VINEYARDS INTEGRATED SMART CLIMATE APPLICATION

Deliverable 2.1: Seasonal forecasts quality assessment report

WP	2	Climatic & Agricultural Data	
Task	2.1	Seasonal forecast forecasts quality assessment	
Dissemination level[1]	PU	**Due delivery date**	30-04-18
Nature[2]	R	**Actual delivery date**	11-05-18

Lead beneficiary	BSC
Contributing beneficiaries	

Version	1.1
Total number of pages	23

[1] Dissemination level: **PU** = Public, **PP** = Restricted to other programme participants (including the JU), **RE** = Restricted to a group specified by the consortium, **CO** = Confidential, only for members of the consortium

[2] Nature of the deliverable: **R** = Report, **P** = Prototype, **D** = Demonstrator, **O** = Other

Version history

Document Version	Date	Author	Comments[3]
1.0	02-05-18	Raül MARCOS (BSC) ; Nube GONZÁLEZ-REVIRIEGO (BSC) ; Verónica TORRALBA (BSC); Albert SORET (BSC)	
1.1	10-05-18	María Navarro (MET)	Review

Deliverable abstract

In this task we have assessed the quality of two seasonal forecast systems (European Centre for Middle-range Weather Forecasting (ECMWF) System 4 and SEAS5), two bias correction approaches (calibration and simple bias correction) and a downscaling strategy (calibration). This has been achieved with a common set of deterministic and probabilistic metrics: the correlation of the ensemble-mean (deterministic), the fair Ranked Probability Skill Score (FRPSS, probabilistic), the fair Continuous Ranked Probability Skill Score (FCRPSS, probabilistic) and with reliability diagrams (probabilistic). These metrics are important for the wine partners because they provide information about different aspects of the predictions that contribute to increase the robustness of their decision-making processes at seasonal time-scales.

More specifically, we have verified monthly / 3-month average predictions of temperature at 2 m, maximum temperature, minimum temperature and precipitation from ECMWF System 4 (S4) and ECMWF System 5 (SEAS5) using the Japanese 55-year reanalysis (JRA-55). This validation has also been performed for the three VISCA demo-sites using the observations provided by the end-users: Raïmat (Codorniu), Quinta do Ataide (Symington) and Mirabella-Eclano (Mastroberardino). In addition, we have applied two bias-correction techniques, calibration and simple bias correction, to adjust the forecast statistical properties of the variables to the reference reanalysis. We have also evaluated the effect of these techniques on the skill and reliability of the predictions compared to their raw counterparts. Furthermore, we have selected the calibration approach as a first version of the downscaling to the demo-sites.

The results obtained show that there is some degree of predictability in the three demo-sites in different variables that can provide value beyond the customary use of climatology. Moreover, the verification / bias-correction / downscaling workflow developed in this task provides the basics of all future refinements that we will conduct during the remaining two years of the project, e. g. through the exploration of new seasonal prediction systems and/or downscaling techniques.

[3] Creation, modification, final version for evaluation, revised version following evaluation, final.

List of acronyms and abbreviations

BSC	Barcelona Supercomputing Center
ECMWF	European Centre for Middle-range Weather Forecasting
S4	ECMWF System 4
SEAS5	ECMWF System 5
JRA-55	Japanese 55-year reanalysis

Contents

1. Introduction

Agriculture, and in particular winegrowing, is a sector highly dependent on heat, sunlight and water, and therefore very sensitive to climate variability, extremes and impacts of climate change. One mitigation strategy consists on using seasonal predictions, e. g. from one to several months into the future, to modify the vegetative cycle of the grapevines to adapt for the climate anomalies of the upcoming months. This is one of the objectives of this VISCA project, which seeks to make European wine industries resilient to climate change, minimizing risks through an improvement of the production management. In order to achieve this objective, predictions from different time-scales (from weather to seasonal) will be integrated in a Decision Support System platform that will help the end users adopt better decisions.

However, although operational seasonal predictions are issued in a number of forecast services[4] [5] [6], winegrowing industry seldom benefits from them. Usually this is because General Circulation Models (GCMs) suffer from substantial systematic biases and have coarser resolutions than those demanded by the end-users. In fact, the wine industry decisions at these time-scales normally rely on climatology, a widespread behavior that might be linked to the tendency of decision makers to reduce the risk of losses by taking conservative decisions. This approach leads to a chronic departure from the optimal management and could have important economic impacts on their activity. Consequently, the need to demonstrate the benefits of seasonal predictions in the management of vineyards offers a window of opportunity to develop novel strategies to approach the problem.

Hence, the estimation of the seasonal forecast quality based on its past performance is a fundamental step towards the construction of climate services[7] to aid end-user decision-making, because it allows quantifying the forecast benefit relative to other prediction approaches. Thus, seasonal predictions have to be systematically compared to a reference (reanalysis or observations) to assess their overall quality in a multi- faceted process known as forecast quality assessment[8]. Three sources of uncertainty in common scoring metrics of probabilistic predictions should be considered: improper estimates of probabilities from small-sized ensembles, insufficient number of forecast cases, and imperfect reference values due to observation errors. A way to alleviate these problems is to use several scoring measures to offer a comprehensive picture of the forecast quality of the system[9] and to apply statistical inference as often as required. This information is also valuable to decide about the

[4] Shafiee-Jood, M., Cai, X., Chen, L., Liang, X.-Z. & Kumar, P. (2014), 'Assessing the value of seasonal climate forecast information through an end-to-end forecasting framework: Application to us 2012 drought in central illinois', *Water Resources Research* 50(8), 6592–6609.

[5] Weisheimer, A. & Palmer, T. (2014), 'On the reliability of seasonal climate forecasts', *Journal of The Royal Society Interface* 11(96), 20131162.

[6] Buontempo, C., Hewitt, C., Doblas-Reyes, F. & Dessai, S. (2014), 'Climate service development, delivery and use in Europe at monthly to inter-annual timescales', *Climate Risk Management* 6, 1–5.

