for several prototypes either for the customary climatic variables (temperature and precipitation) or for user's defined variables (indicators): Ranked Probability Skill Score (RPSS) for terciles, and Relative Operating Characteristic (ROC) area and Brier Skill Score (BSS) for two events (upper/lower tercile). A complete definition of these scores can be found in Wilks (2006).

The **Ranked Probability Skill Score (RPSS)** is a generalization of Ranked Probability Score (RPS) based on a reference forecasting system. The RPS averages squared "error" in the cumulative probabilistic forecasts. Positive values of RPSS indicate more skill than the reference system, usually the climatology. The **Continuous Ranked Probability Skill Score (CRPSS)** is the continuous analog of the RPSS.

ROC curves measure discrimination and skill. If the category of interest is above-normal, the score based on the ROC area indicates the probability of successfully discriminating above-normal observations from normal and below-normal observations. The ROC area ranges from 0% to 100%, with a score of 50% representing no skill, 100% indicating perfect discrimination, and 0% indicating perfectly bad discrimination. It is important to stress that ROC curves are measuring only the discrimination ability between two possible results, but it is not informative about reliability since it is not sensitive to bias.

The **Brier Score (BS)** is the most common verification method for probabilistic forecasts, as it has a mathematical structure similar to the **Mean Square Error (MSE).** BS measures the difference between the forecast probability of an event (p) and its occurrence (o), expressed as 0 or 1, depending on whether the event has occurred or not. As with RMSE, the BS is negatively orientated, i.e. the lower, the "better". The **Brier Skill Score (BSS)** is conventionally defined as the relative probability score compared with the probability score of a reference forecast.

The **Gerrity Skill Score (GSS**, Gerrity Jr (1992); Gandin and Murphy (1992); Materia et al., (2020)) also assesses the quality of a given index provided an additional choice on how high the forecast probability should be to trigger some alarm or to make some decision. This threshold for the forecast probability allows us to treat probabilistic forecasts as if they were deterministic and the GSS is then computed making use of a contingency table.

5.4.- Cost/benefit verification scores

A forecast has "value" if it can be used to help realize some benefit, whether economic, social, or otherwise. **Forecast quality is a prerequisite but not a guarantee of forecast value**: forecasts that have good quality have the potential for being of value, but whether they actually are depends on the impacts of the observed climate, and on the options available for mitigating (or taking advantage of) such impacts (Katz and Murphy, 1997). There is usually an imbalance between the losses and gains realized from the use of forecasts. One may consider, for example, a set of excellent forecasts that are released in a timely manner, and which inform profitable decision-making when the forecasts correspond well with the observed outcomes. It is still possible for these forecasts to have no value if the costs incurred from the occasional "bad" forecast more than offset the benefits from the frequent "good" ones. Very good forecasts can therefore have no, or even negative, value. Conversely, forecasts with low quality can be of immense value if the occasional "good" forecast can be used to good effect. Unfortunately the needed socioeconomic data at the spatial, temporal, user resolution to relate to forecast are frequently not accessible for conducting a verification based on scores measuring the costs/benefits associated with wrong/right forecast.

Anyway, from a user perspective seasonal forecasts are frequently an important element in a decision making process and the estimation of the forecast value using scores taking into account cost and benefits/losses of seasonal forecasts, either deterministic or probabilistic. For a review on this important issue see (Richardson 2003).

A cost/benefit verification has been also applied in the evaluation of some prototypes as, e.g., the prototype providing seasonal forecasts of cereal yield in Castilla y León (Spain) (see Annex 7) and the cases studied in the AgriMet Info Forecast (see Annex 10). The method starts with the definition of the event(s) to be forecasted, e.g., cereal yield below the lower tercile (a bad harvest), and above the upper tercile (a good harvest). Then, for each event, probabilistic seasonal forecasts are transformed into deterministic forecasts of the occurrence or not of the event. The threshold value of probability that triggers the occurrence of the event is determined in a previous step by optimizing the Gerrity equitable skill score value over the hindcast period (Gandin and Murphy, 1992, and Gerrity, 1992, Materia et al., 2020). Once deterministic forecasts of the event have been obtained for each year in the hindcast period, a contingency table for yes/no vs observed/forecasted cases is built. The hit rate H and the false alarm rate F are then obtained. The last step of this assessment consists of the application of a simple Cost/Loss economic model (Richardson, 2000) to obtain the Relative Economic Value of the forecasts (V) (see Annexes 7 and 10 for more details on this calculation).

