



Contents lists available at ScienceDirect

# Science of the Total Environment

journal homepage: [www.elsevier.com/locate/scitotenv](http://www.elsevier.com/locate/scitotenv)

## Use of bias correction techniques to improve seasonal forecasts for reservoirs – A case-study in northwestern Mediterranean

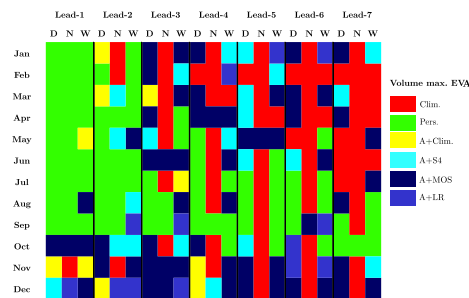
Raül Marcos<sup>a,b,\*</sup>, M<sup>a</sup> Carmen Llasat<sup>b</sup>, Pere Quintana-Seguí<sup>c</sup>, Marco Turco<sup>b</sup><sup>a</sup>Barcelona Supercomputing Center – Centro Nacional de Supercomputación (BSC-CNS), c/ Jordi Girona 29, Barcelona 08034, Spain<sup>b</sup>Department of Applied Physics, University of Barcelona, Av. Diagonal 647, 08028 Barcelona, Spain<sup>c</sup>Observatori de l'Ebre, URL-CSIC, c/ Horta Alta, 38, Roquetes 43520, Spain

### HIGHLIGHTS

- Bias correction of seasonal forecasts at a river basin scale
- Application of MOS-analog, mean bias and linear regression bias adjustments
- Bias correction effect on the seasonal predictability of volume anomalies
- Analysis of the potential for end-users through economic value curves

### GRAPHICAL ABSTRACT

Seasonal predictability of reservoir volume supplies from an economic value area perspective (EVA) and under different conceptual approaches



### ARTICLE INFO

#### Article history:

Received 30 March 2017

Received in revised form 29 June 2017

Accepted 2 August 2017

Available online xxxxx

Editor: D. Barcelo

#### Keywords:

bias correction  
 seasonal forecast  
 water management  
 climate services  
 ECMWF System 4  
 reservoir

### ABSTRACT

In this paper, we have compared different bias correction methodologies to assess whether they could be advantageous for improving the performance of a seasonal prediction model for volume anomalies in the Boadella reservoir (northwestern Mediterranean). The bias correction adjustments have been applied on precipitation and temperature from the European Centre for Middle-range Weather Forecasting System 4 (S4). We have used three bias correction strategies: two linear (mean bias correction, BC, and linear regression, LR) and one non-linear (Model Output Statistics analogs, MOS-analog). The results have been compared with climatology and persistence. The volume-anomaly model is a previously computed Multiple Linear Regression that ingests precipitation, temperature and in-flow anomaly data to simulate monthly volume anomalies. The potential utility for end-users has been assessed using economic value curve areas. We have studied the S4 hindcast period 1981–2010 for each month of the year and up to seven months ahead considering an ensemble of 15 members. We have shown that the MOS-analog and LR bias corrections can improve the original S4. The application to volume anomalies points towards the possibility to introduce bias correction methods as a tool to improve water resource seasonal forecasts in an end-user context of climate services. Particularly, the MOS-analog approach gives generally better results than the other approaches in late autumn and early winter.

© 2017 Elsevier B.V. All rights reserved.

\* Corresponding author at: Barcelona Supercomputing Center – Centro Nacional de Supercomputación (BSC-CNS), c/ Jordi Girona 29, Barcelona 08034, Spain.  
 E-mail address: [raul.marcos@bsc.es](mailto:raul.marcos@bsc.es) (R. Marcos).

## 1. Introduction

The management of water resources is a complex issue, especially in regions prone to hydrological stress and water scarcity (e.g. Pedro-Monzónis et al., 2015). In such situations, end-users have to carefully plan their actions to choose the best decisions and minimize their potential losses (e.g. Wilhite et al., 2000; Bodner et al., 2015). This process is mainly driven by water resource forecasts (Sene, 2010) and the vulnerability degree of the end-user (Downing et al., 2005; Dow et al., 2007). For this purpose, although near-term meteorological forecasts are the most used (e.g. Habets et al., 2008; Adamowski and Karapataki, 2010; Iglesias et al., 2012), the seasonal forecast horizon holds the largest potential, for having information months ahead can substantially increase the resilience of the forecast users (e.g. Block, 2011; Tall et al., 2012; WMO and GFCS, 2016). However, in the operational time-scales, seasonal forecasting is still limited to the use of climatology, a rather conservative approach to diminish the risk of taking misleading decisions (e.g. Cubillo and Garrote, 2008). In fact, regarding dynamical seasonal forecast systems, they are seldom used by any end-user, a preference that can be explained because the skill of these seasonal forecasts is limited in the extra-tropics and for it is difficult to easily communicate the usefulness of the forecasts (e.g. Rayner et al., 2005; Watkins and Wei, 2008). Therefore, it is important to explore new ways to improve seasonal dynamical forecasts and give new insights in their usefulness to provide end-users with the adequate information to decide the best choice for their water management strategies.

The problem of water scarcity, its management and forecast has been already assessed in current literature (e.g. Sene, 2010; Donkor et al., 2014; Bodner et al., 2015). This issue is specially harsh in the Mediterranean where the water deficits of dry summers are often unresolved in the wet season, leading to recurrent drought situations (Blinda et al., 2007; Cook et al., 2016). Under these events resource planning is critical and has to be performed in a seasonal basis. Moreover, this vulnerability is likely to increase in the future due to the rise of drought frequency linked to climate change (e.g. Brewer et al., 2006; Nicault et al., 2008; Turco and Llasat, 2011; Quintana-Seguí et al., 2011) and the growth of water demands (Iglesias et al., 2007). Consequently, in the Mediterranean areas seasonal forecasting could be a very valuable tool to optimize end-user actions and decisions.

Nowadays, although the existing studies confirm the idea that predictability in the extra-tropics is reduced in comparison to the tropics, this does not mean that it is missing at all (e.g. Stockdale, 2000; Quan et al., 2006; Folland et al., 2012; Doblas-Reyes et al., 2013; Scaife et al., 2014; Marcos et al., 2015). However, in spite of the steady advances, seasonal forecasting has to face many theoretical and practical challenges to become an everyday tool in the extra-tropical regions. Actually, the raw forecasts from dynamical systems show biases in comparison to the reference datasets. These biases are the consequence of the inherent limitations of the physical models related to parameterizations, equation simplification and uncertainties in the initialisation procedure (Doblas-Reyes et al., 2013). Such uncertainties can be mitigated through the application of linear and non-linear bias correction techniques that are normally based on statistical methods using antecedent series of forecasts and observations (Weigel et al., 2009; Piani et al., 2010; Peng et al., 2014). However the identification of the best techniques it is not straightforward and might depend on the domain and variable considered as well as in the nature of the bias correction technique itself.

In this paper, we want to study the application of three bias correction approaches (two linear and one non-linear) on the seasonal forecast outcome from the S4 (Molteni et al., 2011). The bias corrected variables then will be introduced in a previously developed (Marcos et al., 2017) seasonal volume-anomaly monthly model for the Boadella reservoir, northwestern Mediterranean, to check whether the use of bias correction techniques is able to improve their

outcome compared to climatology and persistence. Additionally, our implementation will take advantage of the economic value verification as an opportunity to improve the communication of the results to end-users (Marcos et al., 2017).