[7] Doblas-Reyes, F. J., Garćıa-Serrano, J., Lienert, F., Biescas, A. P. & Rodrigues, L. R. (2013), 'Seasonal climate predictability and forecasting: status and prospects', *Wiley Interdisciplinary Reviews: Climate Change* 4(4), 245–268.

[8] Mason, S. J., and O. Baddour (2008), Statistical modelling, in Seasonal Climate: Forecasting and Managing Risk, edited by A. Troccoli et al., Springer, Dordrecht, pp. 167–206.

[9] Jolliffe, I., and D. Stephenson (20012). Forecast Verification: A Practitioner's Guide in Atmospheric Science. Wiley and Sons, 240 pp

application of the optimal bias correction and downscaling methods[10] . Hence, this quality assessment framework seeks to provide the end-users with the tools to understand which approaches could be better for their interests considering three different probabilistic metrics: fair ranked probability skill score (FRPSS), fair continuous ranked probability skill score (FCRPSS), reliability diagrams; and one deterministic: ensemble-mean correlation.

However, the production of statistically consistent and reliable predictions is a necessary condition for the elaboration or climate services[11] . To this aim bias correction and downscaling techniques are essential. Although these methods have been extensively used and critically assessed in climate change applications, their advantages and limitations in seasonal forecasting are not yet well understood and, thus, the importance of studying their effects on the raw predictions through the verification processes.

In our case, to reduce biases we have introduced the simple bias correction method (with a long tradition in seasonal forecasting; see, e.g. Barnston *et al.* 1994 [12]) and the calibration approach[11]. Regarding the spatial scale gap, we have started by using the calibration also as a downscaling adjustment method. These techniques can operate directly both on monthly and daily data and, thus, are suitable for applications demanding predicted series with high temporal and/or spatial resolution.

This deliverable is organized as follows. After this first introductory section there is the 'Data' section that is focused on the description of the different datasets; afterwards, the 'Methodology' presents the different methods used. Subsequently we introduce the 'Results' obtained and, finally, there is the 'Conclusions' section, which contains a summary of the outcomes and some reflections on the task performed.

[10] Ruffault, J. et al., 2013. Projecting future drought in Mediterranean forests: Bias correction of climate models matters! *Theoretical and Applied Climatology*, 117(1), pp.113–122.
[11] Torralba, V. *et al.*, 2017. Seasonal climate prediction: A new source of information for the management of wind energy resources. *Journal of Applied Meteorology and Climatology*, 56(5), pp.1231–1247.
[12] Barnston, A.G. *et al.*, 1994. Long-Lead Seasonal Forecasts—Where Do We Stand? *Bulletin of the American Meteorological Society*, 75(11), pp.2097–2114.

2. Data

2.1. Reference datasets

JRA-55 reanalysis

The JRA-55 reanalysis dataset[13] is the second global atmospheric reanalysis supplied by the Japanese Meteorological Agency (JMA). It ranges from 1958 to nowadays and it is updated also in real-time. It is the first global reanalysis that applies the four-dimensional variational analysis for the last half-century. Besides using the 4D-Var method, it improves the previous reanalysis, JRA-25[14], by different means like the inclusion of Variational Bias correction (VarBC) for satellite radiances, a new radiation scheme or the introduction of dynamic greenhouse concentrations. We have used monthly mean data for mean temperature, maximum temperature, minimum temperature and precipitation.

Demo-site observations

The end-users have provided observations for each of the demo-sites: Raïmat (Codorniu), Quinta do Ataide (Symington) and Mirabella-Eclano (Mastroberardino). All the time-series have been revised by the end-users and university partners to avoid any kind of inhomogeneity. The periods differ from site to site and from variable to variable, but in general they all have more than 20 years of daily data. Actually, they have provided us with daily data for 2m temperature, maximum temperature, minimum temperature and precipitation.

2.2. Prediction datasets

ECMWF System-4 (S4)

S4 seasonal prediction system[15] is a fully coupled general circulation model that provides operational multi-variable seasonal predictions in a real-time basis. In this study we focus on period 1981-2015. Last 35 years of predictions proceed from the combination of the 30-years hindcasts with the 5-years regular contemporary pool of predictions. All predictions have a minimum of 15-member ensemble (51 members for those which start dates are on 1st day of months February, May, August and November, and at every month since May 2011) and 7-months forecast horizon. The predictions used for the discussion are those initialized the 1st of November and the 1st of May. We have used monthly mean data for mean temperature, maximum temperature, minimum temperature and precipitation.

ECMWF System-5 (SEAS5)

SEAS5 is the fifth generation of ECMWF's seasonal forecasting system. It replaces the former ECMWF System 4 (S4) and uses the Integrated Forecast System, IFS, Cycle 43r1. The re-forecast of SEAS5 covers a 36-year period, from 1981 to 2016, with an ensemble of 25 members. Compared to the S4 it includes

---

13 Kobayashi, S. *et al.*, 2015. The JRA-55 Reanalysis: General Specifications and Basic Characteristics. *Journal of the Meteorological Society of Japan*. Ser. II, 93(1), pp.5–48.

14 Onogi, K. *et al.*, 2007. The JRA-25 Reanalysis. *Journal of the Meteorological Society of Japan*, 85(3), pp.369–432.

15 Molteni, F. *et al.*, 2011. The new ECMWF seasonal forecast system (System 4). ECMWF Technical Memorandum, 656(November), p.49.

a number of enhancements in the atmospheric resolution, land-surface initialisation and in the ocean model.

