Skill assessment of a seasonal forecast model to predict drought events for water resource systems

Jaime Madrigal[a,*], Abel Solera[b], Sara Suárez-Almiñana[c], Javier Paredes-Arquiola[d], Joaquín Andreu[e], Sonia T. Sánchez-Quispe[f]

[a,] Research Institute of Water and Environmental Engineering (IIAMA), Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain. e-mail: jaime.madrigal.b@gmail.com*

[b] Research Institute of Water and Environmental Engineering (IIAMA), Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain. e-mail: asolera@upvnet.upv.es

[c] Research Institute of Water and Environmental Engineering (IIAMA), Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain. e-mail: sasual@upv.es

[d] Research Institute of Water and Environmental Engineering (IIAMA), Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain. e-mail: jparedea@hma.upv.es

[e] Research Institute of Water and Environmental Engineering (IIAMA), Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain. e-mail: ximoand@upvnet.upv.es

[f] Faculty of Civil Engineering, Universidad Michoacana de San Nicolás de Hidalgo, Francisco J. Múgica s/n, 58030, Morelia, Michoacán, México. e-mail: soniatsq@hotmail.com

Abstract

Droughts cause significant socio-economic and environmental impacts, so it has become an extremely important element in decision-making within water resource systems. For this reason, the research in this field has increased considerably over the last few decades. In order to be capable of making early decisions and reducing drought impacts, it is necessary to predict the occurrence of such events months or even years in advance. In this sense, various methods have been used to predict the occurrence of droughts. At present, seasonal forecast data can be used to forecast meteorological, hydrological, agricultural and operational droughts. However, the seasonal forecast data of these dynamical ocean-atmosphere coupled models must be analyzed in an exhaustive way, since it is known that these models may not adequately represent the climatic variability at river basin scale. Hence, this paper presents a new methodology for assessing the skill of a climate forecasting system in order to predict the occurrence of droughts by using contingency tables. The indices obtained from the contingency tables are necessary to perform the analysis of the predictive ability of the model in a semi-distributed way. All this taking into account the intensity of droughts using different scenarios based on the threshold below which it is considered to be in drought. Finally, a single value is obtained to determine the predictive ability of the forecasting model for the entire basin. The proposed methodology is applied to the Júcar river basin in Spain. It has been found that the analyzed forecast model shows better results than those obtained using an autoregressive model. Further work is needed to enhance climate forecasting from the perspective of water resources management, however, it should be mentioned that this type of data could be used for drought forecasting, allowing possible mitigation measures.

Keywords

Drought forecasting; forecast verification; contingency table; Jucar river basin.

1. Introduction

Over the last decades, drought events have been defined in a variety of ways, and some of the most common definitions are contained in Dracup et al. (1980), Tate & Gustard (2000) and Mishra & Singh (2010). However, drought can generally be defined as the reduction of water availability in a particular area for a specific period of time. There are several classifications of droughts, for example, Wilhite & Glantz (1985), classify droughts into meteorological, hydrologic, agricultural and socio-economic. Meteorological droughts can generally be defined as a period in which a particular number of days with rainfall less than a certain value (Great Britain Meteorological Office, 1951). This threshold below which a drought event is considered to occur can be the average precipitation value for the time scale analyzed (Hisdal & Tallaksen, 2000).

Droughts are a phenomenon that can be of varying magnitude, duration and intensity and can affect various sectors of society. In water resources management is important to be able to predict a possible drought event in order to have the capacity to make decisions that help minimize the damage of this phenomenon. The application of early mitigation measures is essential to reduce the socio-economic and environmental impacts of drought (Haro et al., 2014a). Drought forecasting is a critical component in risk management, drought preparedness and mitigation, and a major research challenge is to develop suitable techniques for forecasting the onset and termination points of droughts. One of the deficiencies in mitigating the effects of a drought is the inability to predict drought conditions accurately for months or years in advance (Mishra & Singh, 2011).

There are different methodologies in drought forecasting; regression models, time series models, probability models, neural networks models and hybrid models (Bacanli et al., 2009; Cancelliere & Salas, 2004; Fernández et al., 2009; Leilah & Al-Khateeb, 2005; Mishra et al., 2007; Morid et al., 2007). Nowadays, global circulation models and regional climate models are used to produce seasonal forecasts, which can be useful for drought forecasting. The Seasonal Forecast System model (System4) (Molteni et al., 2011), developed by the European Centre for Medium-Range Weather Forecasts (ECMWF), is a dynamical forecasting system. Which produces time series of 7 months from the first day of each month (Wetterhall & Giuseppe, 2017). System4 has been assessed for skill in predicting Asian summer monsoons (Kim et al., 2012b), Northern hemisphere winter (Kim et al., 2012a), below normal rainfall in the Horn of Africa (Dutra et al., 2013), drought forecasting in East Africa (Mwangi et al., 2013), global meteorological drought (Dutra et al., 2014) and for impacts analysis over East Africa (Ogutu et al., 2016). However, seasonal forecast must be analyzed from the point of view of water resources management. To analyze the predictive capacity of a model, several indices (skill scores) can be used, which have existed for more than a century (Peirce 1884, in Bartholmes et al., 2009). Not all indices are suitable for assessing a forecast system and there is no set of indices to obtain all the necessary information, which is why several sets of indicators are often used to cover a wider range of properties of the assessed model (Baldwin & Kain, 2004).

The aim of this paper is to propose a method of forecast verification for seasonal forecast systems by carrying out an assessment of the predictive capacity of the model in drought forecasting as an early warning for a water resources system with high exploitation rates and long lasting droughts. Data from System4 were analyzed against reference data, which in this case are precipitation data from the Spain02. v4 model (Herrera et al., 2012; 2016). In this paper, the data is evaluated by means of contingency tables, proposing a new aggregate index that can be easily used to evaluate the ability of seasonal forecast models to predict drought events. The proposed methodology is applied to the Júcar river basin in Spain.

The remainder of this paper is structured as follows. Section 2 describes the case study and the data used. Section 3 focuses on the proposed methodology. The results and a general discussion are provided in section 4. And finally, section 5 shows the conclusions.

2. Case study

For the analysis of the System4 precipitation data, was used the Júcar River Basin (JRB), located in the eastern of Spain. This basin is comprised of a total surface area of 22,261 km^2 and is the main of 9 water exploitation systems in the Júcar River Basin Demarcation (DHJ), (Haro et al., 2014b). Five sub-basins were considered for this study (see Fig. 1).

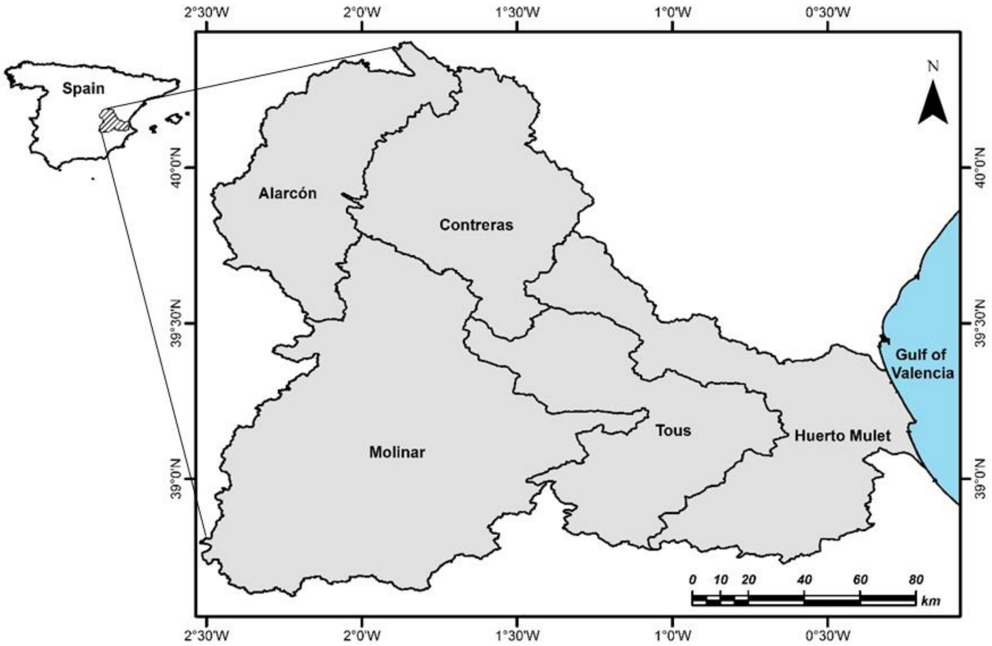

Fig. 1. Study area. Júcar River Basin.

The JRB has an average precipitation of 510 mm/year, and the average temperature is 13.6 °C. The natural resources reach 1,279 hm^3/year. On the other hand, the basin has a total water demand of 1,117 hm^3/year, of which, 88.6% is for irrigated agriculture, thus the basin has a water exploitation index (WEI) of 87% (Pedro-Monzonís et al., 2014).

The JRB is prone to drought events due to its semi-arid environment (Andreu et al., 2013), in the recent decades, four events have been recorded, which caused serious environmental damages and economic losses. The last events recorded are the historic droughts of 1983/84-1985/86, 1992/93-1995/96, 1997/98-2000/01 and 2004/05-2007/08 (Ministerio de Medio Ambiente, 2007).

To mitigate the impacts of these phenomena, the Ministry of Environmental of Spain (MMA) has worked extensively; making special plans of drought management, which consider three scenarios, normal and pre-alert, alert and emergency (Estrela & Vargas, 2008; Ferrer et al., 2008). In addition, the Júcar Basin Agency (JBA) has developed indices that allow determining the appropriate measure to deal with a drought event. These indices called Operative Drought Monitoring Indicators (SODMI), use real-time information provided by the Automatic Data Acquisition System of the DHJ (Estrela, 2006), the data contain information about precipitation and state of reservoirs, aquifers and rivers.