This paper is organized as follows: “Material and methods” 2 section is a comprehensive description of the domain and the data used in this study; “Methodology” 2.3 presents the implementation framework of the bias correction methods and the volume-anomaly model; “Results” 3 contains the verification metrics of the bias corrections applied and the volume-anomaly performance comparison; and, finally, the “Conclusions” 4 section summarises the main results obtained.

## 2. Material and methods

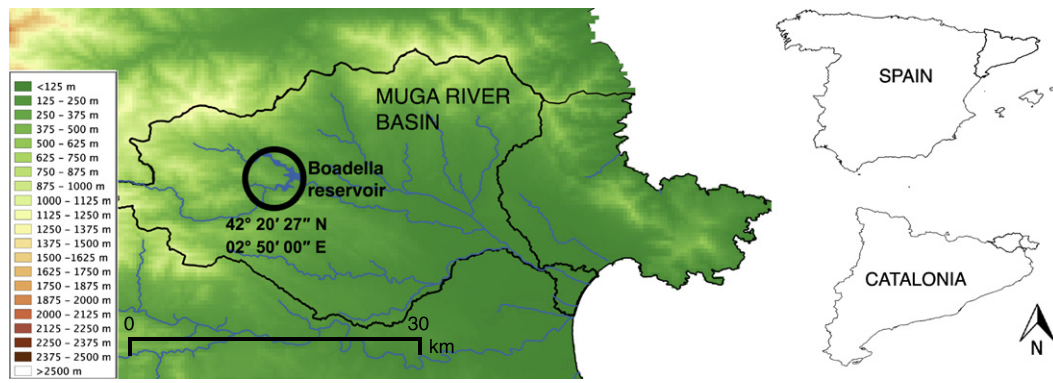
### 2.1. Domain: the Muga Basin and the Boadella reservoir

The Muga River basin is located in the northeastern part of the Iberian Peninsula. It covers a Catalan region delimited to the north by the Pyrenees, and the Mediterranean to the east (Fig. 1). Although its surface is relatively small, about 854 km<sup>2</sup>, there are high contrasts between the mountainous region, with altitudes around 1100 m (summits about 1400 m) and the lower sedimentary plains. The river Muga, which gives the basin's name, is 64 km long and in its upper flow is regulated by the Boadella reservoir, which collects water from a sub-basin of approximately 182 km<sup>2</sup> (Fig. 1). The maximum length of the reservoir is 8.5 km and it has a depth of 54 m. It occupies an area of 364 ha, and its capacity is of 61 hm<sup>3</sup> (Pavón, 2001a,b; Colomer et al., 2004). It accomplishes four goals: a) flood lamination b) irrigation c) urban water supply and d) electricity production.

The mountainous area is mainly covered with forests whereas the lower heights are devoted to agriculture. Main urban areas lie in the lower stream. Its climograph shows that summer is the hottest and driest season (Fig. 2). On the contrary, the wettest season is autumn, with a secondary maximum in late spring (April–May). The coldest season is winter, whereas spring and autumn are milder and show some inertia from the precedent seasons. Temperatures along the year are always positive and tend to be cool in winter (max. ~13 °C; min. ~3 °C), mild in spring–autumn (max. ~17 °C; min. ~10 °C) and hot in summer (max. ~27 °C; min. ~16 °C).

### 2.2. Datasets

In this paper, we use E-OBS v8.0, the European Observational dataset (Haylock et al., 2008), as our reference dataset for atmospheric variables. This is an open dataset covering the European domain at a daily basis, with a resolution of 0.25° × 0.25° from 1950 to nowadays. On the other hand, the seasonal forecasts are given by the European Centre for Medium-range Weather Forecasting System 4 (S4; Molteni et al., 2011), a leading operational seasonal prediction system based on a fully coupled general circulation model that provides operational multi-variable seasonal predictions at 0.75° horizontal resolution. To evaluate the S4 prediction quality and to perform the bias corrections, we also use a set of retrospective forecasts (re-forecasts or hindcasts) emulating real predictions for a 30-year period (1981–2010) with a 15-member ensemble and 7 months forecast horizon for predictions. Each member of the 15-member ensemble comes from the same model (the ECMWF System-4) and is independent of each other. They are the result of running the ECMWF System-4 with slight changes in the initial conditions. Their resolutions are the same and, also, their grid-points (for further details please see Molteni et al., 2011). The S4 original grid (0.75° × 0.75°) has been bi-linearly interpolated to match the 0.25° × 0.25° E-OBS grid and, hence, ease the process of verification and analog search. We have used bilinear interpolation because it is the most straightforward strategy to increase the grid's



**Fig. 1.** Boadella reservoir's location in the Muga River basin (Catalonia, north-eastern Spain). Elaborated from GMTED2010 (Danielson and Gesch, 2011) and Spanish river cover from the *Ministerio de Agricultura, Alimentación y Medio Ambiente*.

resolution without adding information to the source data. In all the cases the variables considered are: a) Daily maximum temperature; b) Daily minimum temperature; and c) Daily precipitation amount.

Hydrological data for the period 1981–2010 comprises mean daily values of total water volume and in-flow measured by the Catalan Water Agency (ACA) in the Boadella reservoir. This dataset has been verified with an automatic checking based on the software HEC-DSS View of the Hydrologic Engineering Center.

We acknowledge that the different spatial resolutions of the seasonal forecast model and the hydrological data might raise some concern. However, since this study focuses on seasonal anomalies at monthly scales, the very low resolution of the seasonal forecasting system is coherent with the low temporal resolution of the analysis. Besides, the monthly statistical model we use to calculate dam volume anomalies works well with low resolution data, as it was shown in our previous paper (Marcos et al., 2017).