In the atmospheric component, the horizontal resolution has been increased from 80km to 36km with 91 vertical levels (the same as S4). Regarding the land-surface initialisation, the SEAS5 includes a new offline recalculation at the native atmospheric resolution with a revised precipitation forcing. Although this initialisation is still not perfect (the reanalysis and real-time assimilation are not the same), the tests performed show a good degree of consistency between the initialisation of SEAS5 re-forecast and real-time predictions. Finally, the SEAS5 uses the new version of ocean model NEMO (Nucleus for European Modelling of the Ocean), with and upgraded model version, ocean physics and resolution. More specifically the resolution has been increased from 1 degree and 42 layers in S4 to 0.25 degrees and 75 layers in SEAS5. The ocean and sea-ice initial conditions are provided by the new ocean analysis and reanalysis ensemble (ORAS5). In the Annex 1 there is the table that summarizes all these upgrades. We have used monthly mean data for mean temperature, maximum temperature, minimum temperature and precipitation.

3. Methodology

3.1. Bias adjustments and downscaling

To reduce the biases associated to the predictions, we have used two bias-adjustment approaches in cross-validation (calibration and simple bias correction) to correct the raw predictions obtained from S4 and SEAS5 considering the JRA-55 reanalysis for the period 1981-2015. Additionally, the calibration has been applied as a first version of downscaling to the demo-sites. This has been achieved by selecting the nearest point to each demo-site after applying the calibration method.

Simple bias correction

Simple bias correction is based on the assumption that both the reference and predicted distributions are well approximated by a Gaussian (normal) distribution, which is, most of the times, reasonable for monthly mean data. The adjustment produces an ensemble of predictions with the same mean and standard deviation as the reference dataset. This is a zero-order approach for correction of the systematic mean error that has been applied in the bibliography to correct temperature and precipitation[16]. The bias correction scheme can be summarized as,

$$y_{ij} = (x_{ij} - \overline{x})\frac{\sigma_{\mathrm{ref}}}{\sigma_e} + \overline{o}$$

where y_{ij} is the bias adjusted prediction of seasonal mean for each year *ith* and ensemble member *jth*; x_{ij} is the raw prediction for year *ith* and ensemble member *jth*; $\bar{x}$ is the seasonal mean obtained from the predictions; σ_{ref} is the interannual standard deviation of the reference dataset; σ_e is the interannual standard deviation of the ensemble members; $\bar{o}$ is the interannual climatological mean of the seasonal obtained from the reference dataset. This is done for each grid cell separately, resulting in a new forecast ensemble that has the same ensemble-mean and standard deviation as the reference.

Calibration

The calibration method can be considered as a way of obtaining predictions with an interannual variance that is equivalent to that of a reference dataset in a similar way to the simple bias correction method but ensuring, simultaneously, an increased reliability of the probabilistic predictions. The method is sometimes referred to as climate conserving recalibration after Weigel *et al.* (2009)[17]. Here we consider the variance inflation EMOS method introduced in Doblas-Reyes *et al.* 2005[18] and recently

[16] Leung, L.R. *et al.*, 1999. Simulations of the ENSO Hydroclimate Signals in the Pacific Northwest Columbia River Basin. *Bulletin of the American Meteorological Society*, 80(11), pp.2313–2329.

[17] Weigel, A.P., Liniger, M. a. & Appenzeller, C., 2009. Seasonal Ensemble Forecasts: Are Recalibrated Single Models Better than Multimodels? *Monthly Weather Review*, 137(4), pp.1460–1479.

[18] 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, Series A: Dynamic Meteorology and Oceanography*, 57(3), pp.234–252.

used by Torralba *et al.* 2017[11] to produce reliable seasonal predictions of wind speed. This calibration strategy has been selected because an inflation of the ensemble spread is required to obtain reliable probabilities. If x_i is the ensemble-mean prediction for any grid point at year *ith* and z_{ij} is the difference of the ensemble member *jth* from the ensemble-mean, at year *ith*; then, the calibrated estimate of the ensemble member *jth*, for a year *ith*, can be expressed as:

$$y_{ij} = \alpha x_i + \beta z_{ij}$$

The coefficients α and β are defined as follows:

$$\alpha = \mathrm{abs}(\rho)\frac{\sigma_{\mathrm{ref}}}{\sigma_{\mathrm{em}}}$$

$$\beta = (1-\rho^2)^{1/2}\frac{\sigma_{\mathrm{ref}}}{\sigma_e}$$

Where ρ is the correlation between the ensemble-mean of the retrospective predictions and the reference dataset; σ_{ref}, is the standard deviation of the reference; σ_{em} is the standard deviation of the ensemble-mean (the time series of x_i) and σ_e is the standard deviation of the ensemble.
The α and β coefficients are found under two constraints: the first is that the standard deviation of the inflated prediction is the same as that for the reference, and the second is that the predictable signal after the inflation is made equal to the correlation of the ensemble-mean with the reference dataset.

3.2. Verification

The quality of the monthly / 3-month average seasonal predictions of 2m temperature, maximum and minimum temperature and precipitation from S4 and SEAS5 have been assessed against both the JRA-55 reanalysis (1981-2015) and the observations at the three demo-site described in the data section. This forecast quality assessment has been achieved through the computation of four verification metrics: the Ensemble-mean Correlation (EnsCor, deterministic), the Fair Ranked Probability Skill Score (FRPSS, probabilistic), the Fair Continuous Ranked Probability Skill Score (FCRPSS, probabilistic) and reliability diagrams (probabilistic). This forecast quality assessment has been performed before and after the application of the bias adjustments techniques to evaluate their effect on the quality of the predictions.

Ensemble-mean Correlation (EnsCor)

The Pearson correlation coefficient[19] 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. This can be defined as:

$$r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 (y_i - \bar{y})^2}}$$

where x_i and y_i are, respectively, the observed and the ensemble-mean predicted values in each season, over the *i=1,2,...,n* years. The $\bar{x}$ and $\bar{y}$ are the average of the ensemble-mean predictions and the observations over the *n* years.

The EnsCor correlation ranges between -1 and 1. If $r_{xy} = 1$ there is a perfect association between the ensemble-mean of the predictions and the observations. When $r_{xy} = 0$ 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 ($r_{xy} < 0$) 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[9] .