6.- Assessment of indicators for prototypes on renewable energy

6.1.- Seasonal predictions of wind capacity factor for Spain in March 2018 (BSC)

The skill scores for surface wind and the three Capacity Factor (CF) indicators for that region can be found in Table 6.1. Looking at wind speed, we see that CRPSS (measuring the forecast distribution quality) is negative for all lead times. However RPSS (measuring the tercile-based forecast quality) is 5% for the 1-month lead time. Albeit being a modest value, it indicates some signal (also seen in ensemble mean correlation) and the possibility to use this forecast with caution. The quality of the extreme probabilities, evaluated with the BSS, is negative for both P10 and P90, indicating that those predictions are not better than climatology in general (see Annex 1 for more details)

	sfcWind			CF IEC I			CF IEC II			CF IEC III		
	Dec	**Jan**	**Feb**	**Dec**	**Jan**	**Feb**	**Dec**	**Jan**	**Feb**	**Dec**	**Jan**	**Feb**
RPSS	−0.06	−0.03	0.05	−0.16	−0.09	−0.01	−0.16	−0.11	−0.06	−0.18	−0.14	−0.07
BS P10	−0.24	−0.06	−0.05	−0.1	−0.14	−0.04	−0.1	−0.09	−0.08	−0.14	−0.04	−0.09
BS P90	−0.03	−0.03	−0.16	−0.18	−0.18	−0.09	−0.18	−0.18	−0.09	−0.13	−0.15	−0.11
CRPSS	−0.17	−0.19	−0.08	−0.15	−0.13	−0.04	−0.15	−0.14	−0.05	−0.13	−0.14	−0.06
EnsCorr	−0.05	−0.45	0.35	−0.36	−0.24	0.16	−0.35	−0.26	0.12	−0.35	−0.28	0.04

Table 6.1.- Skill scores for the surface wind and capacity factor forecasts estimated from a hindcast covering the 1993-2015 period. Forecasts correspond to an area of the north-west of Spain (9W-7W and 42N-44N) for the month of March of 2018 for three turbine types (classes IEC I, II and III) issued up to three months in advance.

Although the surface wind forecasts for March 2018 were quite successful in anticipating a windy month, **the evaluation based on a longer hindcast shows that the potential for using forecasts in this region is modest**. The hindcast-based verification is the only quality measure that was available at the moment of issuing those forecasts, therefore the ability of decision making was limited. The capacity factor forecasts did not perform well in this case.

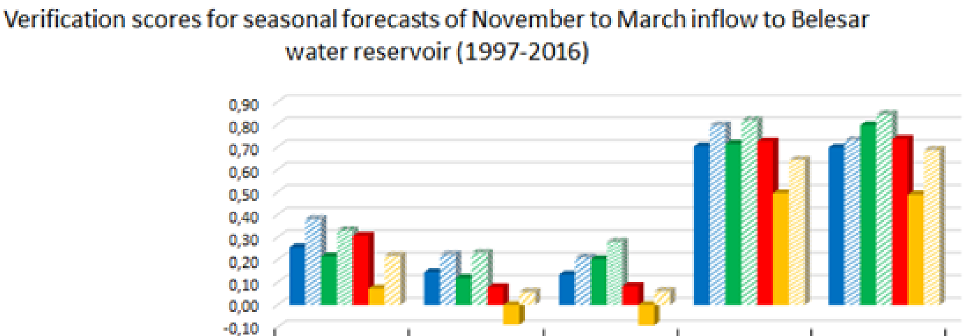
The particular forecasts shown in this case study have also been discussed with the stakeholders in one-to-one conversations. An interesting question that was raised was if it is possible that the predictions that show more evident signals of a climate forcing (i.e. differing substantially from the climatology) could have better predictability than the predictability estimated employing the whole hindcast. More research needs to be done regarding this question.

7.- Assessment of indicators for prototypes on hydrological products

7.1.- Seasonal forecasts of winter inflow for Belesar Water Reservoir (AEMET)

Objective evaluation has consisted in obtaining different standard verification scores of deterministic (ensemble mean) and probabilistic forecasts (lower and upper tercile) over the set of seasonal forecasts between 1997 and 2016 hindcast period. Water inflow observations have been used for this purpose (see Annex 2 for more details).

Figure 7.1 shows the verification scores obtained by the different experiments over the hindcast period 1997-2016. Correlation coefficient is selected to represent the accuracy of the deterministic forecast based on the ensemble mean. Brier skill score (BSS) and ROC area are the scores used to measure the probabilistic forecast skill. It can be seen that both **SIMPA and SURFEX experiments driven by downscaled ECMWF-System 5 seasonal forecasts with ensemble members weighting give the best results and improve the currently used S-ClimWaRe empirical system, both for deterministic and probabilistic forecasts for lower and upper tercile**.



	r	BSS_lower	BSS_upper	ROCarea_lower	ROCarea_upper
SIMPA (equiprobable members)	0,26	0,15	0,13	0,70	0,70
SIMPA(members weighted with Best NAO)	0,38	0,22	0,21	0,80	0,73
SURFEX (equiprobable members)	0,21	0,12	0,20	0,71	0,80
SURFEX (members weighted with Best NAO)	0,33	0,23	0,28	0,82	0,85
S-ClimWaRe	0,31	0,08	0,08	0,73	0,74
ECMWFS5 runoff (equiprobable members)	0,07	-0,09	-0,09	0,49	0,49
ECMWFS5 runoff (members weighted with Best NAO)	0,22	0,06	0,06	0,64	0,69

Figure 7.1.- Verification scores of seasonal forecasted water inflow to Belesar reservoir for the different experiments.