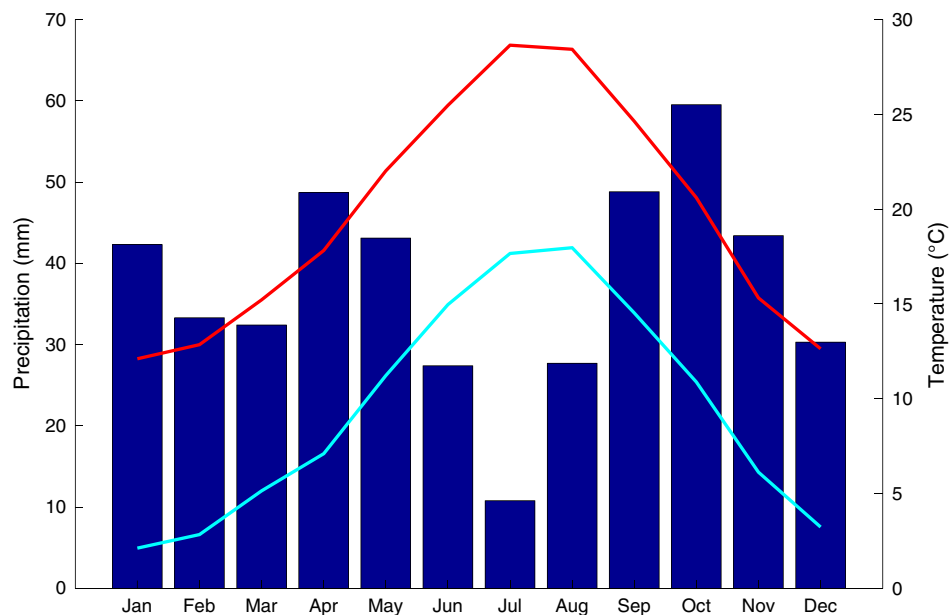
### 2.3. Methodology

#### 2.3.1. Bias correction

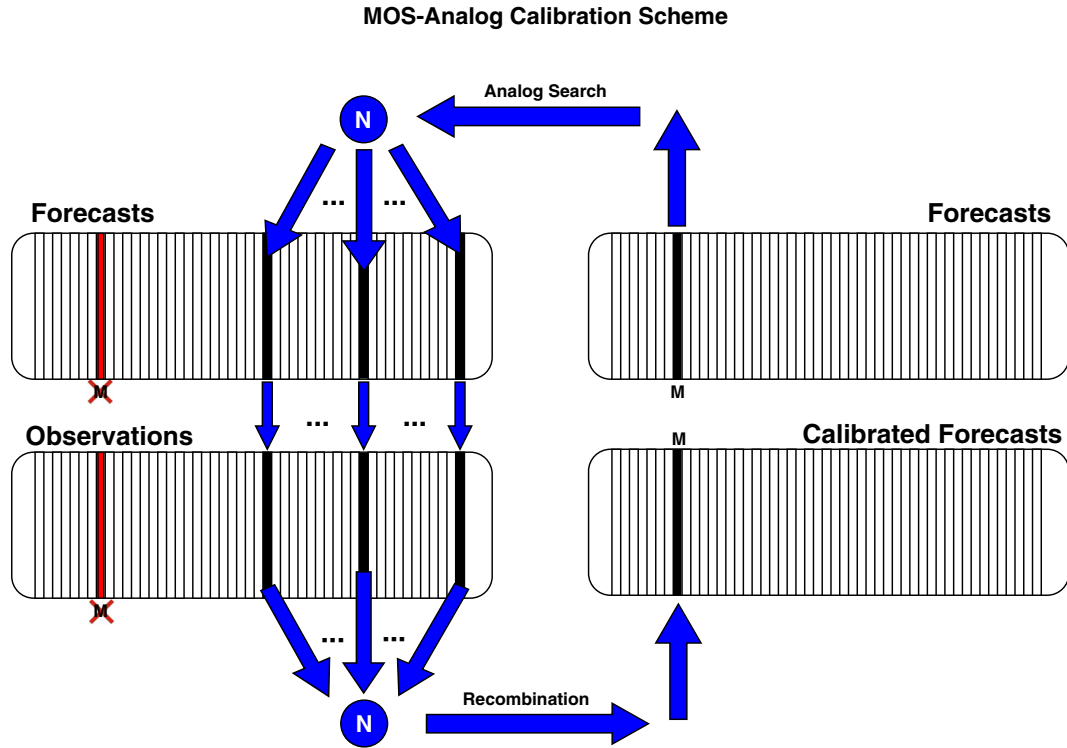
In this study, we have applied three statistical bias correction techniques on the S4 system (they are explained in depth in their corresponding subsection):

- i. MOS-analog (Turco et al., 2011)
- ii. Linear regression
- iii. Mean bias correction

In this way, we could evaluate the efficiency of both non-linear (MOS-analog) and linear (linear regression and mean bias correction) approximations. In the MOS-analog the search for analogs is made considering simultaneously the entire number of region grid-points, whereas in the linear regression and mean bias correction each



**Fig. 2.** Muga basin's climogram. Blue bars refer to monthly rainfall; red line is for monthly maximum temperatures; and cyan, for monthly minimum temperatures. Elaborated from E-OBS v8.0. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 3.** MOS-analog bias correction scheme. For each lead, ensemble member, variable and forecast month we search the  $N$  analogs in the analog pool formed by the same month forecasts taken from the hindcast (excluding the forecast month that we want to calibrate). Afterwards, we take the  $N$  observation fields corresponding to the  $N$  analogs and average them to form a single output that will become our bias corrected forecast.

grid-point series is independently bias corrected. In the three cases, predictors and predictands are referred to the same variables, predictors coming from the S4 previsions and predictands, from E-OBS. Regarding the protocol to assess the performance of the bias corrected forecasts it consists of a Leave-One-Out Cross-Validation (LOOCV). Hence, each year is bias corrected with all the available years except itself. Then the same process is repeated for every year of the 1981–2010 series and the verification metrics are computed.

**2.3.1.1. MOS-analog.** The MOS-Analog technique has been chosen because it is a non-linear method that has already offered interesting results in RCM downscaling (Turco et al., 2011, 2016) and it can be easily transferred to model bias correction. In this case, it consists of seeking, for each member of the S4,  $N$  same member analogs in a pool consisting of forecasts for the same month and horizon from the rest of the hindcast. The analogs are obtained through minimization of the euclidean distance among forecast maps of the same variable. Once found, we select the corresponding E-OBS observation field to be the *true* forecasts. Afterwards, we merge the  $N$  analogs through a simple *pixel mean* procedure to obtain a single *bias corrected* forecast map (Fig. 3).

The MOS application is done upon the principal components computed from every member of the ensemble maintaining a variability explained of the 99% to restrict the influence of its associated error (Gutiérrez et al., 2004). The decision to use principal components instead of direct forecasts in the search for analogs is to reduce the computational burden of the calculations performed. Thus, the MOS-analog bias correction protocol can be summarised in the following steps,

1. Calculation of the principal components of the S4 ensemble monthly forecasts preserving a minimum variability of 99%.
2. Application of the LOOCV MOS-analog process upon the principal components of each variable and for every ensemble

member with the identification of the 5 nearer analogs and their corresponding observed fields.

3. Computation of the mean of the 5 observation fields to obtain a single bias corrected forecast.

It is important to note that the number of analogs retained and the procedure to issue a single bias corrected forecast have to be studied on case basis to obtain the best results (Turco et al., 2011). In Table 1, we attach the possible combinations studied in this work. We have finally chosen the Exp. 7 because it yielded the best results.

**2.3.1.2. Linear regression.** When going for a linear bias correction technique we have chosen the simplest one available, a linear regression of the form,

$$Y = bX. \tag{1}$$

The choice of a regression form without independent term is explained because it is the expected reaction of a perfect model. In fact, without biases or systematic errors, the perfect forecast would

**Table 1**  
MOS-analog experiments.

Exp.	N° analogs	Method*	Analog pool	PC var.
1	1	–	29	95%
2	3	WM	29	95%
3	5	WM	29	95%
4	10	WM	29	95%
5	5	M	29	95%
6	5	M	209	95%
7	5	M	29	99%
8	5	R	29	99%

\* Weighted mean taking into account the euclidean distance (WM)/Mean (M)/Random selection of a single analog (R).



show the same value as the observation. The bias correction is performed through a LOOCV for each variable and for every ensemble member. In each LOOCV approach we have computed the  $b$  parameter taking forecasts as predictors,  $X$ , and observations as predictands,  $Y$ . Afterwards, the forecast to be bias corrected is taken as predictor and with the parameter previously computed we have obtained the final bias corrected forecast. It is important to remember that since this process is repeated for every single year, we end up having  $N$  different calibrating parameters (being  $N$  the number of years).

**2.3.1.3. Mean bias correction.** This strategy consists in searching for the *mean error* in the model forecasts and adding (subtracting) it from every single element of the dataset to have zero mean error. Mathematically the mean error can be written as,

$$ME = \frac{1}{n} \sum_{k=1}^n (f_k - o_k) \quad (2)$$

where  $f_k$  accounts for each of the  $k$  forecast;  $o_k$ , for each of the  $k$  observations; and  $n$  for the total amount of forecast-observation pairs.

### 2.3.2. Verification

The skill of the original S4 forecasts and the bias correction techniques has been evaluated using a LOOCV approach, considering E-OBS observations for the *hindcast* period 1981–2010. In that way, we can simulate the operative way of working since we calibrate the forecasts with all the available past years, except for the current one. Hence, in the verification process the deterministic and probabilistic skills have been assessed. To evaluate the deterministic skill of the S4 and its bias corrections the corresponding parameters have also been computed for *climatology* and *persistence* (at the corresponding lead) in order to identify their added value. The *climatology* value for each month is obtained taking into account all the months from the period 1981–2010 and computing the mean. In the case of *persistence* we consider the forecast as the observed value of the  $n$ -preceding month, where  $n$  is the horizon of the forecast (e.g. the lead 1 forecast for June will be the observed value in May). The selected metrics have been: a) Pearson correlation b) standard deviation and c) Root Mean Squared Error (RMSE).