Ranked Probability Skill Score (RPSS)

A comprehensive measure to evaluate the predictive skill of categorical events from probabilistic seasonal predictions is the ranked probability score (RPS)[19]. The RPS is the sum of the squared distance between the cumulative probabilities of the *n* predictions - reference pairs (for the entire interannual series) for *k* equiprobable forecast categories (e.g. tercile):

[19] Wilks, D.S., 2011. Statistical methods in the atmospheric sciences, London: Academic Press.

$$RPS = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\left[\left(\sum_{j=1}^{k} y_{i,j}\right) - \left(\sum_{j=1}^{k} x_{i,j}\right)\right]^2$$

where y_{ij} and x_{ij} are, respectively, the predicted and observed probabilities assigned by the *ith* forecast (i= 1,... n) to the *kth* category (i= 1,... k). The $x_{ij} = 1$ indicates that the observation is in category *k*, and $x_{ij} = 0$ otherwise.

The RPS is often expressed as a skill score (RPSS) because it allows assessing the prediction's added value relative to the climatology. The RPSS is given by:

$$RPSS = 1 - \frac{RPS}{RPS_{clim}}$$

RPSS ranges from −∞ to one. RPSS values below 0 are defined as unskillful, those equal to 0 indicate that the forecast provides similar information than the climatological forecast, and *RPSS > 0* shows that the predictions are better than the climatology. *RPSS = 1* corresponds to a 'perfect' forecast.

In this deliverable the RPSS has been computed for the verification of terciles (three equiprobable categories associated with the two terciles of the climatological distribution of the reference). The probabilities have been computed as the fraction of ensemble members in the corresponding category.

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[9] . 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 [19.] CRPSS can be expressed as:

$$CRPS = \int_{-\infty}^{\infty} [F(y) - F_0(y)]^2 \, dy$$

where $F(y)$ is the cumulative density function of the predictions and $F_0(y)$ is the cumulative step function that jumps from 0 to 1 at the point where the forecast variable (y) equals to the observation (x):

$$F_0 = \begin{cases} 0, y < x \\ 1, y \geq x \end{cases}$$

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:

$$CRPSS = 1 - \frac{CRPS}{CRPS_{clim}}$$

The CRPSS ranges between −∞ to one. CRPSS values below 0 are defined as unskillful, those equal to 0 indicate that the forecast is similar to the climatology forecast, and *CRPSS > 0* shows that the predictions are better than the climatology. *CRPSS = 1* indicates a 'perfect' forecast.

Fair Scores (FRPSS and FCRPSS)

Fair scores to ensemble predictions have been recently introduced[20] [21]. A skill score is fair when it favors predictions with ensemble members that perform as if they have been sampled from the same distribution than the reference dataset. The fair version of the RPSS (FRPSS) and CRPSS (FCRPSS) has been used in order to give an estimate of what the skill is when an infinite ensemble size is used (a measure of potential skill).

Reliability diagrams

Reliability diagrams are a common diagnostic of probabilistic predictions that assess both reliability and skill. They consist of a plot of the observed relative frequency against the predicted probability of a dichotomous event, providing a quick visual assessment of the impact of tuning probabilistic forecast systems. A perfectly reliable system should draw a line as closely as possible to the diagonal, within a certain measure of uncertainty.

The reliability diagrams have been used to evaluate the three forecast events defined by terciles. To draw a reliability diagram, discretization and grouping into probability bins of the probability predictions have to be done. The reliability diagram also includes information about the frequency of the forecast probabilities in each event, which is known as sharpness diagram. Sharpness is a property of the predictions that gives an indication of the variation in forecast probabilities issued by the prediction system, independently of the observations.

The information provided by the reliability diagram should be interpreted with care because even a perfectly reliable forecast system is not expected to have an exactly diagonal reliability diagram due to the limited samples typical of seasonal forecast systems[9] . To deal with this problem we have included consistency bars [22] in these diagrams. They indicate how likely the observed relative frequencies are, under the assumption that predicted probabilities are accurate.

[20] Fricker, T. E., C. A. T. Ferro, and D. B. Stephenson, 2013: Three recommendations for evaluating climate predictions. *Meteor. Appl.*, 20, 246–255, doi:10.1002/met.1409.
[21] Ferro, C. A. T., 2014: Fair scores for ensemble forecasts. *Quart. J. Roy. Meteor. Soc.*, 140, 1917–1923, doi:10.1002/qj.2270.
[22] Bröcker, J. and L. A. Smith, 2007: Scoring probabilistic forecasts: The importance of being proper. *Wea. Forecasting*, 22, 382–388, doi:10.1175/WAF966.1.

4. Results

In this section we present the forecast quality assessment of temperature at 2 meters height and precipitation averaged over July-August-September (JAS) obtained with seasonal predictions of ECMWF S4 and SEAS5 initialised on 1st of June for the whole Europe against JRA-55 reanalysis. Our analysis is focused on this region, variables and season because Europe is where the three demo-sites are located and, also, because JAS is important for the phenological management of the grapevines. However, it is important to remark that the study has also been conducted at global spatial scale and the results have been obtained for both monthly and 3-month averages with seasonal predictions initialised on 1st day of each month of the year. These results have been obtained before and after the application of each of the two bias adjustment methodologies (simple bias correction and calibration).

We begin by assessing the potential skill of the raw seasonal predictions from SEAS5 compared with the predictions from the previous S4 in Europe for the two variables using the ensemble mean correlation. Then we present the effect on SEAS5 reliability of the simple bias correction and calibration adjustments. After that, two probabilistic verification measurements, FCRRPSS and FRPSS, are analysed for the calibration approach. Finally, we present, for the three demo-sites, the results of 3 metrics (ensemble mean correlation, CFRPSS and FRPSS) both for the raw and downscaled forecasts.

