Integrated approach for the **d**evelopment across **E**urope of user oriented **c**limate **i**ndicators for GFCS high-priority **s**ectors: Agriculture, disaster risk reduction, energy, health, water and tourism

Work Package 6

Deliverable 6.3

Report on the reliability and uncertainties associated with the (hindcast-type) seasonal forecasts of selected sectorial INDECIS indices

This report arises from the Project INDECIS which is part of ERA4CS, an ERA-NET initiated by JPI Climate, and funded by FORMAS (SE), DLR (DE), BMWFW (AT), IFD (DK), MINECO (ES), ANR (FR), with co-funding by the European Union's Horizon 2020 research and innovation programme

Seasonal forecast for the wine sector – Douro Valley case study

Nube González-Reviriego
Raül Marcos-Matamoros

Barcelona Supercomputing Center, Barcelona, Spain

Introduction

Climatic information at seasonal time-scales (up to 7 months ahead) is needed for decision-making in a number of sectors. Compared to the short-to-medium-range forecasts (up to 10 days ahead), seasonal time-scales hold the potential for being of great value for a wide range of users who are affected by variability in climate and who would benefit from understanding and better managing climate-related risks (Bruno Soares et al., 2018; Stoft, 2002).

In Europe, there has been relatively little uptake and use of seasonal forecasts by users for decision making, compared to other parts of the world, such as the USA and Australia, probably due to the relatively limited predictability and quality of the prediction systems in that area (Bennett et al., 2017; Mendoza et al., 2017; Arnal et al., 2018). However, recent advances in the understanding and forecasting of climate have resulted in skillful and useful climatic predictions, which can consequently increase the confidence of agriculture-related prognoses, and improve awareness, preparedness and decision-making from an end-user perspective (Bruno Soares and Dessai, 2016). Winemaking is an example of agriculture applications in which seasonal information can affect decision-making.

The harvest date in the wine sector: decision-making based on climate variables

One of the most critical decisions for the wine sector is setting the harvest date. Depending on it, the quality of the grapes can change and, therefore, the wine made therefrom. The viticultors evaluate not only the grape colour, texture and taste of the grapes but also measure the sugar and acid content of the grapes to set the most adequate harvest date.

An important factor determining the harvest time is climate (Mira de Orduña, 2010; Berbegal et al., 2019), being the global warming the cause of a shift forward of the harvesting time at many wine regions. In the Douro region, the end of August, September and early October are the critical months to look for future climate conditions affecting the harvest date.

In particular, winemakers are quite interested in anticipating how precipitation will be in those months, since continuous precipitation in this period causes overcast skies delaying the ripening of the grapes. In addition to that, an excess of rainfall during these months creates a surplus of water within the grapes which, in turn, decreases the fruit sugar levels and sugar/acid balance that winemakers are looking for (having consequences on the wine flavour). Apart from that, grapes could bloat and even burst (causing spoilage), and, moreover, too much humidity favors mildew and fungus diseases. Several mechanisms can be activated to face such situations, but they all need accurate seasonal forecasts of the August-September-October (ASO) season to optimize decision-making.

In this study, BSC analyses three historical years based on ASO precipitation highlighted by winemaker end-users from the Douro region (MEDGOLD, 2020). Two are good years, 2007 and 2011, and one is a bad one, 2002, attending to a classification based on the quantity and quality of the wine obtained those years.

For each year, a detailed description is provided. This description includes a short explanation of the viticultural year from the wine user's perspective, a description of the ASO season by means of analysing ERA5 reanalysis dataset and the European Centre for Medium-Range Weather Forecasts (ECMWF) System 5 (SEAS5) seasonal forecast available at that time. In the discussion, relevant skill scores are also presented to better understand the potential application of seasonal forecast. Specifications of the data used, bias adjustment applied and forecast verification are described in the Annex.

Summary of the case study

Precipitation over Iberian Peninsula in late Summer and early Autumn could damage grape quality and cause outbreaks of rot.			
Region:	Douro Valley	**Period:**	August-September-October (ASO)
Forecast type:	Seasonal predictions	**Main interest:**	Wine sector
Forecast index:	ASO precipitation		

Year 2002

Viticultural characterization of the year

Late bud-break, due to some extreme episodes of low temperatures in winter and, especially to low winter and early spring precipitation, causing moderate to severe drought episodes early in the season. Disease pressure was almost absent (disease risk period from 14-May to 27-May, 13 days). However, vegetative growth was, thus, stunted, which made for harvests to start in mid-September and last well into October when equinoctial rains severely damaged grape quality, diluting their contents and causing local outbreaks of rot (Botrytis) in some areas.

Analysis of the observed ASO precipitation

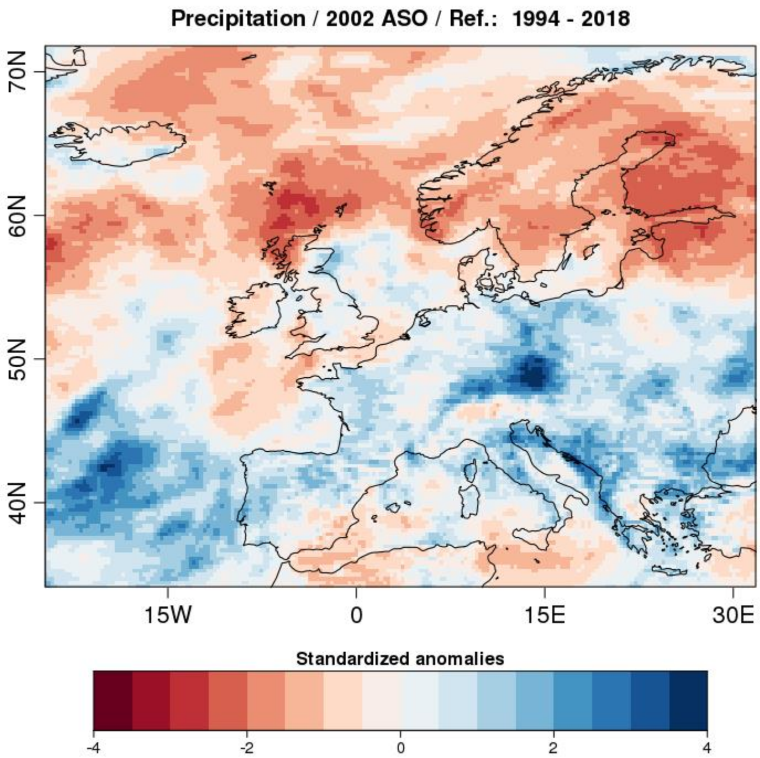

Figure 1. Standardized precipitation anomalies for the ASO 2002 season obtained from ERA5 reanalysis (reference period 1994-2018).