All the metrics have been computed for each grid-point series, and for the 30-years hindcast. Since the analysis is aimed to study individual months, this means that for each forecast horizon we have 12 field maps with 30 temporal values for each grid-point. One advantage of this approach is that it prevents the influence of the annual cycle in the verification process, because it compares observations and forecasts corresponding to the same month of the year. However, the final verification metric has been presented through a single summarising value. This has been achieved by firstly computing the mean of the 30-element series at each grid-point of the re-forecast and, eventually, calculating the spatial mean among the grid-points.

Additionally, we have used Taylor diagrams to depict in a single figure the centered RMSE (cRMSE), the standard deviation and the Pearson correlation. To do so we have rearranged all the grid-point time series for each region to form a single series concatenating one after the other (Taylor, 2001).

Turning to the probabilistic parameters, we have compared the performance of the S4 and its bias corrections against *climatology* assuming percentiles 33 and 66 as thresholds for below, normal or above normal conditions which is a common format for seasonal forecasts (e.g. Mason et al., 1999; O'Lenic et al., 2008). To do so, we have used the discrete Brier Skill Score (dBSS). The differences between the standard BSS and the dBSS is that the dBSS is a

correction of the BSS when working with small ensembles (Weigel et al., 2007).

Regarding the verification of volume-anomaly forecasts, we have used the Economic Value Area, EVA, which is a verification parameter to summarise the quality of a dichotomous probabilistic forecast from a cost/loss perspective (Richardson, 2000; Gutiérrez et al., 2004; Wilks, 2006). It accounts for the area under the Economic Value curve that optimizes the cost/loss ratio for every potential user of the system under study (Gutiérrez et al., 2004). This is an interesting metric because it always compares two forecast systems from a cost/loss perspective (usually, one new versus another established). A zero value EVA implies that it is best for the user to stay with the customary model. Conversely, every positive value of the EVA means that the new model has the potential to overcome the older one. The closer EVA is to one, the sooner the end-user would see this benefit (Wilks, 2006). It is important to note we have established a threshold for EVA of 0.10 below which we consider that this is intrinsic noise (no other skill measure shows positive results below this value).

### 2.3.3. S4 ensemble recombination

One of the reasons that prevent end-users from using seasonal predictions is the interpretation of probabilistic forecasts (e.g. Juanchich and Sirota, 2015). Although some end-users make the effort to learn how to use them, this is not the case for many others. In this work we have restructured the re-forecast ensemble through what we have called the *generalized ensemble* (GE). It is, plainly, a reorganization of the information of the ensemble in a smaller number of maps constructed with values corresponding to a given percentile. In this way, we could maintain the characterization of the probability density function (*pdf*) of the ensemble without working with its every single member. The GE is an attempt to test whether there is a way to issue a deterministic outcome that, even if not being the best of all, could be used to introduce end-users to the use of seasonal forecasts. We condensate the forecast information calculating the percentiles 10, 25, 50 75 and 90 of the ensemble member distribution, thus considering 5 series instead of 15. Besides, from a scientific standpoint, this approach lets us partially retain the information of the ensemble distribution so we could still identify biases or deficiencies in different parts of the probability density function. To our knowledge, this approximation in literature is seldom used deterministically, because the verification process is mainly limited either to the mean/median maps (e.g. Hagedorn et al., 2005; Johnson and Bowler, 2009) or to extremes (Lim et al., 2010). Here, we also retain the other percentiles to develop a deterministic approach and provide information about deficiencies in the original ensemble such as limited variability and the existence of systematic biases in the shape of the *pdf*. The process to build the GE is as follows:

1. For each grid-point and lead-time retrieve the values corresponding to all the forecasts of the ensemble for that grid-point. In our case,

$$z_j = \{x_1, \dots, x_{15}\}. \quad (3)$$

2. Set percentiles of interest to define the general distribution. To cover the IQR (Inter-Quartile Range) and a lower and upper region of the distribution we have chosen,

$$z_j = \{p_{10}, p_{25}, p_{50}, p_{75}, p_{90}\}. \quad (4)$$

3. For each grid-point, draw the values corresponding to those pre-set percentiles to build a new ensemble matching the percentile thresholds defined. In our case,

$$z_j = \{x_{p_{10}}, x_{p_{25}}, x_{p_{50}}, x_{p_{75}}, x_{p_{90}}\}. \quad (5)$$

The choice of the p10 and p90 instead of the more common p95 or p99.9 to represent the extremes pursues the reduction of the induced error coming both from the size of the ensemble and the coarse E-OBS resolution. That said, p10 and p90 are the characteristic measures of moderate climate extremes (e.g. [Moberg and Jones, 2005](#)).

2.3.4. S4 ensemble unification

Current models usually comprise the association of deterministic control output along with an ensemble of perturbed runs. The former is used to embody a deterministic forecast whereas the latter is the base for probabilistic predictions. However, one may ask whether there is a way to take advantage of the benefits of the ensemble to issue a deterministic forecast instead of sticking to the unperturbed run of the model. Here, we will offer an approach based on the construction of a single time series from the GE. To do so, we have to follow these steps:

1. Create a GE from the original ensemble (see [Section 2.3.3](#)). In our case, the GE is formed of 5 member-series corresponding to percentiles 10, 25, 50, 75 and 90.
2. For each element of the grid-point series of the GE evaluate the Mean Absolute Error (MAE) for all the other elements following a LOOCV. Here, we have a hindcast period of 30 years (1981–2010) so each monthly MAE will be evaluated with 29 elements for each of the five members of GE.
3. The forecast for that particular element will be the value of the one in five GE members showing the lowest MAE in the LOOCV.

2.3.5. MLR models for reservoir volume anomalies

The application of seasonal forecast bias correction to water resources has been based on the methodology we had presented in a previous work ([Marcos et al., 2017](#)). That work describes the elaboration of monthly multiple linear regression models (MLR) to foresee volume anomalies in the Boadella reservoir and the application of economic value curves to communicate the usefulness of the method to the end-users. [Table 2](#) shows each monthly model along with its performance in perfect prognosis conditions. In this paper, we forecast the monthly anomalies with those models (for each month of the year and up to lead seven) comparing six strategies for the period 1981–2010 in LOOCV: a) Climatology (Clim.) b) Persistence (Pers.) c) Antecedent observations + climatology (A + Clim.) d) Antecedent observations + S4 anomalies (A + S4) e) Antecedent

observations + MOS-analog bias corrected S4 anomalies (A + MOS) and f) Antecedent observations + LR bias corrected S4 anomalies (A + LR).

Climatology (Clim.) is the operational strategy against which the other approximations are compared. The persistence approach (Pers.) consists in adopting the observed predictand anomaly for the month-*i* as our forecast, where *i* refers to the forecast lead that we are evaluating. E.g. for lead 1 forecast we take as forecast the anomaly value of the precedent month, month-1; for lead 2 we use the anomaly value of the month-2 and so forth for the other leads. The third strategy (A + Clim.) combines, in our MLR models, antecedent (A) information and climatology (Clim.) to study the predictability linked to the memory of the predictors. The difference between A + Clim. and Clim. is that in A + Clim. we use the MLR model including, when necessary, antecedent information, A, plus no-anomaly, Clim., for the future unknown months. With the Clim. approach, on the other hand, we only use the no-anomaly value of the predictand as our forecast (without applying the MLR model). Finally, the last three approaches combine both observations and bias corrected S4 forecasts (simple bias correction, A + S4; MOS-analog, A + MOS; and linear regression correction, A + LR).