4.1 Predictability improvements on SEAS5 compared to S4

Before the application of any bias adjustment technique, the potential skill of the raw predictions obtained from S4 and SEAS5 has been assessed and compared to quantify the quality gain of the newest version of the ECMWF prediction system, SEAS5. To illustrate this assessment, we have focused our attention on the EnsCor. The values for EnsCor of both, 2m temperature and precipitation, between the SEAS5 prediction system and JRA-55 are shown in figures 1a and 1c, respectively, for JAS over Europe. In the case of temperature (Fig. 1a), the predictions are rather good for the entire region with widespread areas showing correlations between 0.25 and 0.50 with a 95% confidence level. In fact, over the Atlantic, Eastern Europe and Scandinavia these correlations increase up to 0.50 - 0.75 (significant at 95% confidence level). Still, there are some areas in the Central Mediterranean where correlations drop to 0-0.25. Regarding precipitation (Fig. 1c), the EnsCor of SEAS5 is lower than for temperatures, especially in central and northeastern part of Europe, where values of EnsCor are below 0. In the Atlantic basin, the Iberian Peninsula, the North of Africa and Eastern Europe, the EnsCor values are positive ranging between 0 and 0.25 and, sometimes showing values above 0.25.

The differences between the EnsCor values obtained with the two forecast systems, SEAS5 and S4, against JRA-55 are shown in figures 1b and 1d. In case of 2m temperature (Fig. 1b), there are positive differences along the northern basin of the continent and, mainly, in the northern Scandinavian Peninsula, which indicates that potential skill of SEAS5 is higher than S4 in those areas.

This might be related to a better description of the atmospheric circulation processes because of the improvement of the spatial resolution in the SEAS5 system. As for precipitation (Fig. 1d) there is no consistent pattern to highlight, with scattered regions where the SEAS5 outperformed the S4 and vice versa.

This behaviour was expected because 2m temperature is a variable that has less uncertainty than precipitation in both seasonal forecast systems and reanalysis, which is reflected in the EnsCor. Considering that SEAS5 shows higher values of EnsCor in temperature than S4 (e.g. JAS), and for precipitation the EnsCor values are similar in both systems, from now onwards the results of this deliverable will be focused on the forecast quality assessment of SEAS5.

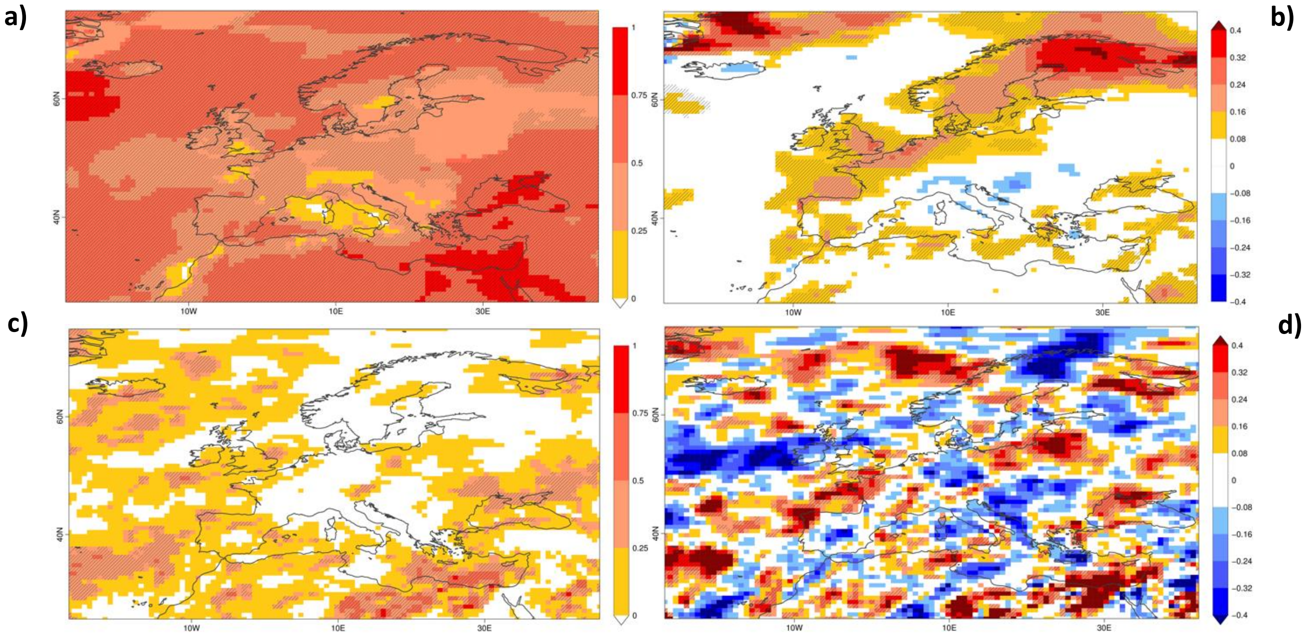

Figure 1. Left column: ensemble mean correlation between JRA55 and SEAS5 for JAS (start date on 1st of June) for (a) 2m temperature and (c) precipitation considering the period 1981-2015. Right column: differences in the ensemble mean correlation obtained from SEAS5 and S4 prediction systems against JRA55 for (b) 2m temperature and (d) precipitation, over JAS season (start date on 1st of June). Red highlights areas where SEAS5 outperforms S4. Areas showing significant correlations or significant differences in correlation (95% confidence level) are hatched.

4.2 Impact of the bias adjustments on reliability

Different methods of bias correction have been tested in this study to obtain predictions with improved statistical properties that can be used to provide a climate service for the wine sector. The impact of the bias adjustment techniques on the predictions' quality have been tested with several scoring measures. In this section, we illustrate the effect of the simple bias correction and calibration methods on the predictions from SEAS5 by the use of reliability diagrams.