7.2.- S-ClimWare (Seasonal Climate predictions in support of Water reservoir management) web viewer (AEMET)

The S-ClimWaRe web viewer provides: i) diagnostics of the hydrological risk linked to climate variability and ii) probabilistic forecasts of meteorological and hydrological parameters in support of decision making by water reservoir managers (http://www.aemet.es/es/serviciosclimaticos/apoyo_gestion_embalses). The new upgraded S-ClimWaRe viewer developed in the MEDSCOPE project toolbox includes seasonal forecasts coming from an empirical forecasting system also called S-ClimWaRe (Voces et al., 2016 and 2019) and from ECMWF System-5 and makes use of two CSTools (for downscaling and for ensemble members weighting (Best NAO estimate method)).

An objective verification against observations has been carried out to assess the skill of the new forecasts introduced. Forecasts for all variables have been obtained for a 20 years hindcast period (1997-2016). The benefits of using downscaling and members weighting have also been shown (see Annex 3 for more details).

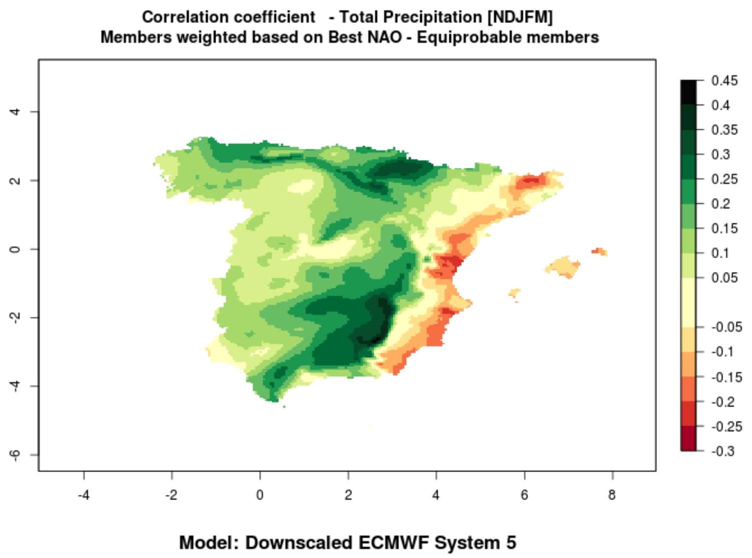


Figure 7.2.- Difference in correlation coefficient for downscaled ECMWF System-5 precipitation forecasts between experiments with weighted ensemble members (based on Best NAO) and equiprobable members. Green/brown shaded areas indicate an improvement/deterioration due to the Best NAO weighting with respect to equiprobable members.

As an example of this comprehensive verification Figure 7.2 shows the difference in correlation coefficient (between Best NAO weighting and equiprobable members) of downscaled precipitation forecast ensemble mean. The forecasted precipitation with Best NAO weighting is much better correlated with observations in most areas except a narrow band along the Mediterranean coast.

7.3.- Seasonal forecasts of Durance river discharge (Météo France)

For each indicator, scores have been calculated over the hindcast period (1993-2016) for each initialisation month at monthly time step using SIM2 as reference (see Annex 4 for more details and also performance of the MSC suite).

For river flow, the **seasonal forecast skill is quite homogeneous for the four scores** (Figure 7.3) **with a better predictability in spring and a lower one in summer**. In June, the month of interest for our prototype, the predictability is quite good, until lead time 4 (September)