3. Results

3.1. Verification of precipitation and temperature forecasts

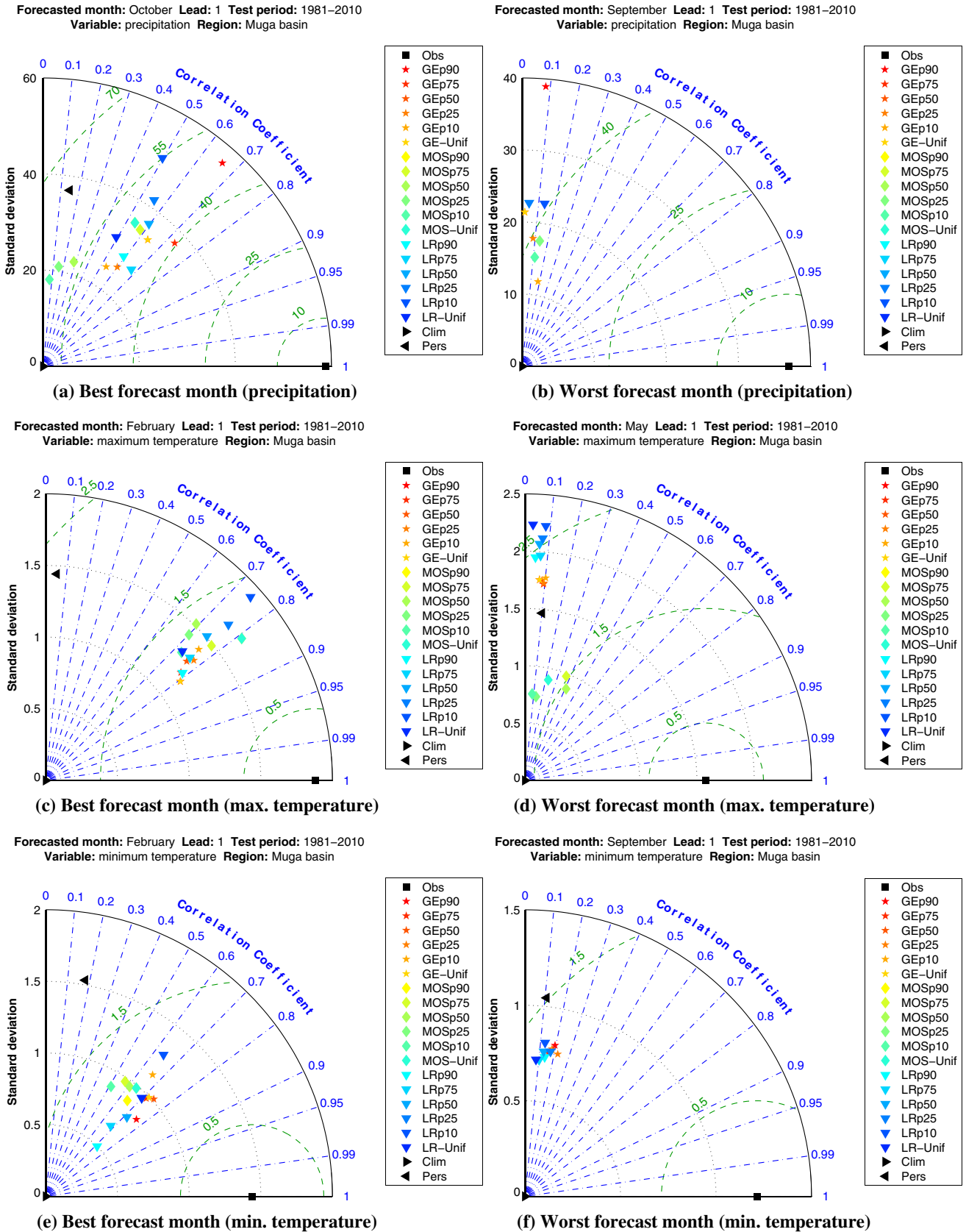
The best results for precipitation are observed at lead one. Actually, the Taylor diagrams show that the original S4 offers generally better results than its LR and MOS-analog bias corrections. [Fig. 4a](#) and [b](#) depict the best and worst months (October and September, respectively). More specifically, the S4 usually surpasses climatology and persistence is the worst option. On the other hand, the largest ameliorations are found in March–June and October–December. When analysing the dBSS ([Fig. 5a](#) and [b](#)), and focusing on the mean bias correction, we find that for the lower tercile (dry conditions; [Fig. 5a](#)) and lead one there is no positive value. Eventually, in the upper tercile the mean bias correction improves February, March, May, June, October, November and December (wet conditions; [Fig. 5b](#)). Several positive results are also observed at other leads, but their lack of consistency points towards the weak signal to noise ratio found at these time scales.

Regarding maximum temperature, the comparison of lead one Taylor diagrams among the original S4 forecasts and its LR and MOS-analog bias corrections displays the finest results for the former, although there are cases where the MOS-analog shows a better behaviour. February is the best predicted month, while May is the worst ([Fig. 5c](#) and [d](#)). Climatology is usually surpassed and persistence is the worst option to choose. Turning to the dBSS analysis of the mean bias corrected S4, we can spot positive values at multiple leads in both categories (lower tercile, [Fig. 4c](#); and upper tercile, [Fig. 4d](#)). In the lowest tercile ([Fig. 5c](#)) the amelioration is observed in January–March, June, August, October and November. As for the upper category ([Fig. 5d](#)) there is an enhancement between December–March and August. An interesting feature of this assessment is the apparent consistency of lead-6 predictability since we have found positive values for the period March–May in the lower tercile and May–June, in the upper. However, the robustness of this result has to be checked further in future experiments.