Figure 2 depicts the SEAS5-JRA55 reliability diagrams for Europe and JAS season (start date on 1st of June) for 2m temperature (left column) and precipitation (right column). The first row contains the raw forecasts (Figs. 2a, b); the second corresponds to the predictions adjusted to the JRA55 with a simple bias correction (Fig. 2c, d); and, finally, the third, is associated with the predictions adjusted with the calibration approach (Fig. 2e, f). According to this figure, the two bias adjustments proposed deteriorate the reliability of the precipitation forecasts (especially for the upper and low tercile categories, Fig. 2d,f). As it has been discussed in previous section, seasonal forecasts of precipitation show limited predictability over Europe, thus the bias adjustments introduce some uncertainty that is translated into a reduction of the precipitation forecast quality. By contrast, the calibration correction does improve the reliability of the raw forecasts of 2m temperature, mainly for the upper and lower terciles. The simple bias correction, on the other hand, does not substantially improve the temperature reliability. This result is related with the definition of the two approaches. While the calibration adjusts the forecast probabilities to be more reliable, the simple bias correction only adjusts the mean and the standard deviation of the predictions. Hence, in the following section we will only present the results for the calibration adjustment.

That said, it is worth noting that the central category shows systematically worse reliability than the other two, both for temperature and precipitation. We might consider that these extreme terciles are often consequence of the additive / persistent nature of the anomalies in the boundary systems (they provide the seasonal predictability to the atmosphere). Thus, the model might find easier to predict these extremes than the central category, for even when there is no dominant anomaly in the boundary systems, the anomalies in temperature / precipitation might as well appear. This result has been previously documented[23] .

[23] Van Den Dool, H.M. & Toth, Z., 1991. Why Do Forecasts for “Near Normal” Often Fail? *Weather and Forecasting*, 6(1), pp.76–85.

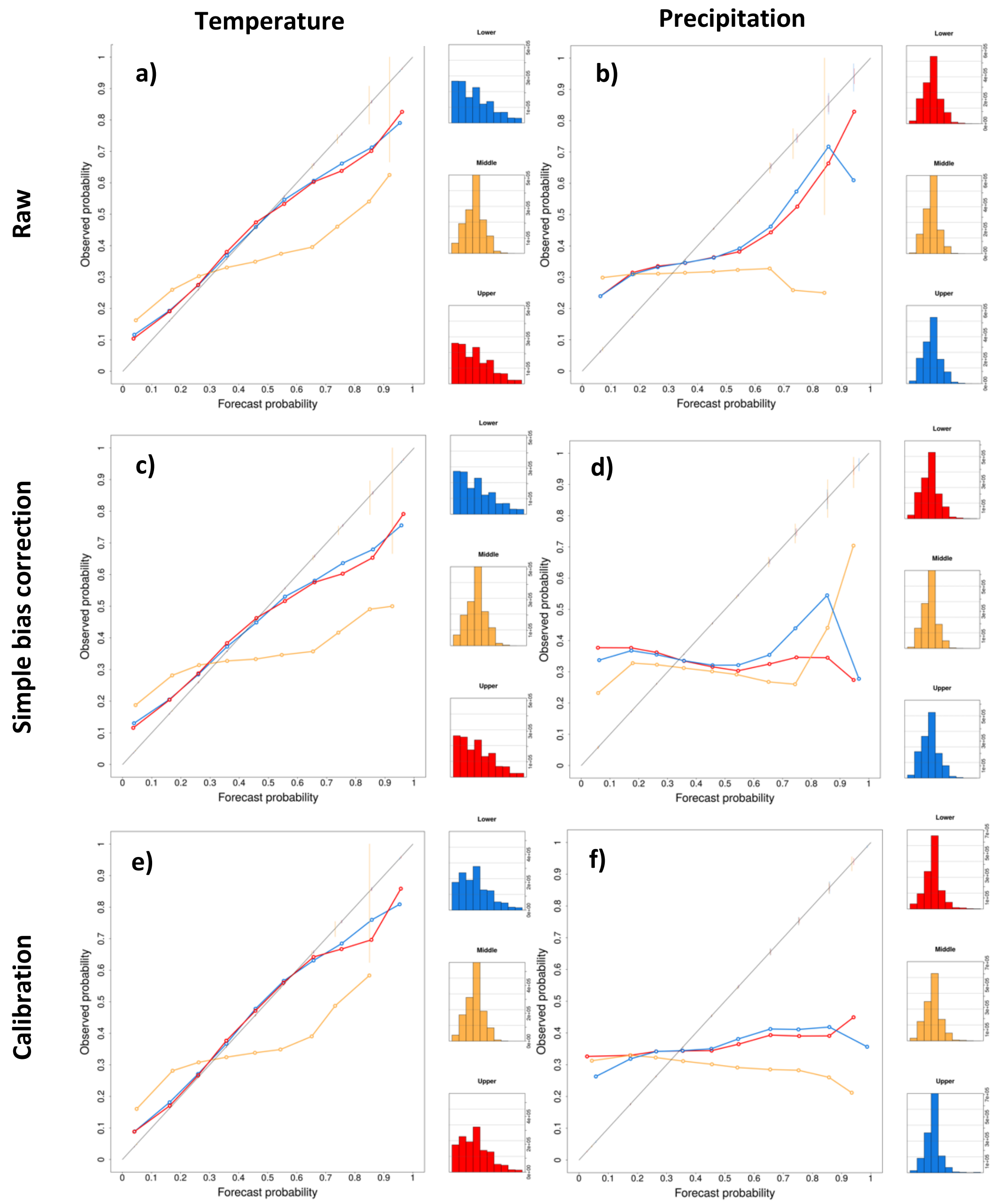

Figure 2. SEAS5-JRA55 reliability diagrams for Europe and JAS season (start date on the 1st of June) for 2m temperature (left column) and precipitation (right column) considering the period 1981-2015. (a-b) Raw forecasts, (c-d) simple bias correction and (e-f) calibration. Red refers to the above-normal category, yellow to the normal and blue, to the above normal.