2.1 Forecast data

The forecast data used in this study come from the System4, developed by the ECMWF. These data are datasets of precipitation and temperature with a lead time of six months. The hindcasts (re-forecasts) start on the 1st of every month for the years 1981-2010, and the ensemble size is 15 members. The grid point calculations of the datasets are on the corresponding reduced N218 gaussian grid, which has about a 0.7 degrees spacing (Molteni et al., 2011). For each of the five sub-basins, two points were taken from the grid of the System4 model, in order to ensure that they were spread over the entire area of interest. Subsequently, the average of the two points was obtained to work with a single dataset per sub-basin. This approach was chosen because an interpolation produces time series with less dispersion than the original series. To obtain the historical time series of the forecast data, the first month of each time series was extracted (0-month lead time hindcast) and the fifteen scenarios generated by the System4 model were used (ensemble members).

2.2 Observed data

To contrast the forecast data, were used the dataset of Spain02 version v4 (Herrera et al., 2016), which are time series of precipitation and temperature in high-resolution grids on a daily scale. These data cover the Iberian Peninsula and the Balearic Islands. The grids correspond to standard grids of EURO-CORDEX: 0.44, 0.22 and 0.11 degrees (Herrera et al., 2012). For this study was selected the thinnest grid of Spain02 that is 0.11 degrees. Inside the datasets exist time series obtained with different interpolation methods, and for this work was used the dataset obtained with the Area-Averaged-monthly trivariate Thin Plate Splines method (AA-3D) (Herrera et al., 2016). Four points were taken in each sub-basin and averaged in order to obtain a representative time series for each sub-basin.

Although the data of the System4 model range from 01/01/1981 to 31/12/2015, the analysis only was performed with the period between hydrological years 1981/82 and 2005/06 (25 years), due to the data range chosen from Spain02, which goes from 01/01/1970 to 31/12/2006. Once the time series were obtained for each sub-basin, they were analyzed on a monthly scale.

3. Methodology

3.1 Statistical analysis

In the evaluation of hydrological models, indices are used to determine their capacity to reproduce reality. Among the most commonly used indicators are Nash-Sutcliffe efficiency (NSE) and the Modified Kling-Gupta Efficiency (KGEM) (Kling et al., 2012; Moriasi et al., 2007; Spalding-fecher et al., 2016). Therefore, a first evaluation of the System4 model data was made, obtaining the NSE and KGEM indicators.

The Nash-Sutcliffe efficiency (Eq. 1) is a metric that determines the relative magnitude of variance of data modeled with observed variance (Nash & Sutcliffe, 1970).

$$NSE = 1 - \left[\frac{\sum_{n=1}^{N}(O_n - F_n)^2}{\sum_{n=1}^{N}(O_n - \overline{O_n})^2}\right] \quad (1)$$

Where N is the total number of time-steps, O_n is the forecasted value at time-step n, F_n is the observed value at time-step n, and $\overline{O_n}$ is the mean of the observed values.

The Modified Kling-Gupta Efficiency (Eq. 2), as well as the NSE, it has a range of -Inf to 1 and its optimal value is 1.

$$KGEM = 1 - \sqrt{(r-1)^2 + (\beta - 1)^2 + (\gamma - 1)^2} \quad (2)$$

Where *r* is the correlation coefficient between forecasted ($_F$) and observed ($_O$) data, β is the ratio of means (μ_F/μ_O) and γ is the ratio of coefficients of variation (CV_F/CV_O). The ideal value of each of the three components is 1 (Gupta et al., 2009; Kling et al., 2012).

In order to obtain the indicators described above, an accumulation of data from the 15 System4 ensemble members was carried out. As these time series are equiprobable, it was decided to work with a unique dataset that includes the characteristics of each forecast. Thus, a single time series that groups the 15 ensemble members was obtained, placing each time series at the end of the previous one. That means a total of 4500 precipitation data on a monthly scale. On the other hand, the time series of observed data was repeated 15 times, so that both data, observed and forecast, have the same length. These aggregate time series of both, were also used for the forecast verification. In addition, the NSE and the KGEM indicators were obtained for the ensemble, in order to analyze the performance of the mean of the 15 ensemble members.

3.2 Forecast verification using a contingency table

Since what is desired is to assess the predictive capacity of the drought event as an early alert, an analysis should be carried out from the perspective of the occurrence of this phenomenon. In the management of water resources it is very important to predict droughts in order to adopt measures to reduce the impacts of these events (Haro et al., 2014a). Thus, the correct prediction of a drought event is more important than the correct amount of the runoff when the prediction is optimal. For this reason, it was decided to analyze the predictive capacity of drought episodes.

Droughts have three main characteristics: intensity, duration and surface area affected (Wilhite, 2000). Intensity refers to the decrease in precipitation and the impacts that this decrease can cause, and it is measured by applying the Palmer drought severity index or through a threshold, which can be a percentage of the mean precipitation and can be arbitrarily selected. Duration is the period of time that precipitation is below the set threshold. For the development of the analysis of this work, five drought scenarios were established, which correspond to 20% (S1), 40% (S2), 60% (S3), 80% (S4) and 100% (S5) of the monthly means. In each of them, the event occurs when the precipitation value for each month is less than the corresponding threshold value. In this way, the measurement of drought intensity is also included in the analysis.

The scenarios were analyzed using 2x2 contingency tables for dichotomous events (Bartholmes et al., 2009), which allow the predictive capacity of the model to be assessed based on meteorological droughts. These type of droughts occur when the monthly precipitation is below the mean precipitation value of the corresponding month (Hisdal & Tallaksen, 2000). In this way, a time series of discrete nonprobabilistic values were obtained for each scenario, since each value can only take a value of four possibilities, *Hit*, *Miss*, *False alarm* and *Correct negative*. Table 1 shows a contingency table.

Table 1. Contingency table. (Wilks, 2006).

		Spain02.v4	
		YES	NO
System4	YES	*Hits (a)*	*False alarms (b)*
	NO	*Misses (c)*	*Correct negatives (d)*

Where, the *Hits (a)* represent the coincidence of drought of both series, the *Misses (c)* correspond to the presence of a drought in the data of Spain02 and the absence of this event in the data of System4. The *False alarms (b)* are presented when in the time series of Spain02 there is no drought and if there is in System4, and finally, the *Correct negatives (d)* are the months in which there is no drought in both models.

From the contingency table, it can be concluded that a perfect forecast is presented when the value of *False alarms (b)* and *Misses (c)* is equal to zero. However, given that an actual forecast is imperfect, some metrics are required to determine the degree of correspondence between the predicted value and the observed value, thus obtaining, different characteristics of the predicted time series (Wilks, 2006). There are a large number of metrics developed for model verification using 2x2 contingency tables (Mason, 2003; Murphy & Daan, 1985).

In this paper, seven different scores were used in order to assess the forecasting system. The Proportion Correct score (PC) shown in Eq. 3, proposed by Finley (1884), since this score determines the correct portion of the time series and it is a widely used metric, being the most direct and intuitive. However, this score does not differentiate between *Hits* and *Correct negatives*, so it is necessary to use other scores.

The Threat Score (TS) is a very useful metric when the event to predict (drought) occurs less frequently than the non-occurrence is (Eq. 4). As well as the PC, the TS has the worst possible value of 0 and 1 as the best possible value. On the other hand, the Bias score (BIAS) is the comparison of the mean forecast with the mean observation, as evident in Eq. (5). The BIAS score is not a precision indicator since it does not provide information on the correspondence between forecasts and observations when is dealing with mean values. A value below 1 indicates that the drought event was less frequently predicted than observed, while a value above 1 indicates that the event was over-expected. BIAS ranges from zero to infinite and the best expected value is 1, which would indicate that the predicted time series is unbiased.

The reliability of the model can be evaluated by using the False Alarm Ratio (FAR), which represents the *yes* portion of the System4 model that is wrong, as shown in Eq. (6), so a better expected value is zero and the worst is 1. On the other hand, the Success Ratio (SR) provides information on the probability that an observed event will be predicted (Eq. 7). This score is sensitive to *False alarms* and ignores *Misses*. This metric can also be represented as 1-FAR, in other words, it is complementary to the False Alarm Ratio.

The Probability Of Detection score (POD) is the ratio between the correct forecasts and the number of times this event has occurred (Eq. 8). On the other hand, the Probability of False Detection score (POFD) is the ratio of *False alarms* to the total number of *no* drought events (Eq. 9).

$$PC = \frac{hits + \ correct\ negatives}{n} \quad (3)$$

$$TS = \frac{hits}{hits + misses + false\ alarms} \quad (4)$$

$$BIAS = \frac{hits + false\ alarms}{hits + misses} \quad (5)$$

$$FAR = \frac{false\ alarms}{hits + false\ alarms} \quad (6)$$

$$SR = \frac{hits}{hits + false\ alarms} \quad (7)$$

$$POD = \frac{hits}{hits + misses} \quad (8)$$

$$POFD = \frac{false\ alarms}{correct\ negatives + false\ alarms} \quad (9)$$

In order to obtain an analysis with the obtained indices, the performance diagram is used (Roebber, 2009), where BIAS, SR (1-FAR), POD and TS are evaluated.

3.3 Contingency table modified

Despite the different indices used, there is information that cannot be collected by them since it should not only be seen as an analysis of the possible occurrence, or not, of the drought event. Consideration should also be given to the possibility that the event may be predicted early or late, in addition to the permanence of the drought over the months. In order to collect these possibilities of dichotomous events, the creation of a 3x3 contingency table is proposed, leaving it as follows:

Table 2. Contingency table modified.