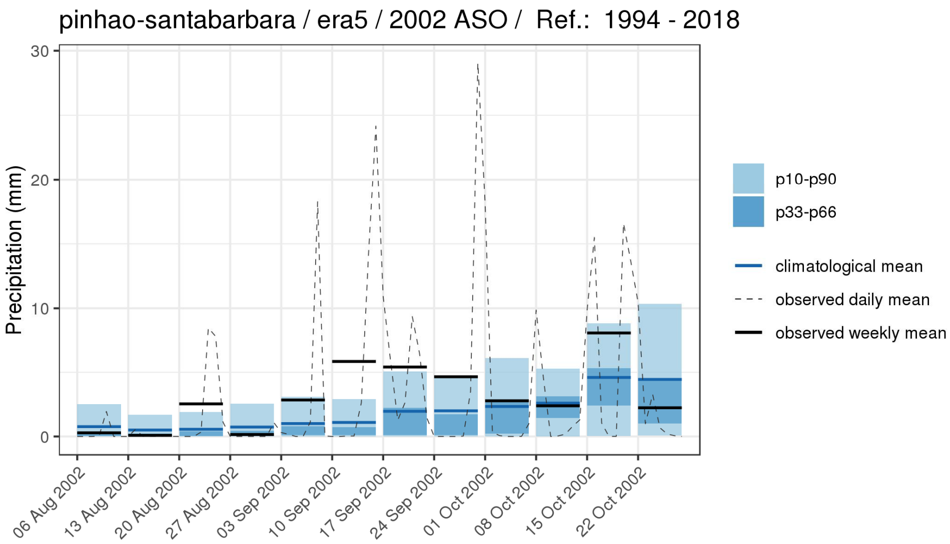

Figure 2. Weekly evolution of the observed precipitation for the Pinhao - Santa Barbara nearest grid point during ASO 2002 season compared to the climatological distribution. Values obtained from ERA5 reanalysis (reference period 1994-2018).

In 2002, the precipitation amount across most areas of central and western Europe during the August-September-October period (ASO) was higher than the climatological mean (Figure 1). This behaviour was also observed in the Douro Valley, where the analysis of the Pinhao-Santa Barbara grid point precipitation time series enables the identification of the weekly accumulations that lead to this humid anomaly (Figure 2). More specifically, the period comprised between the 3rd of September and the 1st of October displayed a continuous positive precipitation anomaly. It is worth stressing that observed high precipitation values on the week starting at 10th September are extreme values outside the climatological 90th percentile. Other weeks with rainfall above climatological values in ASO were the week starting on the 20th of August and the one starting the 15th of October (see Figure 2).

Available forecasts for ASO precipitation

Figure 3 shows a sample of seasonal forecasts that were available months in advance for this event (ASO 2002 seasonal precipitation). The observed precipitation standardized anomalies (purple diamond in Figure 3) for the ASO 2002 season were in the above- normal tercile category for the Pinhao-Santa Barbara region (humid conditions). Figure 3 demonstrates the behaviour of a forecast with positive low skill (see Fair RPSS values of Table 1 and also explanation below). At lead times in excess of one month, the predictions are similarly distributed (around 30%) between the three categories, being the most dominant tercile, the above- and below- normal category for predictions issued respectively in May and June 2002. In this situation, it is difficult to use these predictions for making a decision, since the probabilities are very close to each other. At shorter lead times (1 month ahead - July 2002) the forecast is able to predict the above- normal tercile category (corresponding with the observed tercile category) with a probability of 41%.