4.3 Probabilistic skill scores of bias adjusted predictions

Seasonal predictions are probabilistic in nature, so once the predictions are bias adjusted a probabilistic forecast quality assessment has been performed. This evaluation aims to evaluate different aspects of the forecasts, for which more than one metric is recommended. In this work we have selected two different metrics: FRPSS and FCRPSS. FRPSS allows to measure the skill associated to the predictions of categorical events, that here have been defined by terciles. FCRPSS has been used to evaluate full probability density function of the forecasts.

Figure 3 depicts the FRPSS and FCRPSS for JAS season (start date on 1st of June) of SEAS5 calibrated with JRA55 reanalysis of 2m temperature (Fig. 3a, b) and precipitation (Fig. 3c, d) over Europe. In the temperature case, the FRPSS is higher than zero for most of the regions in Europe, which means that the use of seasonal predictions provides an added value with respect to climatology. More specifically, FRPSS lies between 0 and 0.25 in most of the continental Europe whereas over the Atlantic and the south-eastern part of Europe, the values increase up to 0.25-0.50. Regarding the FCRPSS, its spatial distribution is very similar to the FRPSS, though with lower values. Precipitation performance is rather poor as shown by the scattered patches in the FRPSS with values between 0 and 0.25. The situation for the FCRPSS is even worse, and scarcely any region has values higher than zero.

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.

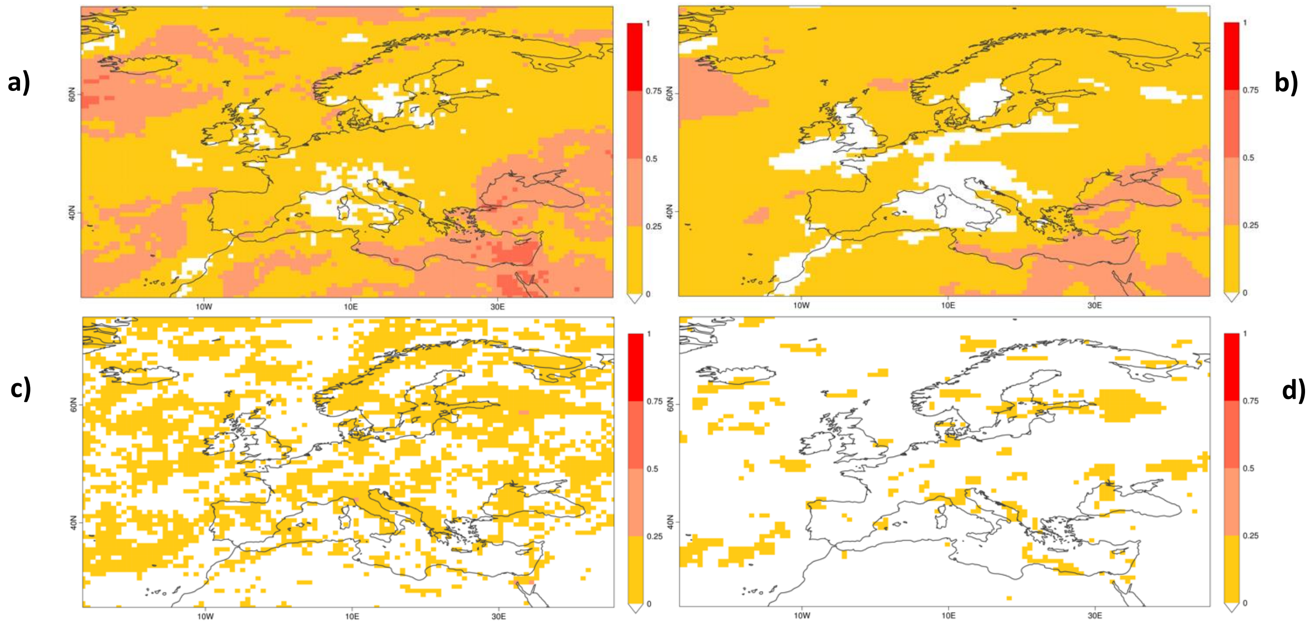

Figure 3. Left column: FRPSS of 2m temperature (a) and precipitation (c) obtained from SEAS5 calibrated with JRA-55 reanalysis for JAS season (start date on 1st of June) over Europe considering the period 1981-2015. Right column: Same as previous column but for FCRPSS.

4.4 Downscaling

To produce climate predictions able to satisfy specific users' needs, a first attempt to downscale the seasonal climate predictions to demo-site observations has been performed. The calibration method has been used as an order-zero downscaling approach. The idea is to take advantage of the increased resolution of the JRA55 reanalysis with respect to the SEAS5. Table 1 compares the verification metrics EnsCor, FRPSS and FCRPSS obtained with raw and calibrated predictions from SEAS5 against the demo-site observations.

The ensemble mean correlation shows that the calibration produces a slightly degraded correlation values for temperature and precipitation. The only exception is found for Mirabella-Eclano where the behaviour is somewhat the opposite. Regarding the FRPSS, the calibration worsens the precipitation score in the three sites, whereas it slightly improves the temperature counterparts in some cases. Finally, in the case of FCRPSS, the precipitation scores are maintained, whereas the temperature ones show some mixed results with Raïmat and Mirabella improving and a small deterioration for Quinta do Ataide. Besides, we can observe that temperature scores are generally better than precipitation and climatology (as indicated by the positive scores).

Table 1. Ensemble mean correlation, FRPSS and FCRPSS for raw and calibrated predictions for JAS (start date on 1st of June against each of the three demo-sites: Raïmat (Codorniu), Quinta do Ataide (Symington) and Mirabella-Eclano (Mastroberardino).