		Spain02.v4		
		DROUGHT START	DROUGHT STAY	NO DROUGHT
	DROUGHT START	*Drought start hit (e)*	*Late start (f)*	*False start (g)*
System4	DROUGHT STAY	*Early start (h)*	*Drought stay hit (i)*	*False stay (j)*
	NO DROUGHT	*False no drought (k)*	*Early exit (l)*	*No drought hit (m)*

When both time series present the start of a drought event in the evaluated month, the discrete variable corresponds to a *Drought start hit (e)*, but if an onset drought event occurs in System4 and in Spain02 there is a drought initiated prior to the month evaluated, a *Late start (f)* is present. On the other hand, when in the System4 time series there is a start of a drought and Spain02 does not present drought, a *False start (g)* is taken.

When in the Spain02 time series there is the onset drought and in the System4 time series there is a drought that started earlier, the value of the variable is an *Early start (h)*, but if both time series have drought in the month evaluated, but in neither case is the start of said event, a *Drought stay hit (i)* is held. Whereas, if in System4 there is a drought initiated prior to the month evaluated while in the Spain02 data there is no drought, the variable is a *False stay (j)*.

When Spain02 presents a start of a drought while in the System4 there is no drought, the variable is a *False no drought (k)*. If the Spain02 time series is in a drought that started before the month evaluated and the System4 time series does not have a drought, an *Early Exit (l)* is presented, and when neither time series has a drought, the discrete variable takes the value of a *No drought hit (m)*.

However, since the metrics mentioned above have been developed for 2x2 contingency tables,

Table 2 must be transformed from 3x3 to 2x2. To this end, the procedure proposed in Wilks (2006) should be used. For example, if the goal is to work with the event *Drought start hit (e)*, Table 2 is modified, as follows:

Table 3. Contingency table for the event Drought start hit.

		Spain02.v4	
		YES	NO
System4	YES	(a') = (e)	(b') = (f)+(g)
	NO	(c') = (h)+(k)	(d') = (i)+(j)+(l)+(m)

Thus, for any event to be evaluated, it will take the place of a *Hit*, the sum of the two remaining values of the row of the evaluated event corresponds to a *False alarm*, while the sum of the two values of the column where the studied event is located corresponds to a *Miss.* Finally, the sum of the rest of the values will be a *Correct negative*.

Since the most important event in this analysis is the occurrence of the drought event and its onset, the events *Drought hit (e)*, *Early start (h)* and *Late start (f)* are evaluated.

3.4 Assessment of an aggregate index

The analysis described above provides a large number of values that are complex to analyze, therefore it is necessary to reduce the number of indices in order to obtain a unique value capable of showing the drought forecasting skill of the forecast system model in a simplified way.

A first approximation is to average the value of each scenario for each index and for each event. It is important to take into account the water exploitation index (WEI) of the river basin, since when the availability is greater than the WEI, the drought is not worrying, as the volume of resources is greater than the demand. Therefore, in order to obtain the average of the scenarios, only those below the WEI should be considered. Even after having averaged the values, a large number of parameters have to be analyzed independently. Since, the four possibilities that each discrete value can obtain (Hits, False alarms, Misses and Correct negatives) the one that is less interesting is the Correct negatives, an average of the TS and POD indices is obtained in order to obtain an aggregate index. Given that, the TS indicates the portion of Hits of the drought event with respect to the presence of this phenomenon in both series and the POD shows the portion of Hits with respect to the presence of the drought in the observed data. Thus, the number of indices is reduced to one per sub-basin and per event.

Moreover, it should be considered that the drought phenomenon also has a spatial dimension. To take this property into account and also to obtain a single value for the entire basin, it is proposed to obtain an overall index as the weighted average of the 5 series, the annual mean precipitation is proposed as a weighting parameter.

Thus, an aggregate index is obtained for the whole basin and for each particular event, considering the intensity and spatial variability. The aggregate index obtained was contrasted with the Relative Operating Characteristic (ROC). The ROC diagram plots the Probability Of Detection score (POD) and the Probability of False Detection score (POFD). This method is widely used for its ability to graph several thresholds in a single diagram. Nevertheless, it can be convenient to summarize a ROC diagram using a single scalar value, and the usual choice for this purpose is the area under the ROC curve. As ROC curves for perfect forecasts pass through the upper-left corner, the area under a perfect ROC curve includes the entire unit square (the perfect area is equal to 1) and the random forecasts lie along the 45° diagonal of the unit square, the area under the ROC of interest is 2A-1. Where A is the area under the curve obtained (Wilks, 2006).

4. Results and discussion

In this section, spatio-temporal droughts from 15 ensemble member of the model System4 from 1981 to 2006 was compared to the reference data set (Spain02.v4). After a precipitation analysis, the predictive capacity of droughts is analyzed using modified contingency tables in order to obtain an aggregate index for the entire basin. Seven indices were assessed in the analysis (see Fig. 9 and Fig. 10)

4.1 Precipitation

For the comparison of the monthly mean precipitation of the System4 model with respect to the Spain02 model, the ensemble of the 15 scenarios of the forecast time series was used. Figure 2 shows the monthly mean of both models for the five sub-basins analyzed.

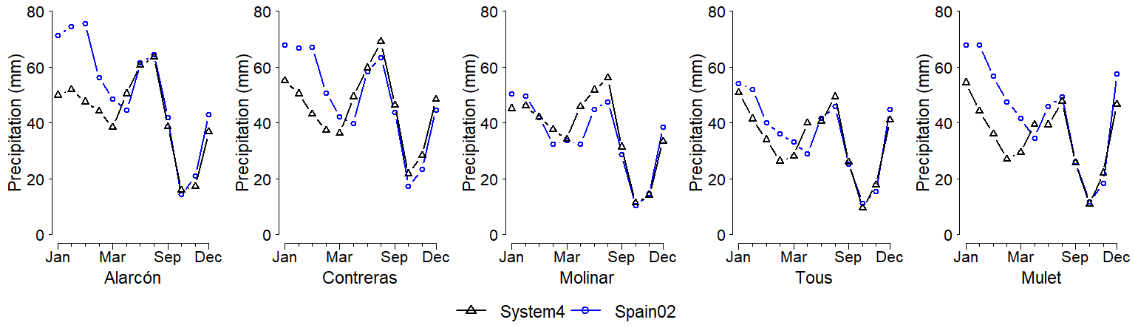

Fig. 2. Monthly mean precipitation.

As can be seen in the previous figure, there is a high correlation between both models from March to September (months with the least precipitation) in the five sub-basins. However, there are notable differences from October to February, when the rainfall is greater. Despite this, in all cases a similar trend of the System4 model data is reproduced with respect to the Spain02 model. The annual precipitation of the System4 model also presents the same trend as the Spain02 model, as can be seen in Figure 3. For this reason, the information can be used, since it is possible to correct differences by applying some method of bias correction.

The data of the ensemble collect the mean values of the 15 System4 model scenarios, however, this time series does not contain the noise presented by each of these scenarios, as can be seen in Figure 3. Therefore, it has been decided to work with the fifteen time series on a monthly scale, obtaining their statistics as a single series.

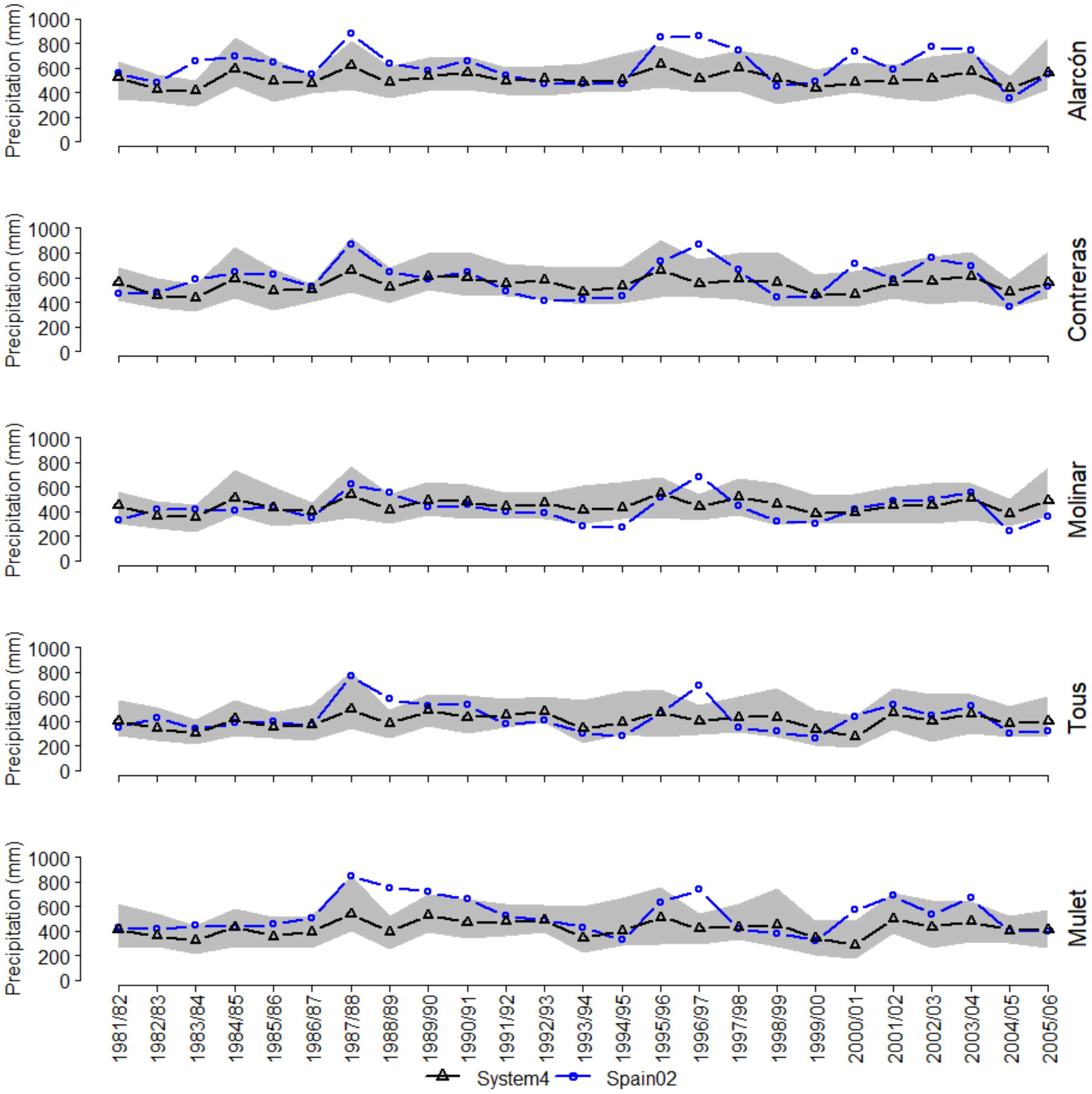

Fig. 3. Annual precipitation. The shaded surface represents the variation of the 15 ensemble members.