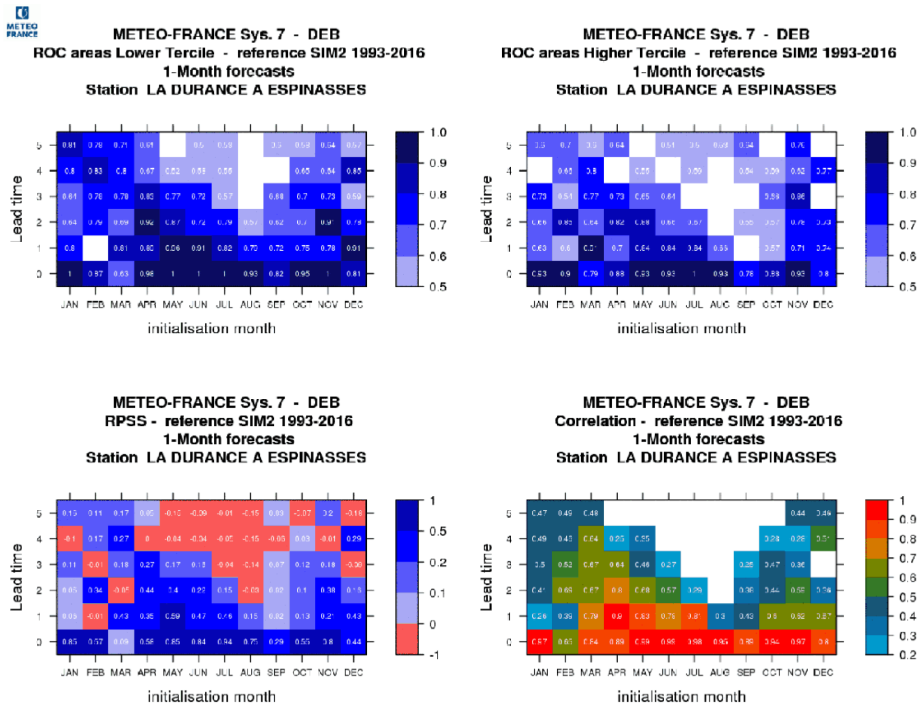


Figure 7.3.- Annual synthesis of seasonal forecast scores for SIM2 discharges at Espinasses station forced by MF Sys7. Upper ROC scores for lower and higher terciles, lower RPSS and correlation. Scores are presented by month of initialization.

Evaluation of snow water equivalent and soil water index is also available in Annex 4

7.4.- Seasonal forecasts of Ebro river discharge (Météo France)

The evaluation of the prototype has been carried out over the hindcast period 1993-2016. Hindcast from the Météo-France System 6 on the period 1993-2016 has been downscaled with ADAMONT for all initialization months from January to June. Results have been used to run the SURFEX-CTRIP chain on the hindcast period. We focus on **monthly river discharges at the outlet of the Ebro basin**. The evaluation is done by comparing **hindcast streamflow with climatological forecast and pseudo-observations** coming from the MESCAN-SURFEX-CTRIP reanalysis (see Annex 5 for more details).

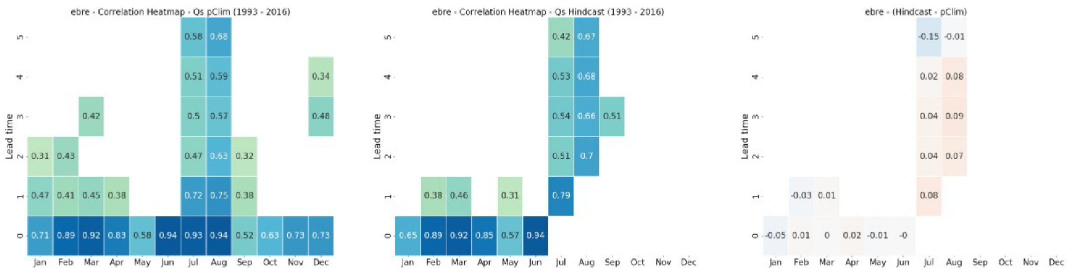


Figure 7.4.- Left: Correlation heatmap for the median of the hindcast, calculation at the outlet of Ebro river over the hindcast period. Values below 0.3 are masked. Each column represents a single validity month forecasted from different lead times (line) Right: Differences between the heatmap on the left and the same for the climatological forecast.

With the hindcast, deterministic scores (correlation calculated for the median of the ensemble, see Figure 7.4) obtained are very similar to the correlation obtained with the climatological forecast. The best performances are generally for the lead time 0, only available from January to June. We obtained **significant results for later lead time in July and August** validity months. Results are not significant from lead time 2 and above except in July and August at the outlet of the Ebro.

Correlations above the lead time 2 are generally limited at the outlet. However, a correlation map in June for the lead time 3 shows that correlation can reach 0.7 in some tributaries of the Ebro. It is mostly the case for tributaries coming from the Pyrenees. It is somehow expected as the snow melt from the Pyrenees is an important source of predictability. This point can be of interest for end users for a potential operational use.

7.5.- Seasonal forecasts of Po river discharge (Météo France)

The evaluation of the prototype has been carried out over the hindcast period 1993-2016. As three major steps had to be passed before running our prototype we first give information about MESCAN reanalysis over Po basin, streamflow modelling over Po basin and downscaling/bias correction using ADAMONT. Finally, the first component of our prototype (climatological forecasts) is evaluated (see Annex 6 for more details)

Hindcast from the Meteo-France System 6 on the period 1993-2016 has been downscaled with ADAMONT for all initialization months from January to June. Those initialization months were selected because they enable us to focus on results on the low flow period. Results have been used to run the SURFEX-CTRIP chain on the hindcast period. We focus on monthly river discharges at the outlet of the Po basin. The evaluation is done by comparing hindcast streamflow with climatological forecast and pseudo-observations coming from the MESCAN-SURFEX-CTRIP reanalysis.

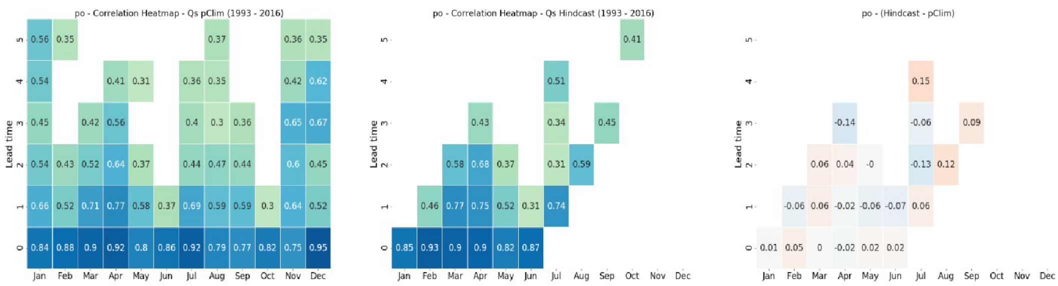


Figure 7.5.- Left: Correlation heatmap for the median of the hindcast, calculation at the outlet of Po river over the hindcast period. Values below 0.3 are masked. Each column represents a single validity month forecasted from different lead times (line) Right: Differences between the heatmap on the left and the same for the climatological forecast.