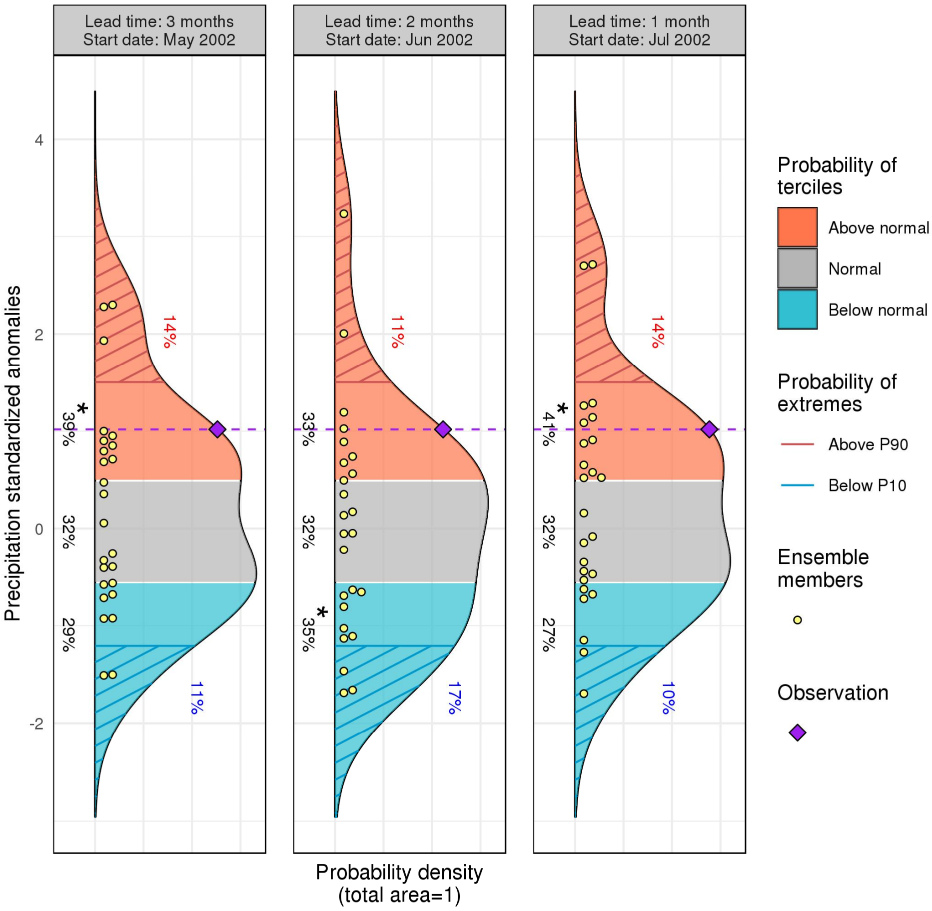

Figure 3. Seasonal predictions for tercile categories (above normal, normal and below normal) of precipitation standardized anomalies for the Pinhao - Santa Barbara nearest grid point. From left to right, seasonal predictions for ASO 2002 season issued in May (3 months in advance), June (2 months in advance) and July (1 month in advance) of the same 2002 year. The percentage values shown on the left side of each category represent the predicted probability of precipitation being in that category. Predicted extreme values are represented with dashed areas and the corresponding probability of being in them are shown on the right side.

Table 1. Verification metrics for SEAS5 predictions of ASO seasonal precipitation issued in May, June and July for the year 2002. The probabilistic metrics considered are the fair version of RPSS, BSS P10, BSS P90 and CRPSS together with the deterministic EnsCorr metric. The hindcast period is 1993-2018.

	May	Jun	Jul
RPSS	0	0.05	0.13
BSS P10	0.02	0.01	0
BSS P90	0.02	0	0
CRPSS	0	0.01	0.06
EnsCorr	0	0.33	0.42

The skill of the seasonal predictions issued one to three months in advance (July to May) to ASO 2002 season show skill compared to climatology in the June and July started month predictions (respective RPSS of 0.05 and 0.13, see Table 1). Other skill metrics such as the ensemble correlation (values of 0.33 and 0.42) and the CRPSS (values of 0.01 and 0.06) also support this inference. The extreme forecast (percentiles 10 and 90), though, show no skill when compared to climatology. This is important because end-users can only trust predictions that have skill regarding the specific predictions that match their decision making (e.g. the RPSS related to decision based on the most probable tercile category). That said, it is important to remember that positive values of the RPSS and CRPSS metrics imply that, statistically, the use of tercile seasonal forecasts give better results than the customary use of climatology. However, since the perfect score for these metrics is 1, the closer the computed metric is to this number the easier for the user to see good results and the other way around when the number is closer to 0.

Year 2007

Viticultural characterization of the year

Winter with normal precipitation made for good water reserves that lasted well into summer because of milder temperatures, not too far above normal values. Disease pressure was moderate to high but still within control, as temperatures did not promote outbreaks (disease risk period from 30-Apr to 16-Jul, 77 days). Regular episodic and moderate rain events during late summer were not enough to cause disease outbreaks and helped maintain a good hydric situation in the grapevines, leading to extremely high-quality grapes, especially in terms of aromatics and color intensity. Harvest started mid-September without any major rain events, sustaining the high quality. This was not a good year for people with heart issues, as we veered in the brink of disaster since Spring to harvest.

Analysis of the observed ASO precipitation

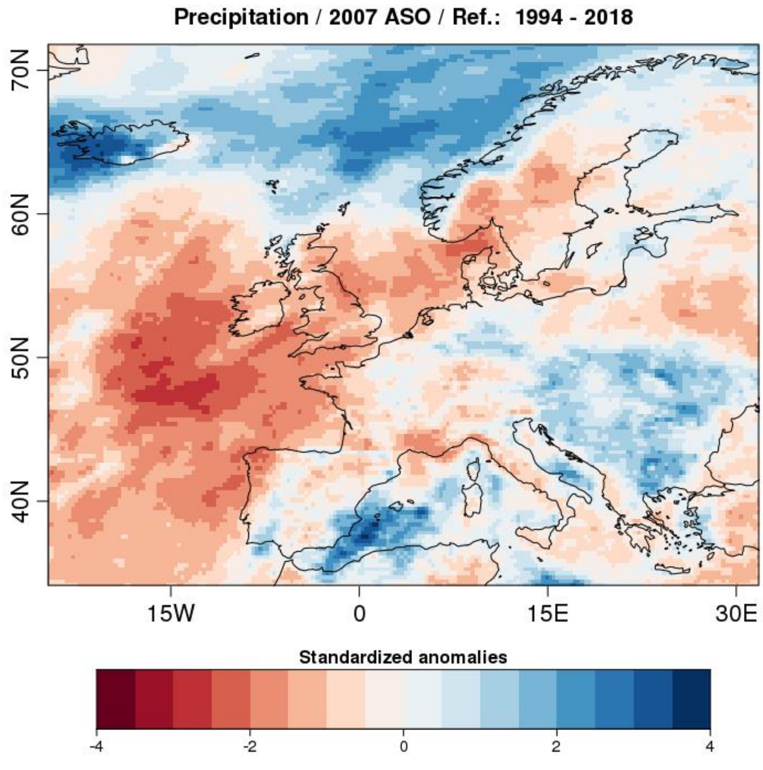

Figure 4. Standardized precipitation anomalies for the ASO 2007 season obtained from ERA5 reanalysis (reference period 1994-2018).

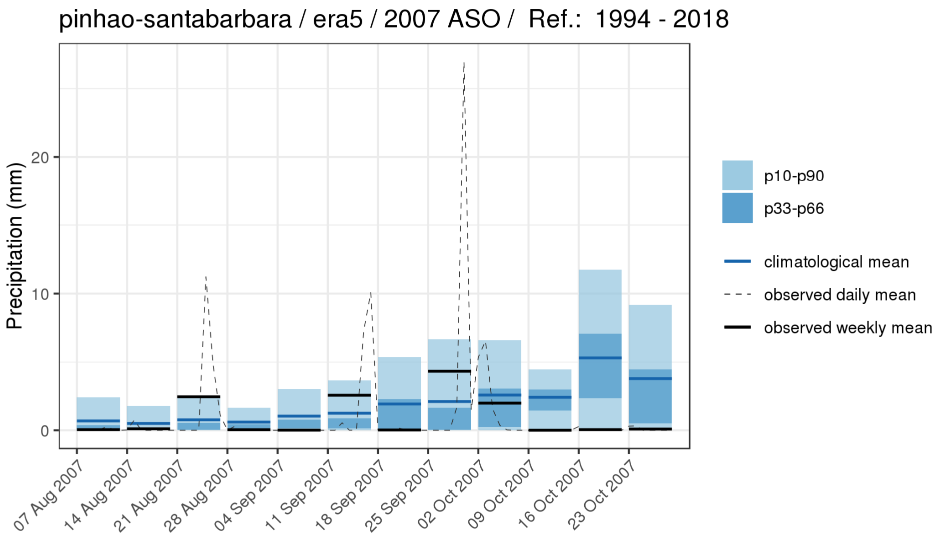

Figure 5. Weekly evolution of the observed precipitation for the Pinhao - Santa Barbara nearest grid point during ASO 2007 season compared to the climatological distribution. Values obtained from ERA5 reanalysis (reference period 1994-2018).