Once the single time series containing the data of the 15 System4 model scenarios has been obtained, it is contrasted with the observed time series and the Nash-Sutcliffe and Kling-Gupta indices are obtained (see Fig. 4 and Fig. 5).

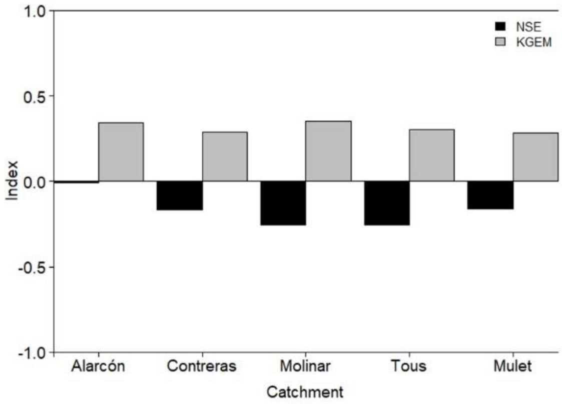

Fig. 4. Nash-Sutcliffe Efficiency and Kling-Gupta Efficiency Modified for the monthly precipitation data.

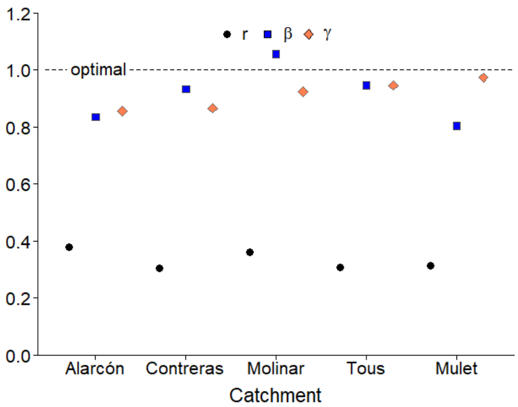

Fig. 5. Kling-Gupta Efficiency Modified elements.

Figure 4 shows low NSE values, between -0.26 and -0.01, while KGEM values are around 0.4. When analyzing the elements of the KGEM (see Fig. 5), it can be seen that the monthly data of the five sub-basins show a relationship close to 1, both for the mean ratios (bias ratio) and the coefficients of variation. On the other hand, the Pearson correlation coefficient is between 0.31 for the Contreras sub-basin and 0.38 in the Alarcón sub-basin. However, in the case of the ensemble, the analysis shows in an NSE between 0.20 for Mulet and 0.27 for Tous, a KGEM between 0.26 for Mulet and 0.33 for Tous, and a Pearson correlation coefficient between 0.48 for Contreras and 0.57 for Alarcón. This indicates that the ensemble mean presents a better similarity to the observed data with respect to the 15 ensemble members. The reason for this similarity is that the ensemble improves the average of the 15 time-series, but the noise of these series is lost. As in the seasonal management of water resources, the noise of the

15 ensemble members is more important than their average, it is important to work with all the series and not only with the ensemble.

4.2 Droughts

Since the time series of occurrence or absence of drought event is a time series of dichotomous data, it can be analyzed using a 2x2 contingency table. The indices obtained are shown in the following tables:

Table 4. Indices obtained from the 2x2 contingency tables for Alarcón, Contreras and Molinar sub-basins.

Catchment	Alarcón					Contreras					Molinar					Perfect
Scenario	*S1*	*S2*	*S3*	*S4*	*S5*	*S1*	*S2*	*S3*	*S4*	*S5*	*S1*	*S2*	*S3*	*S4*	*S5*	*Score*
PC	0.86	0.72	0.64	0.62	0.61	0.86	0.72	0.64	0.61	0.60	0.83	0.71	0.63	0.59	0.60	1
BIAS	0.67	0.76	0.89	1.00	0.99	0.50	0.67	0.90	0.97	1.00	0.61	0.82	0.95	1.04	1.00	1
POD	0.18	0.31	0.43	0.58	0.66	0.13	0.28	0.43	0.56	0.65	0.19	0.33	0.45	0.56	0.65	1
TS	0.12	0.21	0.30	0.40	0.50	0.09	0.20	0.29	0.40	0.48	0.13	0.22	0.31	0.38	0.48	1
SR	0.27	0.40	0.49	0.58	0.67	0.25	0.41	0.48	0.58	0.65	0.31	0.40	0.48	0.54	0.65	1
FAR	0.73	0.60	0.51	0.43	0.33	0.75	0.59	0.52	0.42	0.35	0.69	0.60	0.52	0.46	0.35	0
POFD	0.06	0.14	0.25	0.35	0.45	0.05	0.13	0.25	0.35	0.47	0.07	0.16	0.28	0.39	0.48	0

Table 5. Indices obtained from the 2x2 contingency tables for Mulet and Tous sub-basins.

Catchment	Mulet					Tous					Perfect
Scenario	*S1*	*S2*	*S3*	*S4*	*S5*	*S1*	*S2*	*S3*	*S4*	*S5*	Score
PC	0.84	0.66	0.59	0.57	0.59	0.81	0.67	0.60	0.57	0.57	1
BIAS	1.12	0.98	0.95	0.96	1.01	0.73	0.91	0.93	0.97	1.05	1
POD	0.23	0.35	0.48	0.58	0.68	0.19	0.36	0.48	0.58	0.66	1
TS	0.12	0.22	0.32	0.42	0.51	0.12	0.23	0.33	0.41	0.48	1
SR	0.21	0.36	0.50	0.60	0.67	0.26	0.39	0.51	0.59	0.63	1
FAR	0.79	0.64	0.50	0.40	0.33	0.74	0.61	0.49	0.41	0.37	0
POFD	0.10	0.23	0.33	0.44	0.54	0.09	0.21	0.32	0.44	0.56	0

Table 4 and 5 show that the correct portion (PC) of the analyzed time series is between 86% and 81% for scenario S1, while for scenario S5, the correct portion is between 61% and 57%. In reducing the threshold of droughts, the sub-basin that presented the greatest difference between the PC values for the different scenarios was Mulet, where there was a difference of 27% between the S1 and S5 scenarios.

In general, BIAS shows that the time series of all sub-basins are unbiased or slightly biased. Light over-forecast was obtained for the Mulet-S1 and Tous-S5 scenarios. On the other hand, the greatest biases were obtained in Alarcón, Contreras, Molinar and Tous (all on scenario S1), in addition to Alarcón-S2 and Contreras-S2.

The analysis of the Probability Of False Detection (POFD) and Probability Of Detection (POD) is very important because together they form the conceptual basis for the signal detection approach to verify probability forecasts (Wilks, 2006). The POFD obtained shows that the evaluated data have a *False alarm* percentage with respect to the non-occurrence of the drought event of less than 25% for thresholds S1, S2 and S3, except for the Mulet and Tous sub-basins. In addition, in the previous sub-basins in scenario S3, it is slightly higher than 30%. However, for the rest of the scenarios, the percentage of *False alarms* with respect to non-occurrence of drought fluctuates between 35% and 56%. The POD indicates that, despite having obtained high PC values, the percentage of *Hits* with respect to the total occurrence of the event is low, reaching a maximum of 68% for scenario S5 of the Mulet sub-basin.

The SR and FAR denote the percentage of *Hits* and *False alarms* with respect to the "yes" of the predicted data, respectively. The values found from SR and FAR indicate that the forecast model tends to over-estimate the drought event when considering low thresholds.

Figure 6 shows the performance diagram obtained, where can be seen that for the thresholds of 100% (S5) the results are better, since the optimum forecast is in the upper right part of the diagram. Figure 7 shows the ROC diagram, where it is possible to ascertain what is obtained in the performance diagram, that is, the predictive capacity of the model improves when droughts are less intense (S5). However, this predictive capacity is very low, since the area of interest under the curve is in the range of 0.18 in the Tous sub-basin and 0.27 in the Alarcón sub-basin.

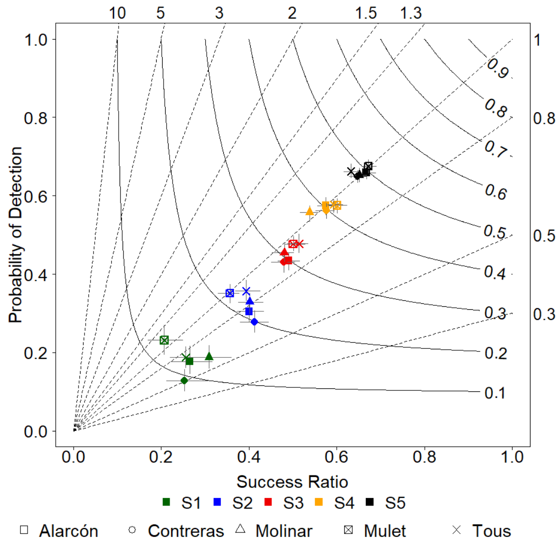

Fig. 6. Performance diagram.