With the hindcast, **deterministic scores** (correlation calculated for the median of the ensemble) obtained are **very similar to the correlation obtained with the climatological forecast.**

Even if streamflow modelling within SURFEX-CTRIP should be improved to be better adapted to the Po watershed, the first results obtained for climatological forecasts over the Po river are quite satisfying for all months except June and October. Results on AUC show slight differences on lower and higher terciles predictability suggesting more potential in forecasting low flows in summer which could be of great interest for end-users.

7.6.- Seasonal forecasts of mountain snow depth and glacier evolution for water management (CNR)

CNR developed two prototypes based on snowpack and glacier models forced by the meteorological variables provided by ECMWF-S5 and MF-S6 seasonal forecast systems. The first prototype is based on the SNOWPACK model (Bartelt and Lehning 2002), which is run every November 1st to simulate the evolution of snow depth and snow water equivalent over the 7 months ahead, up to May 31st of the following year. The second prototype is based on a simple glacier model (Oerlemans, 2001) and it is run at the beginning of May to simulate the variation of glacier length and mass over the summer season ahead. Snow and glacier model forecasts were run with 4 different experiments driven by precipitation forcings with an increasing degree of accuracy and were all evaluated over the hindcast period 1993-2016 using both deterministic and probabilistic metrics (see Annex 9 for more details).

The agreement between snow depth tercile-based forecasts and observations has been evaluated at seasonal and monthly time scales by means of probabilistic skill scores with respect to a simple forecast based on climatology. Table 3 shows an example of probabilistic scores for snow depth forecasts (more scores are available in Annex 9).

		BSS Gastaldi								AUCSS Gastaldi							
		ECMWFS5				MFS6				ECMWFS5				MFS6			
		RAW	QM	RF	QM+RF	RAW	QM	RF	QM+RF	RAW	QM	RF	QM+RF	RAW	QM	RF	QM+RF
Lower Tercile	DJF	0.46	0.36	0.42	0.36	0.14	0.15	0.10	0.16	0.71	0.63	0.73	0.65	0.29	0.25	0.14	0.23
	MAM	0.21	0.19	0.18	0.18	0.10	0.00	0.04	0.07	0.27	0.30	0.27	0.33	-0.03	-0.24	-0.19	-0.05
Mid Tercile	DJF	0.13	0.09	0.17	0.13	-0.03	-0.18	0.11	0.00	0.02	-0.05	0.14	0.10	0.08	-0.26	0.54	0.17
	MAM	-0.03	0.06	0.10	0.10	0.08	-0.11	0.02	0.06	-0.34	-0.20	-0.15	-0.05	0.26	-0.20	0.17	0.12
Upper Tercile	DJF	0.33	0.35	0.34	0.32	0.03	-0.05	0.08	-0.07	0.49	0.49	0.55	0.56	-0.03	0.08	0.11	-0.16
	MAM	0.12	0.14	0.19	0.13	0.03	0.02	0.07	-0.02	0.14	0.21	0.16	0.14	-0.10	0.02	-0.14	0.02
P10	DJF	0.33	0.34	0.31	0.34	0.04	0.02	0.08	0.05	0.83	0.87	0.81	0.81	0.61	0.58	0.67	0.67
	MAM	0.19	0.22	0.20	0.24	-0.12	-0.12	-0.02	-0.03	0.61	0.61	0.59	0.67	-0.22	-0.17	0.05	0.00
P90	DJF	0.02	0.03	-0.08	-0.06	-0.15	-0.11	-0.02	-0.11	0.26	0.23	-0.03	-0.04	-0.23	-0.08	-0.05	-0.16

Table 7.1.- BSS and AUCSS for snow depth seasonal forecasts aggregated at the seasonal scale (DJF and MAM) for ECMWFS5 and MFS6 models, for each experiment, each tercile and for extreme events below the 10th and above the 90th percentile of the distribution, for station Rifugio Gastaldi. Positive values of the scores, indicating an added value of the forecast system with respect to the reference forecast based on climatology, are highlighted in green.

The snowpack prototype provides skillful tercile-based snow depth forecasts, especially for the lower and upper tercile, and especially when driven by ECMWFS5 model forcing. MFS6 forcing shows large temperature and precipitation biases.

With respect to the glacier prototype, in terms of accuracy, it provides a relatively good tercile-based skill in forecasting glacier length change for the lower tercile when driven by ECMWFS5 model forcing. Moreover, a general good skill is detected across all terciles in terms of discriminations among events-non events compared to a simple climatology (see Annex 9).