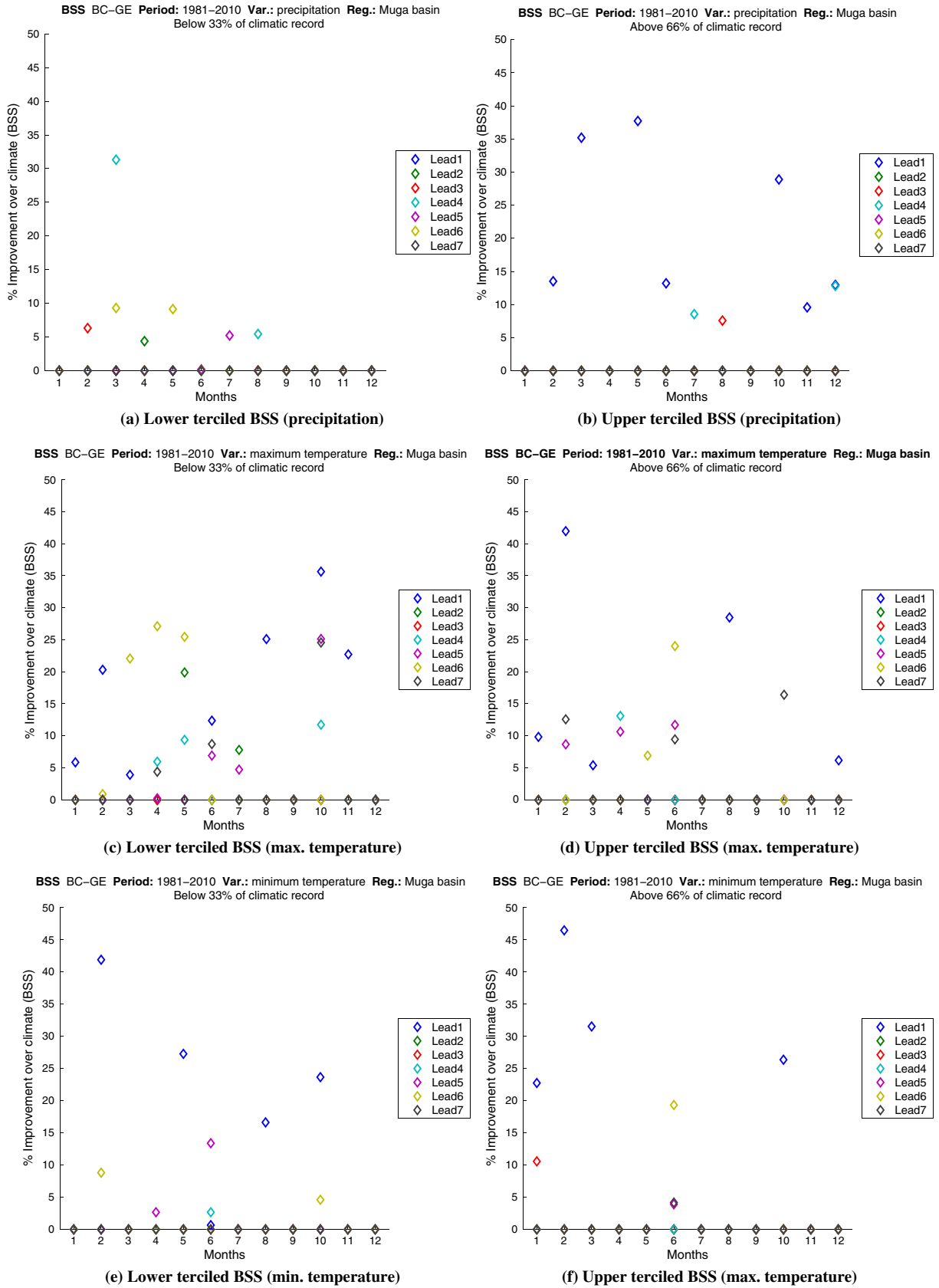
Finally, for minimum temperature the best results are found also at lead one. The corresponding Taylor diagrams show the original S4 forecasts perform better than the LR and MOS-analog bias corrections. The best predicted month is February whereas the worst is September ([Fig. 4e](#) and [f](#), respectively). In some particular cases, such as March, the MOS-analog gives better results. Climatology is not always surpassed; and persistence is the worst choice. The biggest amelioration is observed in the January–March period, while the

**Table 2**  
Performance of the best predictor combinations for volume monthly anomalies in the Boadella reservoir (in-flow, *flwin*; maximum temperature, *Tx*; and minimum temperature, *Tn*). The subscript numbers, *n* and *m*, respectively indicate the first and last month of the period considered for the anomaly computation (up to one year backwards). Each row contains the month, the best predictor combination and the reproduced variance, *R*<sup>2</sup>, in LOOCV perfect conditions. According to [Marcos et al. \(2017\)](#).

Month	Best predictor combination	<i>R</i> <sup>2</sup>
Jan	{ <i>flwin</i> <sub>(5-1)</sub> , <i>Tx</i> <sub>(9-9)</sub> , <i>Tn</i> <sub>(11-11)</sub> , <i>Tn</i> <sub>(2-4)</sub> }	0.76
Feb	{ <i>flwin</i> <sub>(5-2)</sub> , <i>Tx</i> <sub>(10-10)</sub> , <i>Tn</i> <sub>(10-11)</sub> }	0.69
Apr	{ <i>flwin</i> <sub>(2-4)</sub> , <i>flwin</i> <sub>(5-12)</sub> , <i>Tx</i> <sub>(1-1)</sub> , <i>Tx</i> <sub>(6-9)</sub> , <i>Tn</i> <sub>(8-9)</sub> }	0.66
May	{ <i>flwin</i> <sub>(3-4)</sub> , <i>flwin</i> <sub>(6-12)</sub> , <i>Tn</i> <sub>(8-9)</sub> , <i>Tx</i> <sub>(1-1)</sub> }	0.60
Jun	{ <i>flwin</i> <sub>(3-4)</sub> , <i>flwin</i> <sub>(11-12)</sub> , <i>Tn</i> <sub>(9-9)</sub> , <i>Tx</i> <sub>(12-1)</sub> , <i>Tx</i> <sub>(9-9)</sub> }	0.66
Jul	{ <i>flwin</i> <sub>(3-6)</sub> , <i>flwin</i> <sub>(7-7)</sub> , <i>Tn</i> <sub>(6-7)</sub> , <i>Tn</i> <sub>(9-9)</sub> , <i>Tx</i> <sub>(9-9)</sub> , <i>Tx</i> <sub>(7-7)</sub> }	0.67
Aug	{ <i>flwin</i> <sub>(3-5)</sub> , <i>flwin</i> <sub>(6-8)</sub> , <i>Tn</i> <sub>(6-7)</sub> , <i>Tx</i> <sub>(1-1)</sub> , <i>Tx</i> <sub>(9-9)</sub> }	0.79
Sep	{ <i>flwin</i> <sub>(3-6)</sub> , <i>flwin</i> <sub>(7-8)</sub> , <i>Tn</i> <sub>(6-7)</sub> , <i>Tx</i> <sub>(7-8)</sub> , <i>Tx</i> <sub>(9-9)</sub> }	0.76
Oct	{ <i>flwin</i> <sub>(3-4)</sub> , <i>flwin</i> <sub>(6-10)</sub> , <i>Tx</i> <sub>(1-1)</sub> , <i>Tx</i> <sub>(10-10)</sub> , <i>Tx</i> <sub>(9-9)</sub> }	0.82
Nov	{ <i>flwin</i> <sub>(4-4)</sub> , <i>flwin</i> <sub>(7-10)</sub> , <i>Tn</i> <sub>(12-4)</sub> , <i>Tx</i> <sub>(2-2)</sub> }	0.84
Dec	{ <i>flwin</i> <sub>(5-12)</sub> , <i>Tn</i> <sub>(6-7)</sub> , <i>Tx</i> <sub>(2-2)</sub> , <i>Tx</i> <sub>(9-9)</sub> }	0.83

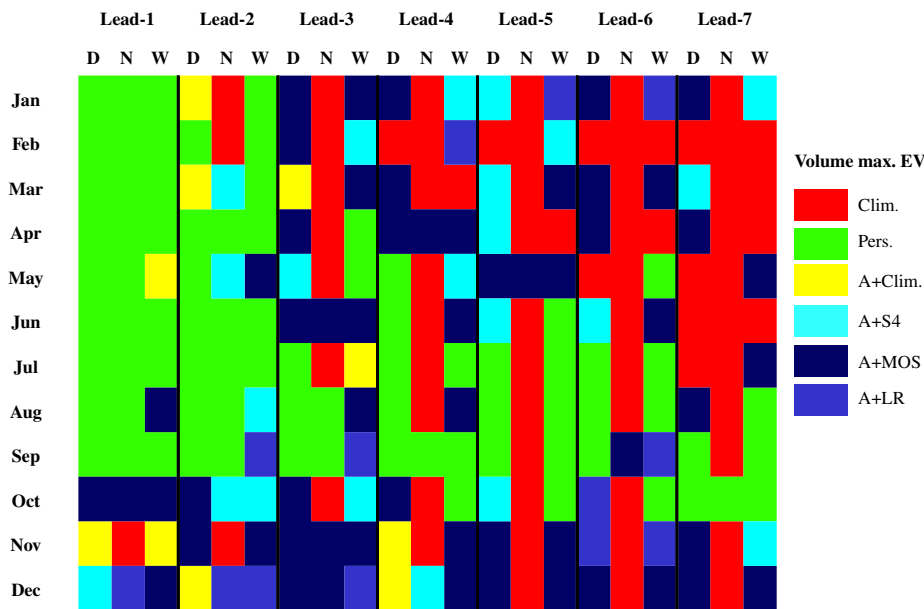


**Fig. 4.** Muga basin's Taylor diagrams at lead one for the best and worst performing months. Each diagram contains the GE for raw S4 forecasts, MOS-analog and MOS-LR S4 bias corrections, their unified deterministic forecast as well as persistence and climatological controls. The period of study is 1981–2010 (a) The best performing month (precipitation): October (b) The worst performing month (precipitation): September (c) The best performing month (max. temperature): February (d) The worst performing month (max. temperature): May (e) The best performing month (min. temperature): February (f) The worst performing month (min. temperature): September.