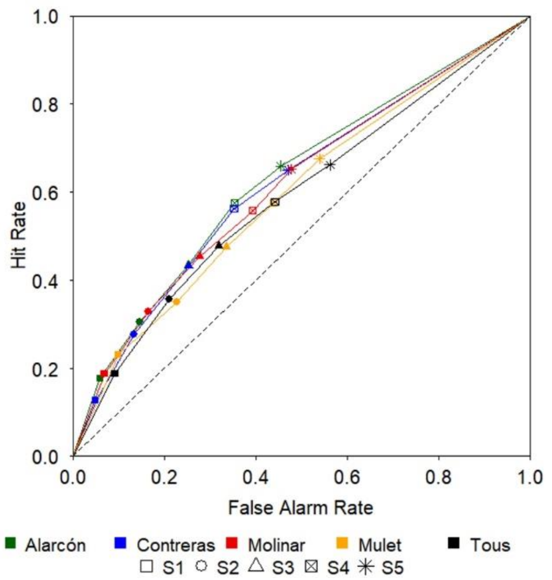

Fig. 7. ROC diagram.

In order to study the coincidences of the System4 model data with respect to the Spain02 model, the 2x2 contingency table is extended to a 3x3 table, with emphasis on the concordances at the onset of each event, these are, *Drought start hit*, *Early start* and *Late start* events. Additionally, the event *Drought stay hit* was analyzed, in order to explore the behavior of the predictions of the System4 model once it has entered a period of drought. Once the 3x3 contingency table has been obtained, it is reduced to 2x2 since the calculated indices have been developed for 2x2 contingency tables. Table 6 shows as an example the transformation of the contingency table from 3x3 to 2x2 for the Alarcón sub-basin and for the *Drought hit* event corresponding to scenario S1.

Table 6. Modified contingency tables for the Alarcón sub-basin, corresponding to scenario S1. Left, 3x3 table. Right, 2x2 table for the Drought start hit event.

		Spain02.v4		
		DROUGHT START	DROUGHT STAY	NO DROUGHT
System4	DROUGHT START	*63 (e)*	*11 (f)*	*207 (g)*
	DROUGHT STAY	*7 (h)*	*4 (i)*	*28 (j)*
	NO DROUGHT	*335 (k)*	*60 (l)*	*3785 (m)*

		Spain02.v4	
		DROUGHT START	NO DROUGHT START
System4	DROUGHT START	*63 (e)*	*218 (f+g)*
	NO DROUGHT START	*342 (h+k)*	*3877 (i+j+l+m)*

Figure 8 shows the performance diagrams of the four events analyzed, where it can be seen that, for all sub-basins, the climate model gives the best results for the S4 and S5 scenarios, while scenario S1 has the worst results. In Figure 8 it can also be seen that the results of the four events are unbiased for the S4 scenario, however, for the *Drought start hit* event, the POD, SR and TS indices are very low. The Contreras sub-basin was the one that obtained the highest values with a POD of 0.39, a SR of 0.35 and a TS of 0.23. On the other hand, for the *Early Start* event, scenario S4 has POD and SR values of around 0.20 while the TS reaches 0.19 and 0.19 for the Mulet and Tous sub-basins respectively.

As for the *Late start* event, the POD and SR values are close to 0.30 and the TS remains around 0.15. Finally, for the *Drought stay hit* event, index values improved slightly; the POD and SR are around 0.40 and the TS around 0.20.

Figure 9 shows the evolution of the seven indices calculated through the five proposed scenarios, this for the Alarcón sub-basin and for the four events analyzed, *Drought start hit, Early start, Late start and Drought stay hit*. For each graph shown, the optimal value of the first five indices is 1 (PC, BIAS, POD, TS and SR), while for the last two, the optimal value is 0 (FAR and POFD).

In the *Drought start hit* event, the correct portion is 88% for scenario S1 and decreases for the other scenarios to 71% for scenario S5. The event is sub-forecast for scenarios S1, S2 and S3 as they present a BIAS of less than 1, however, for scenarios S4 and S5 it is observed that the event is over-forecasted. The percentage of hits with respect to the occurrence of the event in the observed time series is only 36% for scenarios S4 and S5. By eliminating the *Correct negatives* from the analysis, the *Drought start hit* event has a low success rate (TS) of between 10% and 21%.

The *Early start* event presents a more erratic behavior compared to the *Drought start hit* event, since the PC decreases up to 57% for the S5 scenario and with bias from 0.096 for the S1 scenario, to 1.627 for the S5 scenario. However, for the S4 scenario the event is unbiased. In this event there is a high number of *False alarms* regarding the predicted events, as can be seen in the FAR index.

The *Late start* event shows a high over-forecast as it reaches up to a value of 3.75 for scenario S1, besides a very high FAR as well as the *Early start* event.

Finally, the *Drought stay hit* event has a similar behavior to the *Drought start hit* event, with considerable bias in the first three scenarios and a PC close to its optimal value in the S1 scenario. As in all other events, the value of the FAR index is very high, due to the considerable number of *False alarms*.

In Figure 9 it can be seen that after the analysis carried out, there are 5 values for each of the calculated indices, that is, 35 values for each sub-basin and this for each event, making it difficult to fully analyze the ability of the climate model to predict the drought event. Therefore, an aggregate index is proposed. In order to obtain an average of the scenarios analyzed, the scenario S5 was not considered, given that it is higher than the WEI of the JRB is 87% (Pedro-Monzonís et al., 2014). Figure 10 shows the mean of the four scenarios for each sub-basin and event.

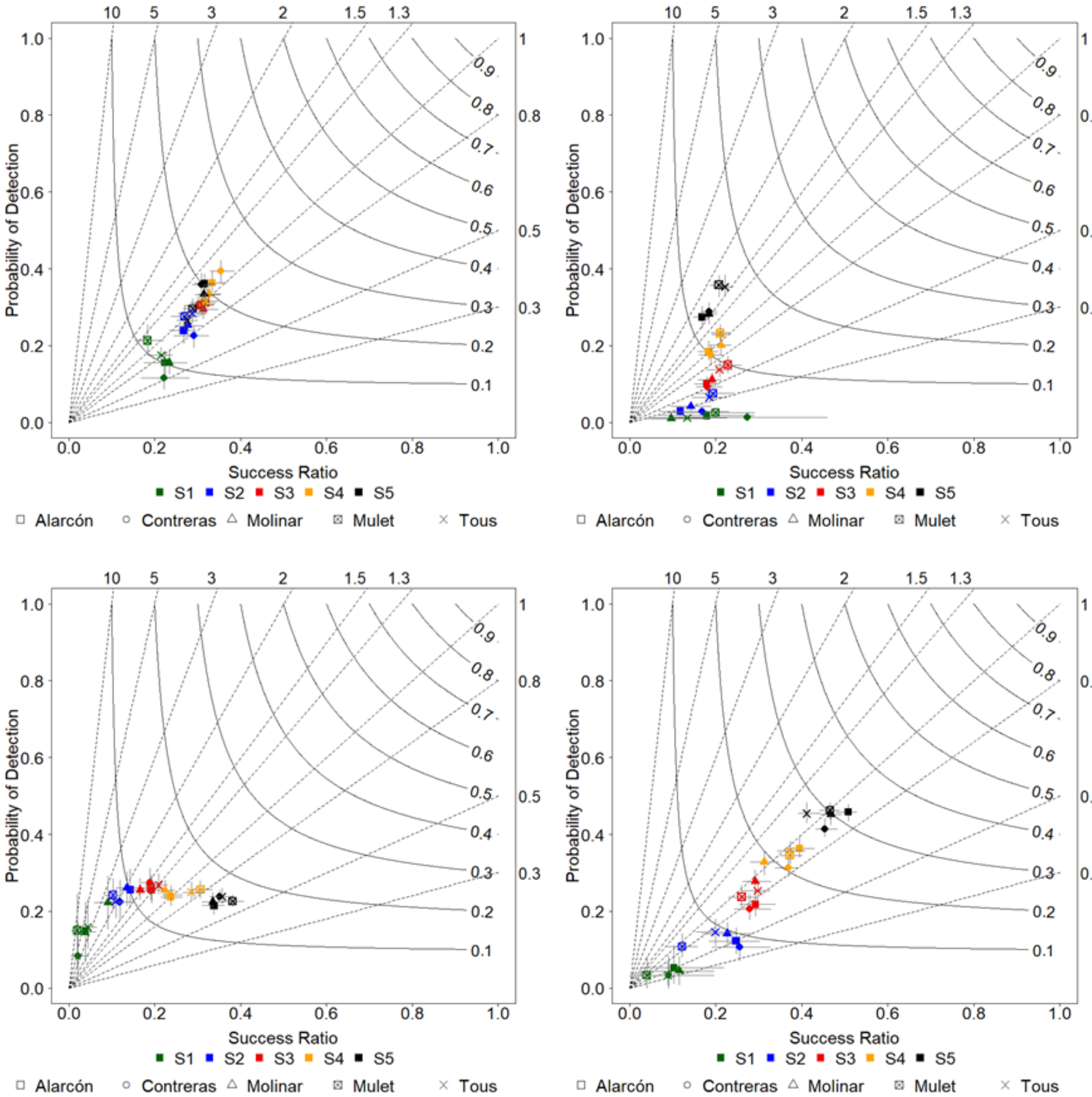

Fig. 8. Performance diagram for the four events analyzed. Top left diagram: Drought start hit event. Top right diagram: Early start event. Bottom left diagram: Late start event. Bottom right diagram: Drought stay hit event.