8.- Assessment of indicators for prototypes on agriculture and forestry

8.1.- Seasonal forecasts for estimation of cereal yield in Castilla y León

This already implemented service (http://cosechas.itacyl.es/) estimates the cereal crops yield in Castilla y León (Spain) making use of April to June seasonal forecasts based either on climatology or on downscaled ECMWF System-5.

A synthetic yield database serving as observed truth (REF) and three different experiments based on Aquacrop model have been set up and run over the hindcast period 1995-2018 (harvest year) on a 5km resolution grid covering Castilla y León (see Annex 7 for more details). Yearly, all experiments start on 25th Sept. and until 1st April, Aquacrop is driven by observational data. The tree experiments carried out differ on the meteorological forcing used to drive Aquacrop from 1st April to 30th June: i) CLIM exp. is based on a 24 members ensemble describing climatological conditions; ii) SEAS exp. with forcing from ECMWF System-5 forecasts; iii) SEAS-D exp. same as SEAS but with downscaled precipitation and temperature.

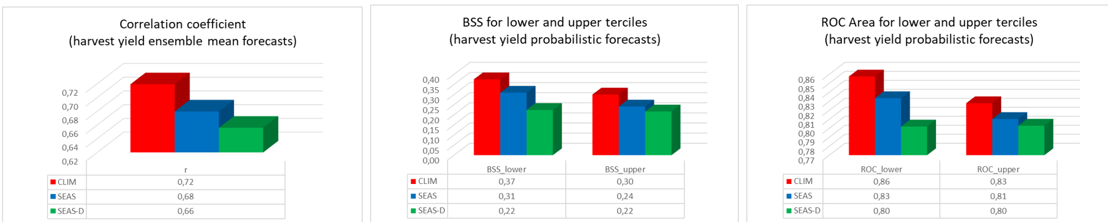


Figure 8.1.- Correlation coefficient between the ensemble mean of wheat harvest yield forecasts and a reference simulation (taken as observation) (left); BSS for lower and upper yield tercile forecasts (middle); ROC area for lower/upper yield tercile forecasts (right) for experiments CLIM, SEAS and SEAS-D over 1995-2018 hindcast period.

All experiments show very good skill according to these verification scores (Fig 8.1). Probabilistic forecasts seem to perform slightly better for lower tercile than for upper tercile. Skill differences between the experiments are smaller in case of upper wheat harvest tercile. The similar performance of three experiments is very likely due to the Aquacrop **memory from past autumn/winter observational forcing**. It also explains the good verification scores obtained by the three experiments. The experiment mimicking the current operations (CLIM) shows the best performance according to verification scores. However, in years with drastic changes in the climatological character between the past autumn-winter and the following spring the usage of seasonal forecasts of meteorological parameters shows to be very promising (provided seasonal forecast is skillful).

8.2.- Seasonal forecasts of climatic indicators for ecosystem management (Fondazione CMCC)

The prototype focuses on agroclimatic indicators related with water availability and fire risk which provide useful information for forest management and agriculture. Two seasonal prediction systems provided data through the Copernicus Climate Change Service (C3S): CMCC Seasonal Prediction System (SPS) v3 and ECMWF Seasonal Forecast system (SEAS) 5. The evaluation of the prototype has been carried out over the hindcast period 1993-2015 through the ERA5 climate dataset. Post-processing correction techniques based on the R package CS_tools were applied and predictions performance were assessed using deterministic and probabilistic evaluation scores. We also investigated the performance of agroclimatic indicators based on the effects of 1) combining several climate variables within the indicators, 2) applying different post-processing techniques for data correction, 3) exploiting different seasonal prediction systems, and 4) exploring predictions accuracy for out-of-the-norm and extreme events.

Results show that correlation between the forecasts and observations varies across climate models, indicators, and regions. Despite these differences, the preliminary evaluation of these data highlights skillful predictions in certain areas in western and Eastern Europe and North Africa. A multi-model combination of ACC results (Fig. 8.2) defines the specific areas and spatial extent where either one or a combination of the two seasonal climate prediction systems provides the best results for the three indicators, together with an assessment of significant correlations by the best model choice and combination.

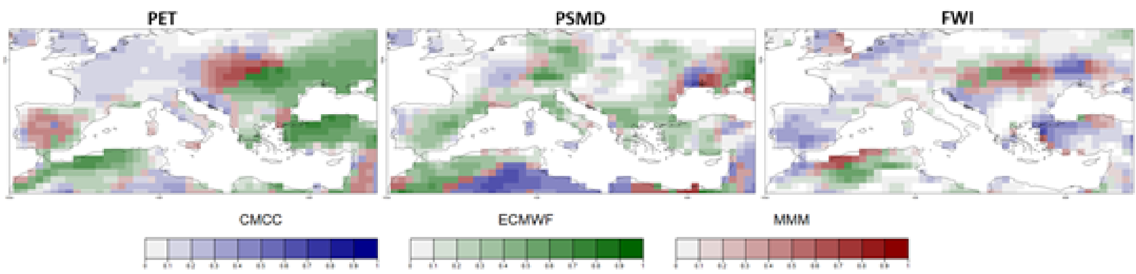


Figure 8.2.- Anomalies cross-correlations between seasonal prediction systems and ERA5 after quantile mapping corrections. The figure shows the greatest ACC among models and the multi-model mean (MMM).