	Ens. Mean Correlation				FRPSS				CFRPSS			
	Raw		Downscaled		Raw		Downscaled		Raw		Downscaled	
	T_m	Prc	T_m	Prc	T_m	Prc	T_m	Prc	T_m	Prc	T_m	Prc
Raïmat	0.20	0.20	0.17	0.05	-0.09	0.04	-0.02	-0.01	-0.14	-0.46	-0.05	-0.46
Quinta do Ataide	0.45	0.21	0.43	-0.03	0.10	0.06	0.10	-0.01	0.08	-0.61	0.02	-0.60
Mirabella-Eclano	-0.37	-0.12	-0.32	0.12	-0.42	-0.07	-0.05	-0.08	-0.56	-0.53	-0.48	-0.52

5. Conclusions

In this work we have assessed the quality of two seasonal forecast systems (European Centre for Middle-range Weather Forecasting, ECMWF, System 4 and SEAS5), two bias correction approaches (calibration and simple bias correction) and a downscaling strategy (calibration). We have verified the monthly / 3-month average of seasonal predictions of 2m temperature, maximum temperature, minimum temperature and precipitation using the JRA-55 reanalysis and demo-site observations (we only showed precipitation and temperature in this deliverable). This analysis has been performed at both global and local scale by applying a common set of deterministic and probabilistic metrics: the Correlation of the Ensemble mean (EnsCor, deterministic), the Fair Ranked Probability Skill Score (FRPSS, probabilistic), the Fair Continuous Ranked Probability Skill Score (FCRPSS, probabilistic) and reliability diagrams (probabilistic). These metrics are important for the wine partners because they provide information about different aspects of the prediction that contribute to increase the robustness of their decision-making processes at these time-scales.

In this deliverable we have presented the results for JAS (1st of June start date) in Europe for 2m temperature and precipitation, but the analysis has been performed for the predictions initialized the 1st of each calendar month (not shown). Our analysis is focused on Europe because is in that region where the three demo-sites are located and we have illustrated our results for JAS as this season is important for the phenological management of the grapevines.

We have started by comparing the performance of the raw seasonal predictions from S4 and SEAS5 in Europe for the two variables using the ensemble mean correlation. This deterministic metric is useful to characterize the predictability of the new version of the system in comparison with the previous version. We have seen the SEAS5 shows some improvement with respect S4, particularly for the temperature. In the case of precipitation, the overall performance is similar in both systems (improving in some regions and worsening in others). This behaviour was expected because the temperature is a variable that has less uncertainty in both the seasonal forecasting systems and the reanalysis. Since temperature is an important variable in the phenological models and the precipitation performance is similar to the S4, we have decided to centre our attention on the SEAS5.

Bias adjustments are required when seasonal predictions are provided to the users, as these predictions are affected by systematic errors. The effect of the bias adjustments (simple bias correction and calibration) on the reliability of SEAS5 predictions has been also explored. Results of the bias correction on precipitation were rather negative in terms of reliability. Nevertheless, the calibration approach improved the reliability of SEAS5 predictions of temperature. These improvements in terms of reliability can be particularly helpful for those users who employ probabilistic information in their decision-making processes, as they need predicted probabilities in agreement with the observed relative frequencies of a particular event.

Results of the FRPSS and FCRPSS for the calibrated predictions reveal that temperature displays widespread better behaviour improving climatology in most areas than precipitation. Finally, we have summarised the results of three verification metrics (ensemble mean correlation, FRPSS and FCRPSS) for both raw and downscaled forecasts in the three demo-sites. We have observed that temperature scores are generally better than precipitation and, also, better than climatology in some cases. Regarding the calibration downscaling it slightly degrades the ensemble mean correlation as a consequence of the uncertainty introduced by the estimation of the parameters used in the calibration. However, the FRPSS shows some improvements for the seasonal predictions of temperature.

The results obtained show that there is some degree of skill in the three demo-sites that can provide value beyond the customary use of climatology. This is an important result for two reasons: the first one is because the end-users will have access to the bias adjusted seasonal predictions of temperature and precipitation together with their associated skill directly from the VISCA platform. This will allow the users to include this seasonal information in their decision making processes, benefiting from the added value these predictions offer compared to the common approach using climatology. The second reason is because these seasonal forecasts will be used to feed the phenological models to provide seasonal phenological information that might be used by the end-users to better control the management of the grapevines. Moreover, the verification / bias-correction / downscaling workflow developed in this task provides the basics to validate all the future refinements that we will conduct during the remaining two years of the project. More specifically it is our plan to assess different downscaling approaches such as perfect-prognosis analog downscaling; other bias correction adjustments, such as quantile-mapping; explore the performance of other prediction systems such as Météo-France System 5 or Met Office GloSea5; and building a multi-model. Finally, it is worth noting that in the final deliverable is our plan to use the annexes to expand the amount of results showed (among the thousands of verification charts obtained), to offer a complete view of the forecast quality assessment to report the full potential of the methodologies applied and developed during the project.

6. Annex

6.1. Upgrades in model and initialization of System5 in comparison to System4.

	S4	SEAS5
IFS Cycle	36r4	43r1
IFS horizontal resolution	T_L255	$T_{CO}319$
IFS Gaussian grid	N128 (80 km)	O320 (35 km)
IFS vertical resolution (TOA)	L91 (0.01 hPa)	L91 (0.01 hPa)
IFS model stochastic physics	3-lev SPPT and SPBS	3-lev SPPT and SPBS
Ocean model	NEMO v3.0	NEMO v3.4
Ocean horizontal resolution	ORCA 1.0	ORCA 0.25
Ocean vertical resolution	L42	L75
Sea ice model	Sampled climatology	LIM2
Atmosphere initialization (Re-forecast/Forecast)	ERA-Interim/Operations	ERA-Interim/Operations
Land Initialization (Re-forecast/Forecast)	ERA-Interim land (36r4)/Operations	ERA-Interim land (43r1)/Operations
Ocean initialization	ORA-S4	ORS-S5
Forecast ensemble size	51 (0-7m) 15 (8-13m)	51 (0-7m) 15 (8-13m)
Re-forecast years	30 (1981-2010)	36 (1981-2016)
Re-forecast ensemble size	15 (0-7m) 15 (8-13m)	25 (0-7m) 15 (8-13m)
Calibration period	1981-2010	1993-2016
Products Release Date	The 8th of each month at 12UTC	The 5th of each month at 12UTC