**Fig. 5.** Muga basin's dBSS improvement over climate for BC-GE S4 considering all months and leads. The period of study is 1981–2010 (a) Lower tercile, precipitation (b) Upper tercile, precipitation (c) Lower tercile, max. temperature (d) Upper tercile, max. temperature (e) Lower tercile, min. temperature (f) Upper tercile, min. temperature.





**Fig. 6.** Volume forecast strategy with the largest Economic Value Area (EVA) for each month, forecast horizon and climatic conditions (D -lower tercile-, N -middle tercile-, and W -upper tercile-) at the Boadella reservoir. The six forecast systems considered are: climatology (Clim.), persistence (Pers.), antecedent observation combined with climatological values (A + Clim.), antecedent observations combined with mean bias corrected S4 (A + S4), antecedent observations combined with MOS-analog bias corrected S4 (A + MOS) and antecedent observations combined with S4 bias corrected with a linear regression procedure (A + LR). Note that forecast strategies different from climatology only appear in the table if they have a minimum EV Area of 0.10.

smallest is in June and November. The dBSS assessment displays the best results for the mean bias adjustment (Fig. 5e and f). In the lowest tercile we have positive values for February, May, June, August and October (Fig. 5e). Finally, for the upper tercile, we find positive values in January, February, March, October and June (Fig. 5f).

### 3.2. Application to reservoir volume anomalies

At first lead persistence is the leading method in all climatic categories except from October to December, where the A + MOS and the A + Clim offer the greatest EVA (Fig. 6). For lead two persistence shows the finest upgrade in all the three terciles in April, June and July. In February, it is the best choice in the lower and upper terciles; in March, for the upper tercile; in May, for the lower; and in August and September, for the lower and middle terciles. At this forecast horizon the A + S4/MOS/LR are the best options in the rest of terciles and months. The middle tercile is not enhanced in January, February and November. At lead three the lower and upper terciles always offer better results than climatology. Yet, we have that the finest is generally A + MOS, followed by persistence and A + S4. The middle tercile is only better modelled than climatology in five months: June, August, September November and December. In June, November and December the best forecast system is A + MOS whereas in August and September persistence is the most advantageous. From August to December there is an upgrade with respect to climatology in the upper and lower terciles up to lead seven, normally by means of persistence but also with A + MOS and S4. This pattern can be also seen in May and June up to lead five; and in July, up to lead six, but in the latter case being persistence the best option. As for November, December and January the predictability of the first and third tercile reaches lead seven, but in this case the dominant forecast system is the A+MOS. This also happens in the lower tercile in April and August; and in the upper tercile in May and July. From lead four to lead seven the middle tercile is normally better forecast by climatology and only in scattered cases is there another forecast approach which gives better results.

## 4. Conclusions

In this paper, we have studied whether different bias correction strategies -mean bias correction, MOS-analog and linear regression- could be useful to improve seasonal forecasting at a river basin scale and transfer this amelioration to the prognosis of water resources. Firstly we have evaluated the skill of the raw S4 output in the Muga river basin (northwestern Mediterranean) and we have compared the original and bias-corrected S4 forecasts with climatology and persistence. To do so we have studied the S4 hindcast period 1981–2010 for each month of the year and up to lead seven considering an ensemble of 15 members rearranged in the so-called *generalized* form. Finally, we have performed a test to check the potential improvement these corrections might have when assessing the seasonal forecast of water resources in the Boadella reservoir through the economic value area metric.

Our study showed that the MOS-analog and LR bias corrections show mixed results when compared to the S4, giving better performance more often for maximum temperature than for precipitation and minimum temperature. The mean bias correction, on the other hand, is more consistent and yields usually better results. That said, the MOS-analog correction sometimes gives greater ameliorations than the mean bias and linear regression corrections. It is worth noting that the mean bias S4 correction enhancements are rather limited to the first lead. When looking to the dBSS of the lower and upper terciles, though, these improvements are more restricted to winter and some scattered months in other seasons. More specifically, we find that for precipitation February, March, October and November show the largest ameliorations. The same months are found for minimum temperature only exchanging November by January. Finally, for maximum temperature, the best months are January–March, August and October–November. Thus, we spot winter skill in the three analysed variables and also in autumn and the early spring.

The recurrent predictability of February and October in regions of the Spanish Mediterranean is highly valuable because both are key months regarding water management from an end-user perspective. For instance, the observed first-lead winter predictability might be

related to the stability of the winter general circulation anomalies, but this is something that has to be checked in future works. In the occasions when there is first lead upgrade beyond climatology the bias study seems to confirm that first order model biases are the cause for the reduction in skill at this horizon. Finally, focusing on persistence, it is surpassed by the S4 in virtually all leads and months. Occasionally, the results have also shown some amelioration beyond climatology at other leads. Yet, this predictability is not as systematic as the corresponding to lead one and, therefore, to discard the possibility of being noise, it has to be independently studied case by case (i. e. lead 6 maximum temperature forecasts for March–May).

Regarding the application of S4 forecasts to water resources, it has been based on the use of previously developed monthly multiple linear regression volume anomaly models for the Boadella reservoir (Marcos et al., 2017). These models proved to be well suited to reproduce and forecast volume anomalies in this reservoir. The analysis of the results showed that the economic value could go beyond climatology up to lead four in all months but February, and even reaching lead seven from August to January. Actually, in late autumn and early winter, the best forecast configuration is the A + MOS, an outcome that could be interesting for end-users since this period holds great part of the precipitation that feeds the reservoir. In the other months, persistence is the best forecast system up to lead two, and reaching lead five from June to September. This points towards the possibility to introduce persistence as another potential predictor in the volume-anomaly models.

These results encourages us to proceed further in our research by seeking the collaboration of end-users to tailor future experimental suites to test the functionality of this methodology even at operational level. Furthermore, future steps will also involve the application of this strategy to other reservoirs to test the transferability of the results obtained.

## Acknowledgments

We thank the Catalan Water Agency for the hydrological data provided. We acknowledge the AEMET and ECMWF for the ECMWF System 4 ensemble re-forecast data. We also acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (<http://ensembles-eu.metoffice.com>) and the data providers in the ECA&D project (<http://www.ecad.eu>). Raül Marcos thanks the Ministerio de Educación Cultura y Deporte for the FPU (grant reference AP2010-0999) and the Agustí Pedro i Pons University Foundation funding for international research projects. Marco Turco was supported by the Spanish Juan de la Cierva Programme (IJCI-2015-26953).