Predictions of out-of-the-norm events (i.e., those over the 66th percentile and assessed using the BSS) follow similar spatial patterns as correlations but with more limited predictability in terms of extent. Forecast of out-of-the-norm Potential Evapotranspiration (PET) shows lower accuracy in the Iberian Peninsula and coastal areas in North Africa. Out-of-the-norm Potential Soil Moisture Deficit (PSMD) forecasts also show lower accuracy in Spain and some central European areas, whereas for FWI, the accuracy is mostly conserved. The accuracy assessment of detecting extreme events (i.e., events over the 95th percentile) highlights that spatial patterns show a general decrease except for some areas where it increases. PET and PSMD accuracy is more robust in some areas of northern Spain and Eastern Europe, whereas for FWI, it is in Algeria.

8.3.- AgriMetInfo seasonal forecasts (INRAE)

The prototypes aim at producing seasonal forecasts (SF, 6 months) tailored for agriculture and forestry sectors where management decisions might be improved by a better knowledge of the climatic conditions during the forthcoming months. Although decisions can be taken throughout the year, the evaluation has focused on a single SF period (6 month from May) covering the main period of active vegetation development. This study has focused on three specific cases covering different issues in the field of agriculture and forestry: i) risk of heat waves or drought events; ii) agriculture water requirements and crop productions using regionalized crop models; iii) risk of plant dehydration and wildfire activity (see Annex 10 for more details)

In all cases, evaluation has analyzed the following aspects: i) memory effect; ii) seasonal forecast evaluation using deterministic and probabilistic scores; iii) evaluation of indicators also using deterministic and probabilistic scores; iii) economic assessment.

Figure 8.3 shows, as an example (see Annex 10 for a detailed evaluation), area under ROC curve (AUC) of 4 fire indicators for the upper and lower tercile using seasonal forecasts and observed data. AUC values are higher on average with seasonal forecasts (0.65 on average) than with average climate data (0.55 on average).

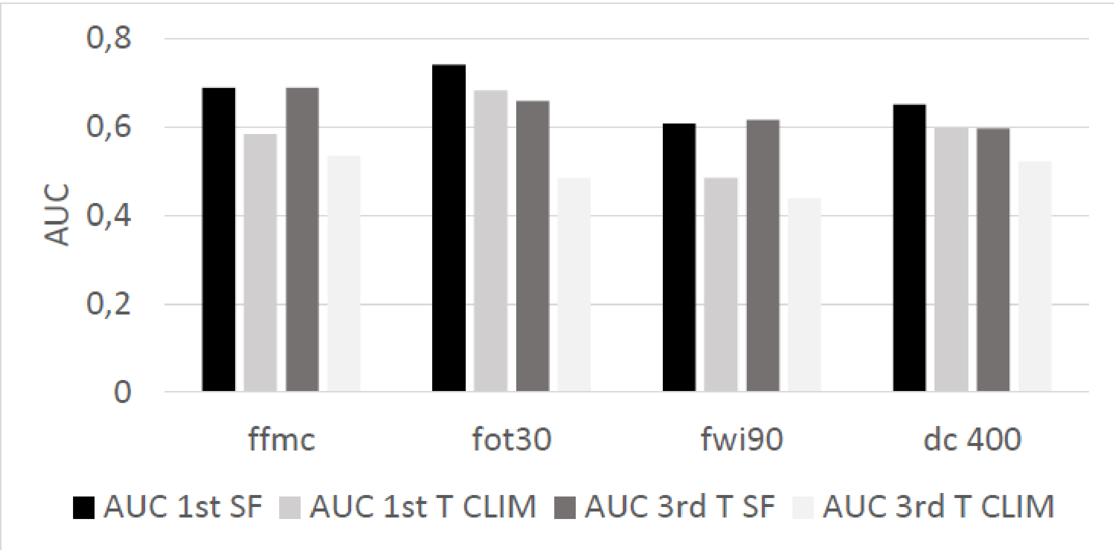


Figure 8.3.- AUC for upper/lower terciles of four FIRE indicators (ffmc, fot30, fwi90, dc 400) computed with averaged observed data and with seasonal forecasts over the French Mediterranean region.

9.- Conclusions

Deliverable D4.1 and the accompanying Annexes have summarized the big effort conducted in MEDSCOPE WP4 to evaluate developed prototypes. Evaluation was mostly done using the existing retrospective forecasts (hindcasts) from operational SFSs. This evaluation has been conducted not only for seasonal forecasts coming from different sources but also for variables and indicators identified by each sector as relevant. In most cases, sectoral indicators are expressed in the form of deterministic and probabilistic forecasts which are finally verified following the standard procedures for seasonal forecast verification including cost/benefit scores. Most details of probabilistic and deterministic scores for each prototype are collected in the Annexes, presenting the main text only a brief summary of evaluation results and cross cutting issues.