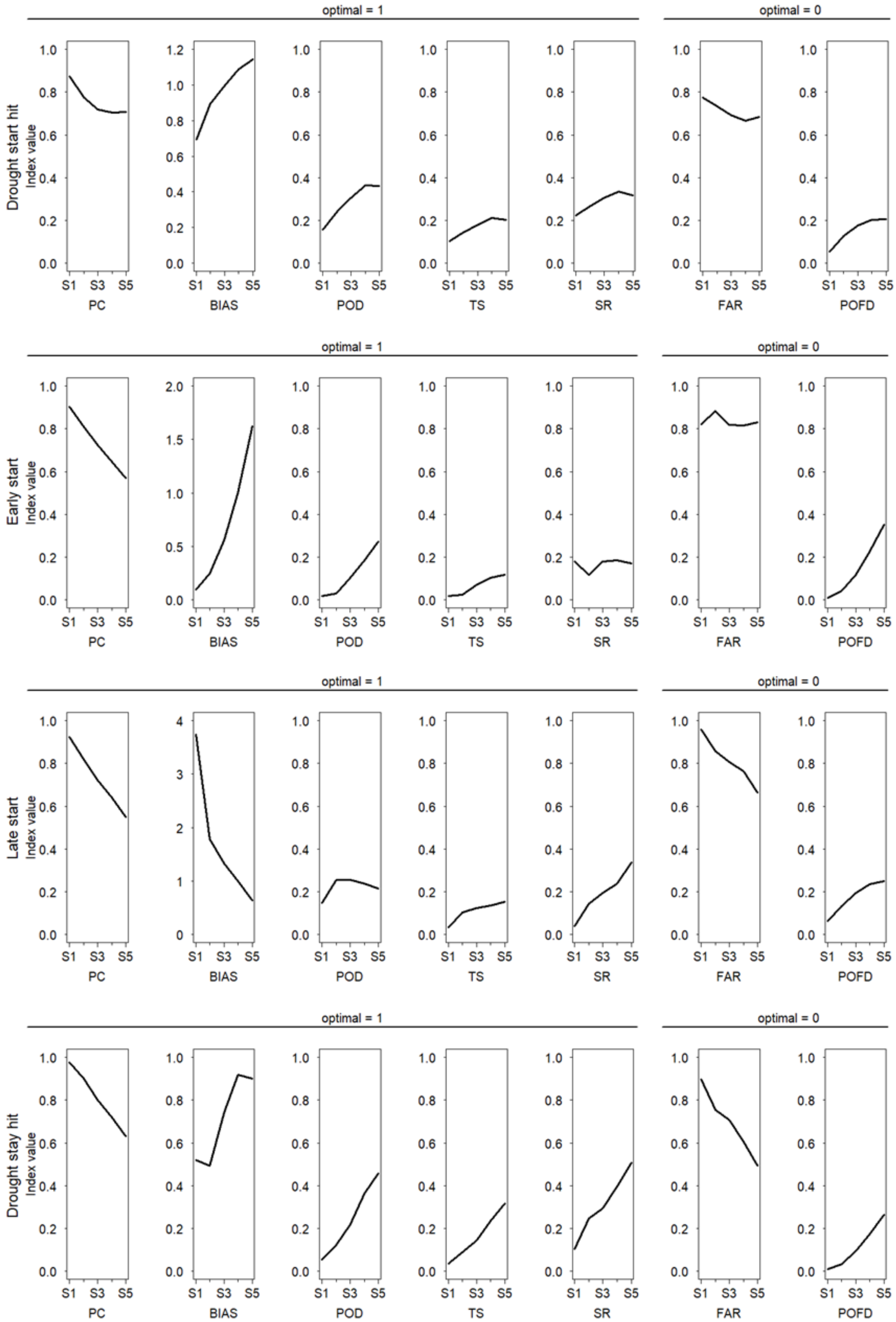

Fig. 9. Indices obtained for the four events analyzed, corresponding to the Alarcón sub-basin. The x-axis corresponds to the five scenarios.

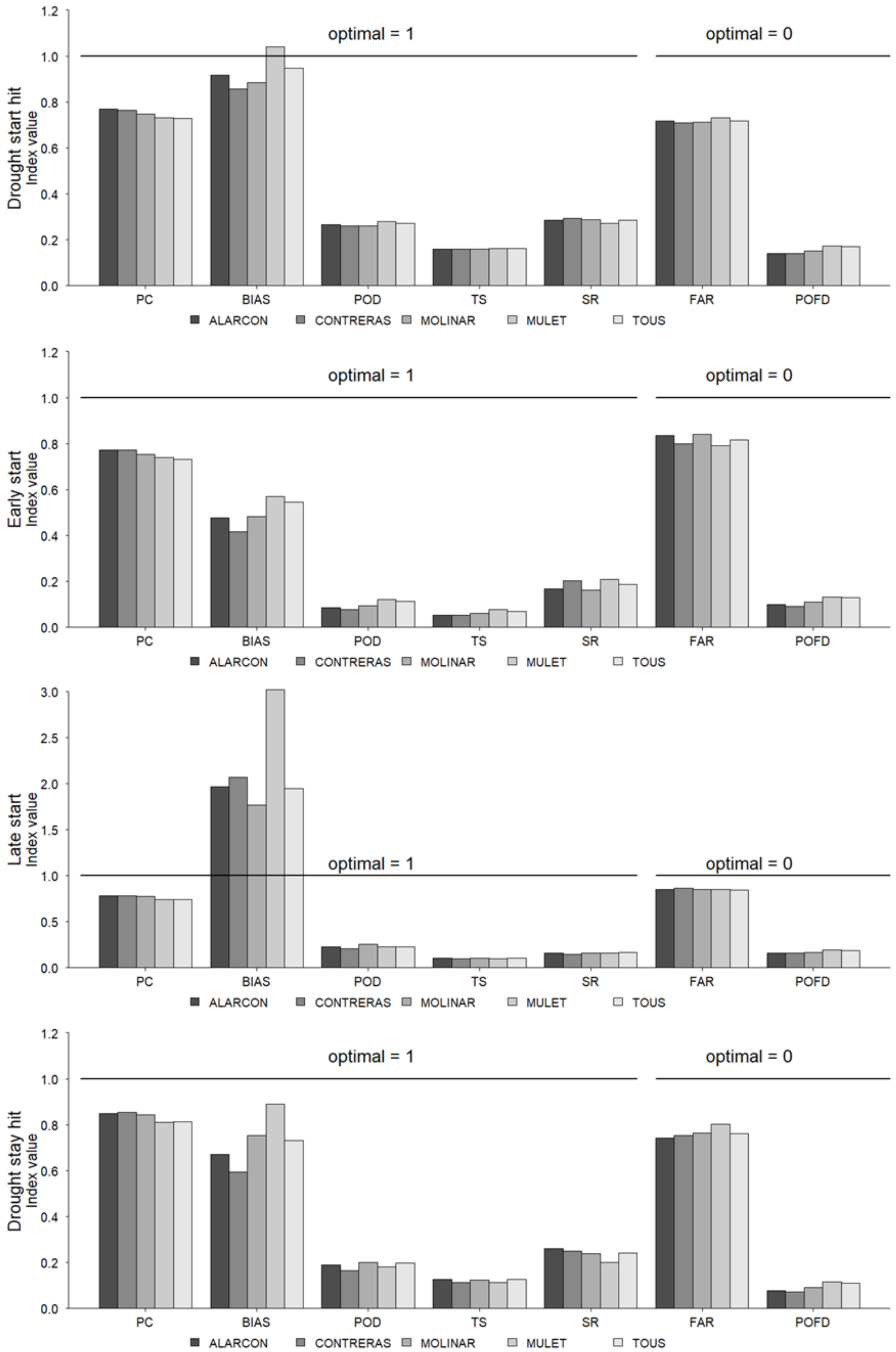

Fig. 10. Average values for each calculated index.

By reducing the number of indices considering only TS and POD, compound indices are obtained for each sub-basin, which are shown in Table 7.

Table 7. Compound indices.

	Index by event			
Catchment	Drought start hit	Early start	Late start	Drought stay hit
Alarcón	0.21	0.07	0.16	0.16
Contreras	0.21	0.06	0.15	0.14
Molinar	0.21	0.07	0.18	0.16
Mulet	0.22	0.10	0.16	0.15
Tous	0.22	0.09	0.16	0.16

When calculating a weighted average, considering the annual mean precipitation, an aggregate index is obtained for the entire basin. The results are shown in the following table.

Table 8. Aggregate indices for the Júcar River Basin.

	Index by event			
Catchment	Drought start hit	Early start	Late start	Drought stay hit
Júcar	0.21	0.08	0.16	0.15

The assessment of the areas under the curves of the ROC diagram for the Drought start hit event, considering only the first four scenarios (0-80%) results in values ranging from 0.11 for the Tous sub-basin to 0.17 for Alarcón. If a weighted average is obtained in the same way as the aggregate index, it has a value of 0.19 (Figure 11).

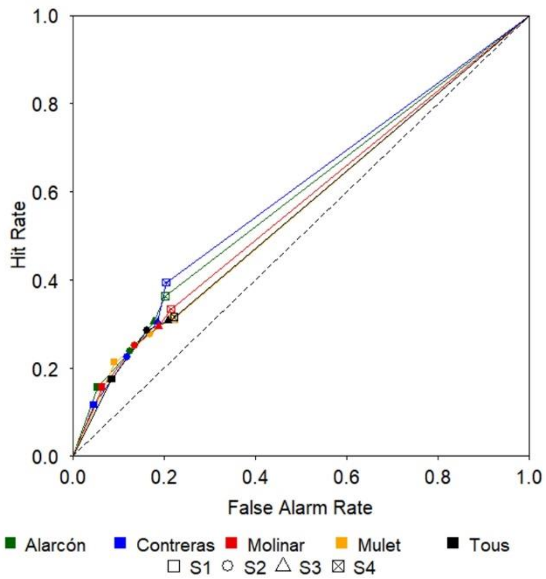

Fig. 11. ROC diagram for Drought start hit event.

The above results may be considered unoptimistic when describing the predictive capacity of the analyzed model. However, this conclusion was predictable given the known uncertainty in the occurrence of future precipitation events. As a result, these cannot be judged on their own, they need to be compared with other alternatives in order to obtain forecasts. For this purpose, the classical method has been chosen, which has already been used on some occasions in the management of the JRB (Andreu et al., 2013; Haro-Monteagudo et al., 2017; Suárez-Almiñana et al., 2017). A stochastic model AR(1), shown in Eq. (10), has been calibrated for the 5 sub-basin time series obtained from Spain02.

$$X_t = \varphi_1 \cdot X_{t-1} + \theta_0 \cdot \varepsilon \quad (10)$$

Where X_t and X_{t-1} are the variables, φ_1 is an autocorrelation matrix, θ_0 is an matrix of coefficients that multiplies the random *N(0,1)* values vector represented by ε. With this AR(1) model, the same number of monthly forecasts have been generated for the same historical period that has been analyzed with the System4 model forecasts. The results obtained are shown below.