In 2007, the precipitation amount across the Atlantic shores of western and southwestern Europe during the August-September-October period (ASO) was lower than the climatological mean (see Figure 4). This behaviour was also observed in the Douro Valley, where the analysis of the Pinhao-Santa Barbara grid point precipitation time series enables the identification of the weekly accumulations that lead to this dry anomaly (Figure 5). More specifically, the period comprised between the 7th of August and the 4th of September displayed an overall negative anomaly of precipitation. Other weeks with no rainfall and below climatological values in ASO were the three weeks ranging from 9th to 30th October 2007 (see Figure 5).

Available forecasts for ASO precipitation

Figure 6 shows a sample of seasonal forecasts that were available for ASO precipitation in 2007. The observed precipitation standardized anomalies (purple diamond in Figure 6) for the ASO 2007 season were in the below-normal tercile category for the Pinhao-Santa Barbara region (dry conditions). Figure 6 demonstrates the behaviour of a forecast with limited skill. At high lead times (3 month ahead - May 2007), the probabilities of the three categories are quite similar with a slightly higher probability (36%) for above- normal tercile category – i.e., the observed weaker than normal precipitation standardized anomalies are not predicted by the majority of the ensemble members (consistent with the zero fair RPSS value shown in Table 2). However, the predictions issued on June and July 2007 tend to suggest likely lower-than-normal precipitation standardized anomalies, though members underestimate the severity of the event and the normal tercile category is estimated to be more probable than below-normal, with respective probabilities of 56% and 41%.

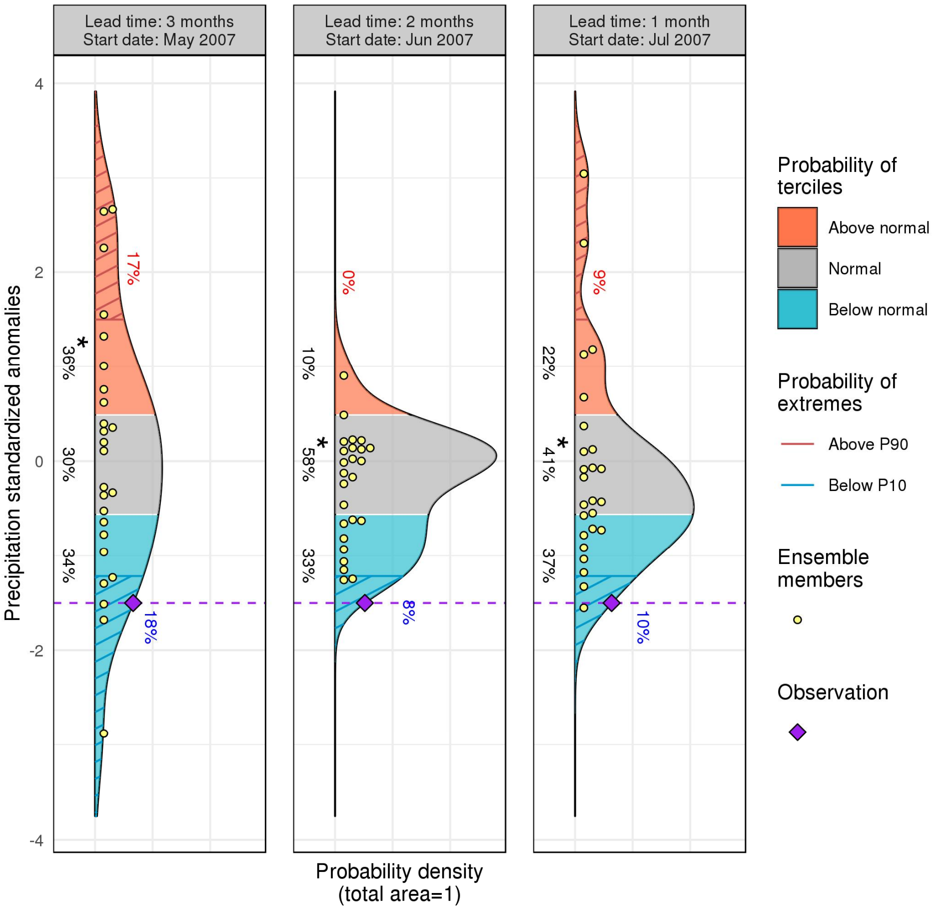

Figure 6. Seasonal predictions for tercile categories (above normal, normal and below normal) of precipitation standardized anomalies for the Pinhao - Santa Barbara nearest grid point. From left to right, seasonal predictions for ASO 2007 season issued in May (3 months in advance), June (2 months in advance) and July (1 month in advance) of the same 2007 year. The percentage values shown on the left side of each category represent the predicted probability of precipitation being in that category. Predicted extreme values are represented with dashed areas and the corresponding probability of being in them are shown on the right side.

Table 2. Verification metrics for SEAS5 predictions of ASO seasonal precipitation issued in May, June and July for the year 2007. The probabilistic metrics considered are the fair version of RPSS, BSS P10, BSS P90 and CRPSS together with the deterministic EnsCorr metric. The hindcast period is 1993-2018.

	May	Jun	Jul
RPSS	0	0.03	0.12
BSS P10	0	0	0
BSS P90	0.02	0	0
CRPSS	0	0.01	0.06
EnsCorr	0	0.26	0.44

The skill of the seasonal predictions issued one to three months in advance (July to May) to ASO 2007 season shows no skill for precipitation at May and low positive skill compared to climatology at June and July started month predictions (respective RPSS of 0.03 and 0.12, see Table 2). Other skill metrics such as the ensemble correlation (values of 0.26 and 0.44) and the CRPSS (values of 0.01 and 0.06) also support this inference. The extreme predictions (percentiles 10 and 90), though, show no skill when compared to climatology.