Questions linked to the usage of application models -typically used to translate climate variables into users' defined variables and indicators - were thoroughly discussed in Annexes for different prototypes. Users' variables and indicators cover a wide typology and depend strongly on the specific sector under consideration.

The main text also includes and discusses a number of issues common to many prototypes, such as, the usage of verifying (actual versus synthetic) observations, the comparison against a reference forecast (usually based on climatology), the important role of memory (in agriculture/forestry prototypes), etc.

Although predictability and SFSs skill over the Mediterranean region is relatively low, some prototypes have benefited from untapped windows of opportunity for certain seasons, regions or SFSs or from the predictability coming from other sources such as the memory of certain variables carried by application models. This last case is especially notable in the case of agricultural growth models.

9.- References

- Balsamo G, Viterbo P, Beljaars A, van den Hurk B, Hirschi M, Betts AK, Scipal K., 2009. A revised Hydrology for the ECMWF model: verification from field site to terrestrial water storage and impact in the integrated forecast system. Journal of Hydrometeorology 10: 623– 643.
- Bartelt, P., & Lehning, M. , 2002. A physical SNOWPACK model for the Swiss avalanche warning: Part I: numerical model. Cold Regions Science and Technology, 35(3), 123-145.
- CEDEX. 2016. Evaluación de los Recursos Hídricos en España. Monografía CEDEX. ISBN: 978-84-7790-578-3. 374 pags. .
- Gandin, L. S., and A. H. Murphy, 1992: Equitable skill scores for categorical forecasts. Mon. Wea. Rev., 120, 361–370, https://doi.org/10.1175/1520-0493(1992)120,0361:ESSFCF.2.0.CO;2 .
- Gerrity, J. P., Jr., 1992: A note on Gandin and Murphy's equitable skill score. Mon. Wea. Rev., 120, 2709–2712, https://doi.org/10.1175/1520-0493(1992)120,2709:ANOGAM.2.0.CO;2.
- Habets, F., Boone, A., Champeaux, J. L., Etchevers, P., Fran-chistéguy, L., Leblois, E., Ledoux, E., Le Moigne, P., Martin, E.,Morel, S., Noilhan, J., Quintana Segui, P., Rousset-Regimbeau,F., and Viennot, P., 2008: The SAFRAN-ISBA-MODCOU hydrometeorological model applied over France, J. Geophys. Res., 113,D06113, doi:10.1029/2007JD008548.
- Lledó, Ll., Torralba, V., Soret, A., Ramon, J., and Doblas-Reyes, F. (2019). Seasonal forecasts of wind power generation. Renewable Energy, 143:91–100.
- Masson, V., P. Le-moigne, E. Martin, S. Faroux, A. Alias et al., The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes, Geosci. Model Dev, vol.6, pp.929-960, 2013.
- Materia, S., Muñoz, Á. G., Álvarez-Castro, M. C., Mason, S. J., Vitart, F., & Gualdi, S. (2020). Multimodel Subseasonal Forecasts of Spring Cold Spells: Potential Value for the Hazelnut Agribusiness. Weather and Forecasting, 35(1), 237-254. https://doi.org/10.1175/WAF-D-19-0086.1
- Oerlemans J, Jania J, Kolondra L., 2011. Application of a minimal glacier model to Hansbreen, Svalbard. Cryosphere 5:1–11. doi: 10.5194/tc-5-1-2011
- Sánchez-García, E., Voces-Aboy, J., Navascués, B. and Rodríguez-Camino, E.: Regionally improved seasonal forecast of precipitation through Best estimation of winter NAO. Adv. Sci. Res., 16, 165-174. https://doi.org/10.5195/asr-16-165-2019 , 2019.
- Viel, C., Beaulant, A.-L.,Soubeyroux,J.-M., Céron, J;-P., 2016, How seasonal forecasts could help a decision maker:an example of climate service for water resource management, Adv. Sci. Res., 13, 51–55, doi:10.5194/asr-13-51-2016
- Voces, J., Sanchez-Garcia, E., Navascués, B., Franco, F., Rodriguez-Camino, E., 2016: Sistema estadístico de predicción estacional para la gestión de los embalses en España, Nota Técnica. nº 21, AEMET. http://hdl.handle.net/20.500.11765/4431 , 2016
- Wilks, D. S. (2011). Statistical methods in the atmospheric sciences (Vol. 100). Academic press.

- World Meteorological Organization, 2018: Guidance on Verification of Operational Seasonal Climate Forecasts (WMO-No. 1220). Geneva. https://library.wmo.int/index.php?lvl=notice_display&id=20618#.XvcCI20zYs4
- World Meteorological Organization, 2020: Guidance on Operational Practices for Objective Seasonal Forecasting (WMO-No. 1246). Geneva. https://library.wmo.int/index.php?lvl=notice_display&id=21741#.XvcDEG0zYs4