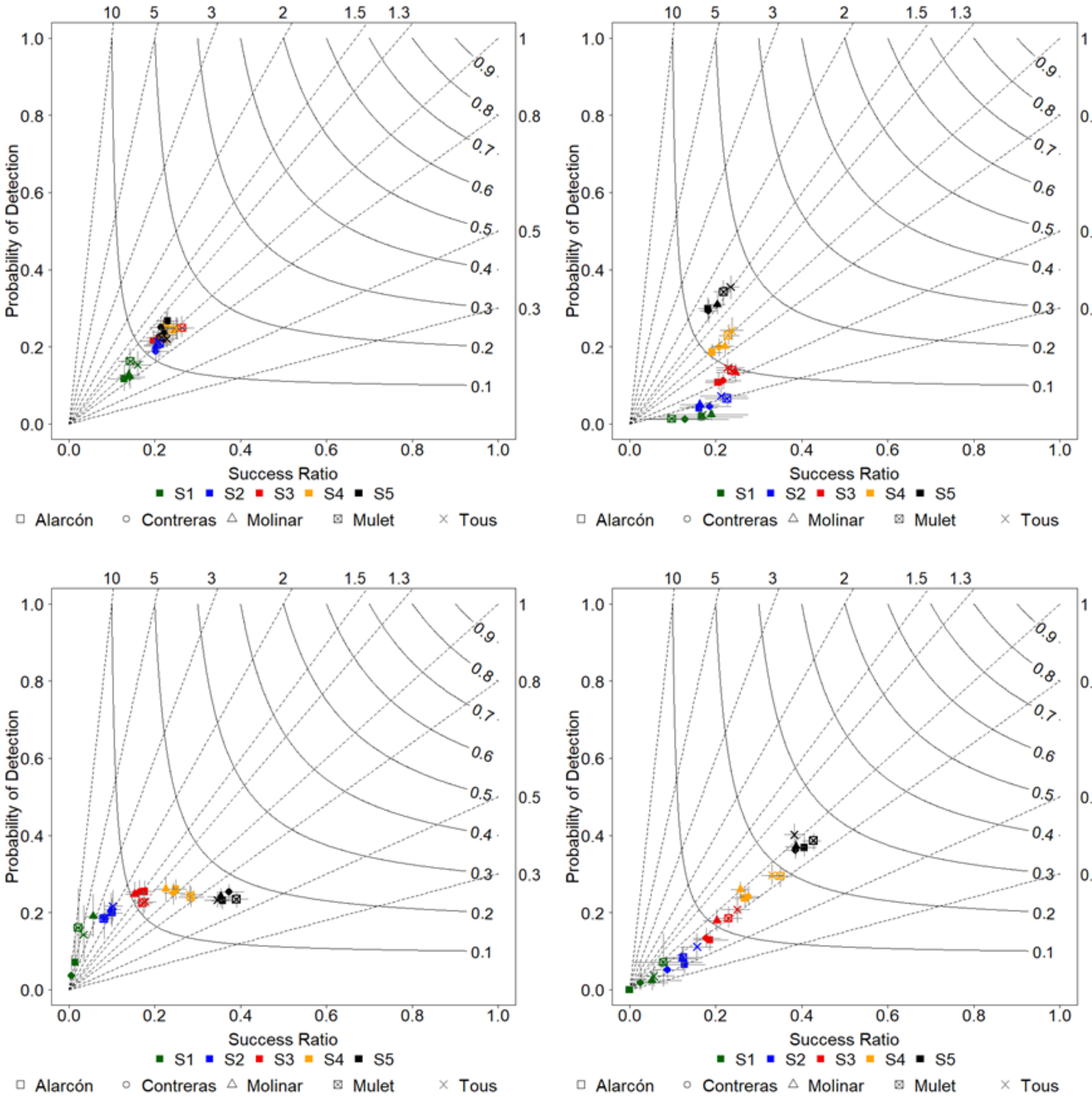

Fig. 12. Performance diagram for the four events analyzed in the autoregressive series.

Top left diagram: Drought start hit event. Top right diagram: Early start event. Bottom left diagram: Late start event. Bottom right diagram: Drought stay hit event.

Table 9. Compound indices. AR(1) model.

	Index by event			
Catchment	Drought start hit	Early start	Late start	Drought stay hit
Alarcón	0.15	0.07	0.14	0.09
Contreras	0.15	0.08	0.13	0.09
Molinar	0.15	0.09	0.16	0.11
Mulet	0.17	0.09	0.14	0.13
Tous	0.17	0.10	0.15	0.13

Table 10. Aggregate indices for the Júcar River Basin. AR(1) model.

Catchment	Index by event			
	Drought start hit	Early start	Late start	Drought stay hit
Júcar	0.16	0.09	0.14	0.11

When comparing the results of the System4 model with the time series generated with the autoregressive model, it can be seen that for the *Drought start hit* event, the AR(1) model presents less bias, however, for the rest of events the performance is very similar between both models (see 8 and Fig.). When comparing the results of the aggregated indices, it can be observed that although the System4 model values are low, they are higher than those obtained for the autoregressive model by 34% for the *Drought start hit* event and up to 42% for the *Drought stay hit* event. Nevertheless, the improvement is between 8% and 12% for Early start and Late start events.

Since the results of the aggregate index obtained for the droughts onset analysis for the System4 and the AR(1) models are similar, it is necessary to perform a contrast test to determine if there is a significant difference. In order to determine if exist this difference a Mann-Whitney U test was performed. The series of both models that were compared were obtained from the POD and TS indices with 40 values each. For each of the two indices, there is a value for each of the five sub-basins and for each of the four scenarios considered (0-80%). The Mann-Whitney U test resulted in a p-value of 0.0037. Therefore, there is a significant difference between the indices of the System4 and the AR(1) models.

In order to make predictions for basin management, it is also important to consider predictive capacity over longer time periods. The cumulative forecast for the entire 7-month period could be analyzed from the System4 data sets. Nonetheless, increasing the forecast period will result in a loss of prediction reliability, so using the first month is likely to be the most reliable.

Analysis in terms of flow rates is also relevant, but in this field it is foreseeable that the AR(1) model will provide a high degree of representativeness due to the subterranean component, which in the case of the Júcar river is very high. Therefore, it will be useful to explore how to improve this predictive capacity, using hydrological models or autoregressive moving average models that include an exogenous variable (ARMAX).

5. Conclusions

In the area of watershed management, greater attention should be paid to drought prediction, so the proposed methodology is skillful to value predictions.

Regardless of the result obtained for the case study, climate predictions are a potentially very valuable tool in basin management, and it is necessary to continue working on improving the generation methods and their validation.

The methodology proposed in this work has an important application for the evaluation of the predictive skill of drought events of climatic and stochastic models, since it has been demonstrated that with its application it is possible to determine the quality of the forecast of this phenomenon.

Based on the analysis of the System4 model data and despite the low values obtained from NSE, KGEM and the aggregated indices, it is concluded that the dynamical System4 coupled model can be used for forecasting drought events given that the aggregated indices obtained were better than those produced by the classical method (autoregressive).

Of the four events analyzed, the most important is the *Drought start hit*, since what we are looking for is that the model can predict when a drought will start, from this point of view, the climate model System4 is more accurate than the AR(1) model, since the aggregate index is 0.21, while for the autoregressive model it is 0.16, on a scale of 0 to 1, where 1 is the optimal value.

The analysis of the Relative Operating Characteristic (ROC) allows verifying the results obtained in the calculation of the aggregate index, since a weighted area below the curve was obtained with a value of 0.19 for the Drought start hit event, when, as in the aggregate index, the optimum value is 1. In order to improve the characterization of the climate model, it is necessary to make a bias correction before applying the methodology developed in this paper, since as was observed in section 4.1, monthly precipitation data show bias in the months with higher precipitation.

It is important, once the precipitation of the model has been analyzed, to carry out an evaluation using this method of the runoff data, as this would allow the hydrological and operational droughts of a water resources management system to be assessed.

6. Acknowledgements

The authors thank AEMET and UC by the data provided for this work (Spain02 gridded data set).

The authors thank the Spanish Research Agency (MINECO) for the financial support to ERAS project (CTM2016-77804-P, including EU-FEDER funds). Additionally, we also value the support provided by the European Community's in financing the projects SWICCA (ECMRWF-Copernicus-FA 2015/ C3S_441-LOT1/SMHI) and IMPREX (H2020-WATER-2014–2015, 641811).

Author Contributions: This manuscript is a result of the Doctoral research of Jaime Madrigal.

Jaime Madrigal received a lot of contributions of the co-authors that are the following.

Abel Solera raised the main idea of develop an aggregate index. Abel Solera, Javier Paredes-Arquiola, Joaquin Andreu and Sonia T. Sánchez-Quispe supported revised methodology, equations, results, literature review and final revisions. Sara Suárez-Almiñana made revisions of the structure of manuscript, English editing and literature review.

Declarations of interest: none.