Year 2011

Viticultural characterization of the year

Mild temperatures during the whole growth cycle caused for a relatively late bud-break and a smooth, regular vegetative growth with hardly any disease pressure (disease risk period from 20-Apr to 30-May, 40 days). The absence of the usual high temperature extremes during the maturation phase, together with small amounts of precipitation, not enough to cause disease pressure but sufficient to support the grapevine hydric status led to the development of grapes very balanced and even with more concentration than 2007, that were harvested untouched and in pristine condition because of overall dry harvest conditions. Considered as a perfect year.

Analysis of the observed ASO precipitation

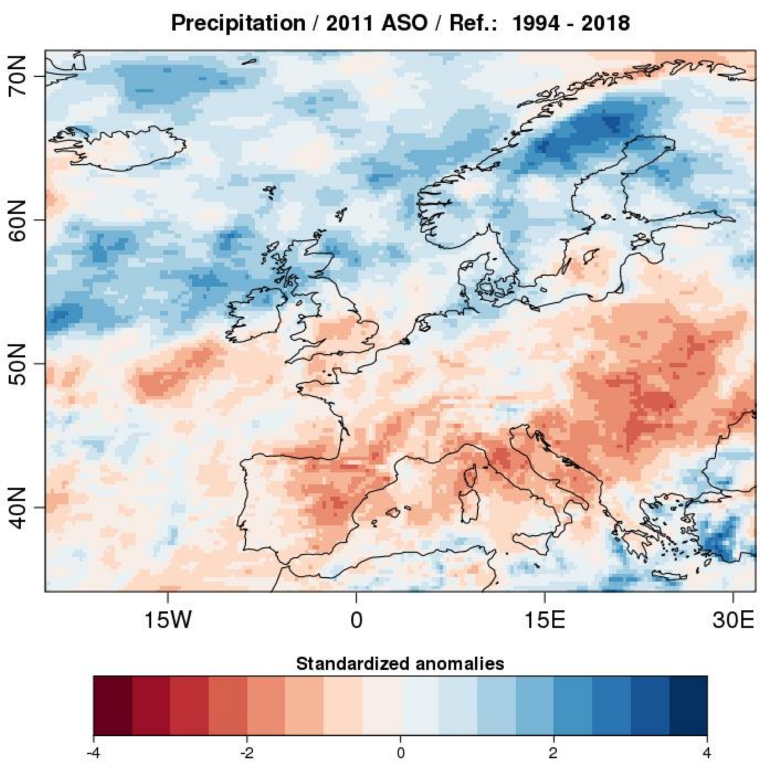

Figure 7. Standardized precipitation anomalies for the ASO 2011 season obtained from ERA5 reanalysis (reference period 1994-2018).

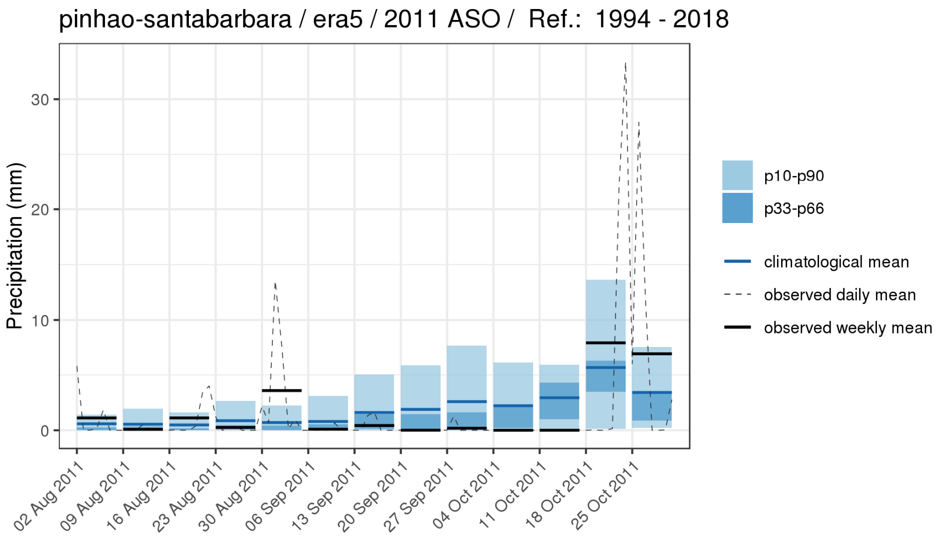

Figure 8. Weekly evolution of the observed precipitation for the Pinhao - Santa Barbara nearest grid point during ASO 2011 season compared to the climatological distribution. Values obtained from ERA5 reanalysis (reference period 1994-2018).

In 2011, the precipitation amount across most of central, south and western Europe during the August-September-October period (ASO) was lower than the climatological mean (see Figure 7). This behaviour was not that clear in the Douro Valley, where the analysis of the Pinhao-Santa Barbara grid point precipitation time series (Figure 8) enables the identification of dry and humid weekly accumulations that mostly compensate in overall, leaving a slight dry anomaly for the 3-month aggregation. More specifically, the period between the 6th of September and the 18th of October was very dry, with almost no rain at all. On the contrary, during the interval between the 18th of October and the 1st of November, the rainfall was high enough to mainly compensate for the dry anomaly and leave ASO season with a small dry anomaly (see Figure 8). Nevertheless, it is worth pointing out that the precipitation event that happened at the end of October was outside of the harvest dates for this year 2011 according to the end-user description.

Available forecasts for ASO precipitation

Figure 9 shows a sample of seasonal forecasts that were available for ASO precipitation in 2011. The observed precipitation standardized anomalies (purple diamond in Figure 9) for the ASO 2011 season were in the normal tercile category for the Pinhao-Santa Barbara region (normal conditions). Figure 9 shows that at high lead times (3 month ahead - May 2011), the predictions are similarly distributed, being the probabilities of the three categories quite similar with a slightly higher probability (38%) for below- normal tercile category – i.e., the observed normal precipitation standardized anomalies are not predicted by the majority of the ensemble members (consistent with the zero fair RPSS value shown in Table 3). Nevertheless, at shorter lead times (2 and 1 months ahead - June, July 2011) the forecasts are able to predict the normal tercile category (corresponding with the observed tercile category) with a probability of 40% and 44% (see Figure 9, middle and right panels).