7. References

Andreu, J., Ferrer-Polo, J., Pérez, M.A., Solera, A., Paredes-Arquiola, J., 2013. Drought Planning and Management in the Júcar River Basin, Spain. In: Schwabe K, Albiac J, Connor JD, et al. (eds) Drought in Arid and Semi-Arid Region. Springer, pp 237–249

Bacanli, U.G., Firat, M., Dikbas, F., 2009. Adaptive Neuro-Fuzzy Inference System for drought forecasting. *Sci Total Environ* 23:1143–1154. doi: 10.1007/s00477-008-0288-5

Baldwin, M.E., Kain, J.S., 2004. Examining the Sensitivity of Various Performance Measures. In: 2.9 (ed) 17th Conf. on Probability and Statistic in the Atmospheric Sciences, 84th AMS Annual Meeting. Seattle, WA, pp 1–8

Bartholmes, J.C., Thielen, J., Ramos, M.H., Gentilini, S., 2009. The european flood alert system EFAS – Part 2 : Statistical skill assessment of probabilistic and deterministic operational forecasts. *Hydrol Earth Syst Sci* 13:141–153. doi: 10.5194/hess-13-141-2009

Cancelliere, A., Salas, J.D., 2004. Drought length properties for periodic-stochastic hydrologic data. *Water Resour Res* 40:1–13. doi: 10.1029/2002WR001750

Dracup, J.A., Lee, K.E., Paulson, E.G., 1980. On the Definition of Droughts. *Water Resour Res* 16:297–302

Dutra, E., Magnusson, L., Wetterhall, F., Hanna, L., Cloke, G.B., Boussetta, S., Pappenberger, F., 2013. The 2010 – 2011 drought in the Horn of Africa in ECMWF reanalysis and seasonal forecast products. *Int J Climatol* 33:1720–1729. doi: 10.1002/joc.3545

Dutra, E., Pozzi, W., Wetterhall, F., Di Giuseppe, F., Magnusson, L., Naumann, G., Barbosa, P., Vogt, J., Pappenberger, F., 2014. Global meteorological drought - Part 2 : Seasonal forecasts. *Hydrol Earth Syst Sci* 18:2669–2678. doi: 10.5194/hessd-11-919-2014

Estrela, T., 2006. La gestión de las sequías en España. *Ing y Territ* 74:52–57

Estrela, T., Vargas, E., 2008. Drought Management Plans in the Spanish River Basins. In: López-Francos (ed) Drought management: scientific and technological innovations. Options Méditerranéennes : Série A. Séminaires Méditerranéens; n. 80, Zaragoza : CIHEAM, pp 157–162

Fernández, C., Vega, J.A., Fonturbel, T., Jiménez, E., 2009. Streamflow drought time series forecasting : a case study in a small watershed in North West Spain. *Stoch Environ Res Risk Assess* 23:1063–1070. doi: 10.1007/s00477-008-0277-8

Ferrer, J., Pérez, M.A., Honrubia, M.A., Perez, F., 2008. Drought administrative actions , drought statutory laws and the Permanent Drought Commission in the Júcar River Basin Authority. In: López-Francos (ed) Drought management: scientific and technological innovations. Options Méditerranéennes : Série A. Séminaires Méditerranéens; n. 80, Zaragoza : CIHEAM, pp 221–226

Finley, J.P., 1884. TORNADO PREDICTIONS. *Am Meteorol Journey* 1:85–88

Great Britain Meteorological Office, 1951. The Meteorological Glossary. Chemical Publishing Co., New

York, USA

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J Hydrol* 377:80–91. doi: 10.1016/j.jhydrol.2009.08.003

Haro-monteagudo, D., Solera, A., Andreu, J., 2017. Drought early warning based on optimal risk forecasts in regulated river systems : Application to the Jucar River Basin (Spain). *J Hydrol* 544:36–45. doi: 10.1016/j.jhydrol.2016.11.022

Haro, D., Solera, A., Paredes, J., Andreu, J., 2014a. Methodology for Drought Risk Assessment in Within-year Regulated Reservoir Systems . Application to the Orbigo River System (Spain). *Water Resour Manag* 28:3801–3814. doi: 10.1007/s11269-014-0710-3

Haro, D., Solera, A., Predo-Monzonís, M., Andreu, J., 2014b. Optimal Management of the Jucar River and Turia River Basins under Uncertain Drought Conditions. *Procedia Eng* 89:1260–1267. doi: 10.1016/j.proeng.2014.11.432

Herrera, S., Fernández, J., Gutiérrez, J.M., 2016. Update of the Spain02 gridded observational dataset for EURO-CORDEX evaluation : assessing the effect of the interpolation methodology. *Int J Climatol* 36:900–908. doi: 10.1002/joc.4391

Herrera, S., Gutiérrez, J.M., Ancell, R., Pons, M.R., Frías, M.D., Fernández, J., 2012. Development and analysis of a 50-year high-resolution daily gridded precipitation dataset over Spain (Spain02). *Int J Climatol* 32:74–85. doi: 10.1002/joc.2256

Hisdal, H., Tallaksen, L.M., 2000. Drought Event Definition. In: ARIDE Technical Report 2000. University of Oslo, Oslo, Norway

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

Kim, H., Webster, P.J., Curry, J.A., Toma, V.E., 2012b. Asian summer monsoon prediction in ECMWF System 4 and NCEP CFSv2 retrospective seasonal forecasts. *Clim Dyn* 39:2975–2991. doi:

10.1007/s00382-012-1470-5

Kling, H., Fuchs, M., Paulin, M., 2012. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *J Hydrol* 424–425:264–277. doi: 10.1016/j.jhydrol.2012.01.011

Leilah, A.A., Al-Khateeb, S.A., 2005. Statistical analysis of wheat yield under drought conditions. *J Arid Environ* 61:483–496. doi: 10.1016/j.jaridenv.2004.10.011

Mason, I.B., 2003. Binary events. In: Jolliffe IT, Stephenson DB (eds) Forecast Verification. Wiley, pp 37–76

Ministerio de Medio Ambiente, 2007. Plan especial de alerta y eventual sequía en la Confederación Hidrográfica del Júcar

Mishra, A.K., Desai, V.R., Singh, V.P., 2007. Drought Forecasting Using a Hybrid Stochastic and Neural Network Model. *J Hydrol Eng* 12:626–638. doi: 10.1061/(ASCE)1084-0699(2007)12

Mishra, A.K., Singh, V.P., 2010. A review of drought concepts. *J Hydrol* 391:202–216. doi: 10.1016/j.jhydrol.2010.07.012

Mishra, A.K., Singh, V.P., 2011. Drought modeling – A review. *J Hydrol* 403:157–175. doi: 10.1016/j.jhydrol.2011.03.049

Molteni, F., Stockdale, T., Balmaseda, M.A., Balsamo, G., Buizza, R., Ferranti, L., Magnusson, L., Mogensen, K., Palmer, T., Vitart, F., 2011. The new ECMWF seasonal forecast system (System 4). Tech. Memo. 656, Reading, UK

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Trans ASABE* 50:885–900. doi: 10.13031/2013.23153

Morid, S., Smakhtin, V., Bagherzadeh, K., 2007. Drought forecasting using artificial neural networks and time series of drought indices. *Int J Climatol* 27:2103–2111. doi: 10.1002/joc

Murphy, A.H., Daan, H., 1985. Forecast evaluation. In: Murphy AH, Katz RW (eds) Probability, Statistics, and Decision Making in the Atmospheric Sciences. Westview, Boulder, pp 379–437

Mwangi, E., Wetterhall, F., Dutra, E., Di Giuseppe, F., Pappenberger, F., 2013. Forecasting droughts in East Africa. *Hydrol Earth Syst Sci* 10:1–22. doi: 10.5194/hessd-10-1-2013

Nash, J.E., Sutcliffe, J.V., 1970. River Flow Forecasting Through Conceptual Models Part I - A Discussion of Principles*. *J Hydrol* 10:282–290. doi: 10.1016/0022-1694(70)90255-6

Ogutu, G.E.O., Franssen, W.H.P., Supit, I., Omondi, P., Hutjes, R.W.A., 2016. Skill of ECMWF system-4 ensemble seasonal climate forecasts for East Africa. *Int J Climatol* 37:2734–2756. doi: 10.1002/joc.4876

Pedro-Monzonís, M., Ferrer, J., Solera, A., Estrela, T., Paredes-Arquiola, J., 2014. Water Accounts and Water Stress Indexes in the European Context of Water Planning: the Jucar River Basin. In: *Procedia Eng* pp 1470–1477

Peirce, C.S., 1884. The numerical measure of the success of predictions. *Science* (80-) 4:453–454. doi: 10.1126/science.ns-4.93.453-a

Roebber, P.J., 2009. Visualizing Multiple Measures of Forecast Quality. *Weather Forecast* 24:601–608. doi: 10.1175/2008WAF2222159.1

Spalding-fecher, R., Chapman, A., Yamba, F., Walimwipi, H., Kling, H., Tembo, B., Nyambe, I., Cuamba, B., 2016. The vulnerability of hydropower production in the Zambezi River Basin to the impacts of climate change and irrigation development. *Mitig Adapt Strateg Glob Chang* 21:721–742. doi: 10.1007/s11027-014-9619-7

Suárez-Almiñana, S., Pedro-Monzonís, M., Paredes-Arquiola, J., Andreu, J., Solera, A., 2017. Linking Pan-European data to the local scale for decision making for global change and water scarcity within water resources planning and management. *Sci Total Environ* 603–604:126–139. doi: 10.1016/j.scitotenv.2017.05.259

Tate, E.L., Gustard, A., 2000. Drought definition: a hydrological perspective. In: Vogt J V, Somma F (eds) Drought and Drought Mitigation in Europe. Kluwer Academic Publishers, the Netherlands, pp 23–48

Wetterhall, F., Di Giuseppe, F., 2017. The benefit of seamless forecasts for hydrological predictions over

Europe. *Hydrol Earth Syst Sci*. doi: 10.5194/hess-2017-527

Wilhite, D.A., 2000. Chapter 1 Drought as a Natural Hazard : Concepts and Definitions. In: 69th edn. Drought Mitigation Center Faculty Publications

Wilhite, D.A., Glantz, M.H., 1985. Understanding the Drought Phenomenon : The Role of Definitions.

Wilks, D.S., 2006. Statistical Methods in the Atmospheric Sciences, Second Edi. San Diego, CA