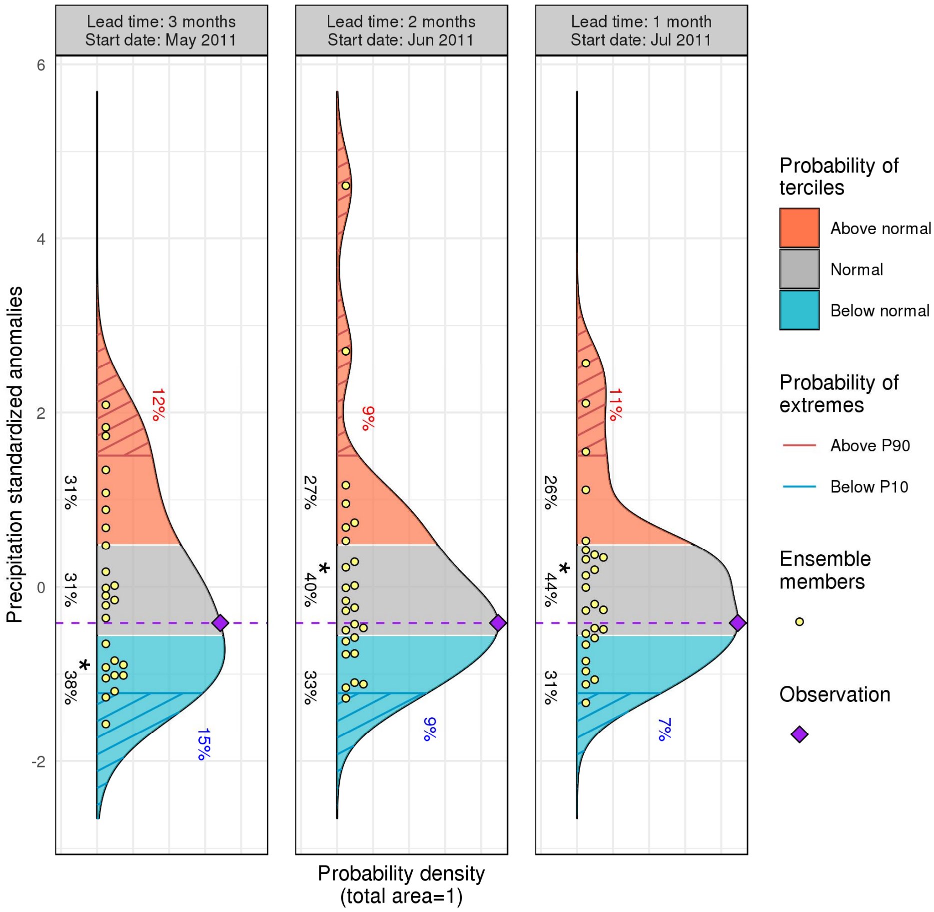

Figure 9. Seasonal predictions for tercile categories (above normal, normal and below normal) of precipitation standardized anomalies for the Pinhao - Santa Barbara nearest grid point. From left to right, seasonal predictions for ASO 2011 season issued in May (3 months in advance), June (2 months in advance) and July (1 month in advance) of the same 2011 year. The percentage values shown on the left side of each category represent the predicted probability of precipitation being in that category. Predicted extreme values are represented with dashed areas and the corresponding probability of being in them are shown on the right side.

Table 3. Verification metrics for SEAS5 predictions of ASO seasonal precipitation issued in May, June and July for the year 2011. The probabilistic metrics considered are the fair version of RPSS, BSS P10, BSS P90 and CRPSS together with the deterministic EnsCorr metric. The hindcast period is 1993-2018.

	May	Jun	Jul
RPSS	0	0.04	0.12
BSS P10	0.01	0.01	0
BSS P90	0.02	0	0
CRPSS	0	0.01	0.06
EnsCorr	0	0.34	0.45

The skill of the seasonal forecasts issued one to three months in advance (July to May) show no skill at May and positive skill compared to climatology in the June and July started month predictions (respective RPSS of 0.04 and 0.12, see Table 3). Other skill metrics such as the ensemble correlation (values of 0.34 and 0.45) and the CRPSS (values of 0.01 and 0.06) also support this inference. The extreme forecast (percentiles 10 and 90), though, show no skill when compared to climatology.

Summary and Conclusions

The analysis conducted in this report aims to illustrate the potential of the seasonal forecast in predicting the ASO precipitation index for the Pinhao - Santa Barbara region at Douro Valley (Portugal) for helping winemakers in making informed decisions regarding harvest. BSC has analysed three different years, selected by the winemakers, as examples where climate conditions lead to having good (2007 and 2011) or bad (2002) years in terms of quality and quantity of wine production.

The study begins by climatically describing the viticultural year as a whole, to subsequently focus on the August-September-October accumulated rain anomaly. Firstly, the general pattern is framed by looking at the observed European 3-month precipitation standardized anomalies. Afterwards, the analysis is focused on the Pinhao-Santa Barbara nearest grid point, to identify the evolution of the observed weekly precipitation amounts that account for the total ASO anomaly. In this way it is possible to highlight which periods contributed most to the overall behaviour of the 3-month interval. Finally, after this characterization, the bias-adjusted seasonal predictions for tercile categories of precipitation standardized anomalies during ASO season issued three months in advance are introduced as well as the skill metrics for each year.

The seasonal predictions for tercile precipitation events over the Pinhao - Santa Barbara region at Douro Valley (Portugal) are skillful one month in advance, based on the Fair RPSS values obtained. At higher lead times, the skill (Fair RPSS) decreases, being still positive two months in advance but unskillful 3 months in advance. The positive skill found in this report considering the hindcast period 1993-2018 implies an added value in a long term use of seasonal forecasts against the customary use of climatology as a future estimation of ASO precipitation.

When looking at individual years, in two of them, 2002 and 2011, the one-month ahead most likely predicted tercile category for precipitation over the Pinhao - Santa Barbara region coincides with the observation. Conversely, in 2007 the most likely predicted tercile doesn't match with the finally observed. This illustrates the uncertainty linked to the seasonal forecasts with limited skill (positive low values of Fair RPSS).

Although BSC analysis has demonstrated the potential of seasonal forecasts issued in July and June for predicting the ASO precipitation and, hence, the potential added value of seasonal forecasts on decision-making for the wine industry, it is important to highlight that an economic assessment is always needed to analyse its impact for each particular decision. Nevertheless, this is out of the scope of this report and remains open for future assessments.

References

Arnal, L., Cloke, H. L., Stephens, E., Wetterhall, F., Prudhomme, C., Neumann, J., ... & Pappenberger, F. (2018). Skilful seasonal forecasts of streamflow over Europe?. *Hydrology and Earth System Sciences*, *22*(4), 2057-2072.

Bennett, E. M. (2017). Research frontiers in ecosystem service science. *Ecosystems*, *20*(1), 31-37.

Berbegal, C., Fragasso, M., Russo, P., Bimbo, F., Grieco, F., Spano, G., & Capozzi, V. (2019). Climate changes and food quality: The potential of microbial activities as mitigating strategies in the wine sector. *Fermentation*, *5*(4), 85.

C3S, Copernicus Climate Change Service (2017). ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS), Available at: https://cds.climate.copernicus.eu/cdsapp#!/home

De Orduna, R. M. (2010). Climate change associated effects on grape and wine quality and production. *Food Research International*, *43*(7), 1844-1855.

Doblas-Reyes, F.J., Hagedorn, R., Palmer, T.N. (2005). The rationale behind the success of multi-model ensembles in seasonal forecasting II. Calibration and combination. Tellus A 57(3):234–252, doi: 10.1111/j.1600-0870.2005.00104.x

Doblas-Reyes, F.J., García-Serrano, J., Lienert, F., Biescas, A.P., Rodrigues, L.R.L. (2013). Seasonal climate predictability and forecasting: status and prospects. Wiley Interdisciplinary Review Climate Change 4(4):245–268, doi: 10.1002/wcc.217

Epstein, E. S. (1969). A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology*, *8*(6), 985-987

Ferro, C. A. T. (2014). Fair scores for ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society*, 140(683), 1917-1923, doi: 10.1002/qj.2270

Jolliffe, I. T., & Stephenson, D. B. (Eds.). (2012). *Forecast verification: a practitioner's guide in atmospheric science*. John Wiley & Sons.

Manzanas, R., Gutiérrez, J.M., Bhend, J., Hemri, S., Doblas-Reyes, F.J., Torralba, V., Penabad, E., Brookshaw, A. (2019). Bias adjustment and ensemble recalibration methods for seasonal forecasting: a comprehensive intercomparison using the C3S dataset. Climate Dynamics: 1-19, doi:10.1007/s00382-019-04640-4.

MEDGOLD (2020). D3.2: Report on the Methodology followed to implement the wine pilot services. MEDGOLD project H2020-SC5-01-2017, 79pp.

Mendoza, P., Wood, A. W., Clark, E., Rothwell, E., Clark, M. P., Nijssen, B., ... & Arnold, J. R. (2017). An intercomparison of approaches for improving operational seasonal streamflow forecasts.

Soares, M. B., Alexander, M., & Dessai, S. (2018). Sectoral use of climate information in Europe: A synoptic overview. *Climate Services*, *9*, 5-20.

Soares, M. B., & Dessai, S. (2016). Barriers and enablers to the use of seasonal climate forecasts amongst organisations in Europe. *Climatic Change*, *137*(1-2), 89-103.

Stoft, S. (2002). Power System Economics: Designing Markets for Electricity, Wiley-IEEE Presshttp://dx. doi. org/10.1109/9780470545584.

Torralba, V., Doblas-Reyes, F.J., MacLeod, D., Christel, I., Davis, M. (2017). Seasonal climate prediction: a new source of information for the management of wind energy resources. Journal of Applied Meteorology and Climatology. 56(5):1231–1247, doi:10.1175/JAMC-D-16-0204.1

Weisheimer, A., & Palmer, T. N. (2014). On the reliability of seasonal climate forecasts. *Journal of the Royal Society Interface*, *11*(96), 20131162, doi: 10.1098/rsif.2013.1162

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (Vol. 100). Academic press. (m/s)

ANNEX

Observed and forecasted precipitation datasets

ERA5

ERA5 is the latest climate reanalysis produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), providing hourly data on many atmospheric, land-surface and oceanic climate variables for the period from 1979 to the present. This reanalysis replaces the ERA‐Interim reanalysis which was started in 2006. ERA5 is based on the Integrated Forecasting System (IFS) Cy41r2 (operational in 2016). ERA5 benefits from a decade of developments in model physics, core dynamics and data assimilation. In addition to a significantly enhanced horizontal resolution of 31 km (compared to 80 km for ERA‐Interim), ERA5 has hourly output, and an uncertainty estimate from an ensemble (3‐hourly at half the horizontal resolution). The reanalysis includes 137 levels in the atmosphere (up to 0.01hPa) on a regular latitude-longitude grid. For more information on this reanalysis see (C3S, 2017).

ECMWF SEAS5

The ECMWF SEAS5 seasonal predictions dataset obtained from the Climate Data Store of the Copernicus Climate Change Service (CDS-C3S) is the fifth generation of seasonal prediction which replaces the former System 4 and uses the Integrated Forecast System (IFS) Cycle 43r1. There are 25 ensemble members for the hindcast period from 1993 to 2016. SEAS5 includes enhancements in the land-surface initialization, atmospheric resolution and the ocean model when compared to S4. For instance, regarding the land-surface initialization, the former includes a new offline recalculation at the native atmospheric resolution with an improved precipitation forcing. The performance tests show a good degree of consistency between the initialization of SEAS5 re-forecast and the real-time predictions while the initialization is not perfect (i.e., the real-time assimilation is not identical as reanalysis). In addition, the SEAS5 uses the new version of ocean model NEMO (Nucleus for European Modelling of the Ocean) with upgraded ocean physics and resolution. Finally, the ocean and sea-ice initial conditions are provided by the new ocean analysis and reanalysis ensemble (ORAS5).

Precipitation bias-adjustment protocol

Climate predictions are affected by systematic errors resulting from the inability of global circulation models to reproduce all the relevant processes responsible for climate variability and the uncertainty affecting the initial conditions (Doblas-Reyes et al., 2013, and references therein). Hence, the bias adjustment is a fundamental step to reduce forecast errors and produce usable, tailored and high-quality climate predictions. Bias adjustment methods aim to adjust the statistical properties of climate predictions to those of an observational reference (Torralba et al., 2017), allowing wine businesses to integrate bias-adjusted climate predictions in their decision-making processes.

The climate predictions of precipitation included in this case study have been bias-adjusted with the variance inflation method (or calibration) and the reference dataset is the ERA5 reanalysis. The variance inflation method produces predictions that will have interannual variance that is equivalent to that of the reference dataset. The method is described in Doblas-Reyes et al. (2005), and has been tested against a simple bias correction for the case of wind speed in Torralba et al. (2017) and with five more methods for temperature and precipitation in Manzanas et al. (2019). The main advantage of this method is the inflation of the ensemble spread, which ensures that predictions have reliable probabilities.

The bias adjustment parameters are determined by comparing the past forecasts with the observational reference (ERA5 reanalysis in this case). The bias adjustment has been applied in cross-validation. Cross-validation adjusts a forecast for a particular year without using the information corresponding to that specific year. For example, when bias-adjusting a collection of 25 hindcasts runs (e.g. 1993-2018), the parameters used to calibrate a single year (e.g. 2005) are extracted from the other 26 years. This procedure is useful to emulate as closely as possible real-time forecast situations in which the observational reference is not available. The real-time forecast is calibrated with the full available hindcast period. The calibrated hindcasts are used to evaluate the skill of the forecast products. In the case of seasonal predictions, a fixed historical period of 26 years (1993-2018) is used for the real-time forecast calibration.

Forecasted precipitation quality assessment

The estimation of the forecast quality based on its past performance is a fundamental step in the prediction process, because it allows the quantification of the added value of that forecast relative to other prediction approaches (Doblas-Reyes et al., 2013). Moreover, the evaluation of the forecast quality is based on the comparison of the hindcasts and past observations. This evaluation of past forecasts can inform users about the expected performance of future forecasts (Weisheimer and Palmer, 2014).

A wide range of verification metrics are available in the literature (e.g. Jolliffe and Stephenson, 2012). For this case study different verification metrics have been selected to inform on the quality of the predictions, (i) the Ensemble-mean Correlation (EnsCor) for the ensemble mean prediction, (ii) the Fair Ranked Probability Skill Score (Fair RPSS) for tercile events, (iii) the Fair Continuous Ranked Probability Skill Score (Fair CRPSS) for the entire distribution and (iv) the Brier Skill Score for extreme events. A detailed description of these metrics can be found below.

It is important to highlight that each verification metric represents an aspect of the forecast and so the suitability of a prediction depends on the needs of the end-user. For instance, if the end-user only needs the probabilities of each tercile category from the prediction, the FRPSS shows an accurate picture about what the user will get when compared to climatology. Conversely, if the user needs all the values of the distribution, then the FCRPSS is the metric to look at when assessing the value of the prediction.

In this case study, the fair version of the RPSS, CRPSS and the BSS has been used as it compensates for the effect of the number of members on the score, since it rewards ensembles with members that behave as if they are sampled from the same distribution as the observational reference (Ferro et al., 2014).

The RPSS, CRPSS and the BSS range is from $-\infty$ to 1. Skill scores below 0 are defined as unskilful, those equal to 0 are equal to the climatology forecast, and anything above 0 is an improvement upon climatology, up to a score of 1, which indicates a "perfect" forecast. In the case study, these values are expressed as percentages, where a skill of 1 would equal to 100% skill.

(i) Ensemble-mean Correlation (EnsCor)

The Pearson correlation coefficient (Wilks, 2011) between the predicted ensemble-mean and the reference data set has been used as a measure of the linear correspondence between the retrospective predictions and the reference. The EnsCor correlation ranges between -1 and 1. If it is equal one there is a perfect association between the ensemble-mean of the predictions and the observations. When it is equal zero indicates that there is no association between the ensemble-mean of the predictions and the reference dataset, which in turn, shows that the ensemble-mean of the predictions does not provide any added value relative to the retrospective climatology. Values of EnsCor inferior to zero indicate that the observed climatology should be used instead of the predictions. A positive EnsCor value is the minimum requirement for seasonal predictions to have some potentially useful information because it depends not only on the potential predictability but also on the precise distribution of the data (Jolliffe and Stephenson 2012).

(ii) Ranked Probability Skill Score (RPSS)

A particular measure of the predictive skill for the probabilistic forecasts for tercile events is the Ranked Probability Skill Score (RPSS) (Epstein, 1969; Wilks, 2011), which is a squared measure comparing the cumulative probabilities of categorical forecast (i.e. terciles) and observational references relative to a climatological forecast strategy. The RPSS is based on the ranked probability score (RPS), which is a measure of the squared distance between the forecast and the observed cumulative probabilities. The RPSS shown in the DST has been computed based on categorical forecasts for terciles. The individual observations in the verification time series can fall in any of the three categories, with probability determined by the probability density function.

(iii) Continuous Ranked Probability Skill Score (CRPSS)

The Continuous Ranked Probability Skill Score (CRPSS) is a commonly used probabilistic skill score that allows the predictive skill assessment of the full probability distribution (Jolliffe and Stephenson 2012). It is based on the continuous ranked probability score (CRPS), a score that reduces to the mean absolute error if a deterministic forecast is used (Wilks 2011). The CRPS measures the difference between the predicted and observed cumulative distributions and it can be converted into a skill score (CRPSS), measuring the performance of a forecast relative to the climatology.

(iv) Brier Skill Score (BSS)

To provide an indication of the quality of the forecast probabilities under the 10th percentiles and those exceeding the 90th percentile, the Brier skill score (BSS) has been used. The BSS is based on the Brier score, which verifies whether the observed events that were forecasted with a specific probability have occurred or not, particularly events corresponding to the observations falling below the 10th percentile or above the 90th percentile. The smaller the Brier scores, the better the agreement between the predicted probabilities and the observed frequency of occurrence. The BSS indicates the degree of improvement of the Brier score of a particular extreme forecast compared to the Brier score of a climatological forecast