## References

- Adamowski, J., Karapatakis, C., 2010. Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: evaluation of different ANN learning algorithms. *J. Hydrol. Eng.* 15 (10), 729–743.
- Blinda, M., Boufaroua, M., Carmi, N., Davy, T., Detoc, S., Froebrich, J., Gabina, D., Iglesias, A., Khalid, A., Khalifa, E., Margat, J., Moneo, M., Tagar, Z., 2007. Mediterranean Water Scarcity and Drought. *Tech. Rep.* European Commission.
- Block, P., 2011. Tailoring seasonal climate forecasts for hydropower operations. *Hydrol. Earth Syst. Sci.* 15 (4), 1355–1368.
- Bodner, G., Nakhforoosh, A., Kaul, H.P., 2015. Management of crop water under drought: a review. *Agron. Sustain. Dev.* 35 (2), 401–442.
- Brewer, S., Alleaume, S., Guiot, J., Nicault, A., 2006. Historical droughts in Mediterranean regions during the last 500 years: a data/model approach. *Climate Past Discuss.* 2 (5), 771–800.
- Colomer, J., Serra, T., Casamitjana, X., Baserba, C., 2004. L'embassament de Boadella. *Revista de Girona* 52–54.
- Cook, B.I., Anchukaitis, K.J., Touchan, R., Meko, D.M., Cook, E.R., 2016. Spatiotemporal drought variability in the Mediterranean over the last 900 years. *J. Geophys. Res. Atmos.* 121 (5), 2060–2074.
- Cubillo, F., Garrote, L., 2008. Drought risk and vulnerability in water supply systems. *Proc. Drought Manag. Sci. Technol. Innov.* 80 (80), 21–28.
- Danielson, J.J., Gesch, D.B., 2011. Global multi-resolution terrain elevation data 2010 (GMTED2010). *U.S. Geol. Surv. Open File Rep.* 2011-1073 2010, 26.
- Doblas-Reyes, F.J., García-Serrano, J., Lienert, F., Biescas, A.P., Rodrigues, L.R.L., 2013. Seasonal climate predictability and forecasting: status and prospects. *Wiley Interdiscip. Rev. Clim. Chang.* 4 (4), 245–268.
- Donkor, E.A., Mazzuchi, T.A., Soyer, R., Alan Roberson, J., 2014. Urban water demand forecasting: review of methods and models. *J. Water Resour. Plan. Manag.* 140 (2), 146–159.
- Dow, K., O'Connor, R.E., Yarnal, B., Carbone, G.J., Jocoy, C.L., 2007. Why worry? Community water system managers' perceptions of climate vulnerability. *Glob. Environ. Chang.* 17 (2), 228–237.
- Downing, T.E., Aerts, J., Soussan, J., Barthelemy, O., Bharwani, S., Hinkel, J., Ionescu, C., Klein, R.J., Mata, L.J., Matin, N., Moss, S., Purkey, D., Ziervogel, G., 2005. Integrating social vulnerability into water management.
- Folland, C.K., Scaife, A.A., Lindesay, J., Stephenson, D.B., 2012. How potentially predictable is northern European winter climate a season ahead? *Int. J. Climatol.* 32 (6), 801–818.
- Gutiérrez, J.M., Cano, R., Cofiño, A.S., Sordo, C.M., 2004. *Redes Probabilísticas Y Neuronales Aplicadas a Las Ciencias Atmosféricas*. Centro de Publicaciones. Secretaría Técnica. Ministerio de Medio Ambiente, Santander.
- Habets, F., Boone, A., Champeaux, J.L., Etchevers, P., Franchistéguy, L., Lebouis, E., Ledoux, E., Le Moigne, P., Martin, E., Morel, S., Noilhan, J., Seguí, P.Q., Rousset-Regimbeau, F., Vienne, P., 2008. The SAFRAN-ISBA-MODCOU hydrometeorological model applied over France. *J. Geophys. Res. Atmos.* 113 (6), 1–18.
- Hagedorn, R., Doblas-Reyes, F.J., Palmer, T.N., 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting – i. Basic concept. *Tellus Ser. A Dyn. Meteorol. Oceanogr.* 57 (3), 219–233.
- Haylock, M.R., Hofstra, N., Klein Tank, A.M.G., Klok, E.J., Jones, P.D., New, M., 2008. A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *J. Geophys. Res. Atmos.* 113 (20), 1–12.
- Iglesias, A., Garrote, L., Cancelliere, A., Cubillo, F., Wilhite, D.A., 2012. *Coping with Drought Risk in Agriculture and Water Supply Systems: Drought Management and Policy Development in the Mediterranean*. vol. 26. Springer Science & Business Media.
- Iglesias, A., Garrote, L., Flores, F., Moneo, M., 2007. Challenges to manage the risk of water scarcity and climate change in the Mediterranean. *Water Resour. Manag.* 21 (5), 775–788.
- Johnson, C., Bowler, N., 2009. On the reliability and calibration of ensemble forecasts. *Mon. Weather Rev.* 137 (5), 1717–1720.
- Juanchich, M., Sirota, M., 2015. How to improve people's interpretation of probabilities of precipitation. *J. Risk Res.* 9877, 1–17. (January).
- Lim, Y.K., Cocke, S., Shin, D.W., Schoof, J.T., LaRow, T.E., O'Brien, J.J., 2010. Downscaling large-scale NCEP CFS to resolve fine-scale seasonal precipitation and extremes for the crop growing seasons over the southeastern United States. *Climate Dynam.* 35 (2), 449–471.
- Marcos, R., Llasat, M.C., Quintana-Seguí, P., Turco, M., 2017. Seasonal predictability of water resources in a Mediterranean freshwater reservoir and assessment of its utility for end-users. *Sci. Total Environ.* 575, 681–691.
- Marcos, R., Turco, M., Bedía, J., Llasat, M.C., Provenzale, A., 2015. Seasonal predictability of summer fires in a Mediterranean environment. *Int. J. Wildland Fire* 24 (8), 1076–1084.
- Mason, S.J., Goddard, L., Graham, N.E., Yulaleva, E., Sun, L.S., Arkin, P.A., 1999. The IRI seasonal climate prediction system and the 1997/1998 El Niño event. *Bull. Am. Meteorol. Soc.* 80 (9), 1853–1873.
- Moberg, A., Jones, P.D., 2005. Trends in indices for extremes in daily temperature and precipitation in central and western Europe, 1901–99. *Int. J. Climatol.* 25 (9), 1149–1171.
- Molteni, F., Stockdale, T., Balmaseda, M.A., Balsamo, G., Buizza, R., Ferranti, L., Magnusson, L., Mogensen, K., Palmer, T.N., Vitari, F., 2011. The new ECMWF seasonal forecast system (system 4). *ECMWF Tech. Memo.* 656, 49. (November).
- Nicault, A., Alleaume, S., Brewer, S., Carrer, M., Nola, P., Guiot, J., 2008. Mediterranean drought fluctuation during the last 500 years based on tree-ring data.
- O'Lenic, E.A., Unger, D.A., Halpert, M.S., Pelman, K.S., 2008. Developments in operational long-range climate prediction at CPC. *Weather Forecast.* 23 (3), 496–515.
- Pavón, D., 2001a. L'embassament de Boadella: antecedents d'una obra, transformacions d'un territori. *Revista de Girona* 38–45.
- Pavón, D., 2001b. L'embassament de boadella: antecedents, execució i repercussions locals d'una infraestructura hidràulica. *AIEE* 34, 179–217.
- Pedro-Monzonis, M., Solera, A., Ferrer, J., Estrela, T., Paredes-Arquiola, J., 2015. A review of water scarcity and drought indexes in water resources planning and management. *J. Hydrol.* 527, 482–493.
- Peng, Z., Wang, Q.J., Bennett, J.C., Schepen, A., Pappenberger, F., Pokhrel, P., Wang, Z., 2014. Statistical calibration and bridging of ECMWF system4 outputs for forecasting seasonal precipitation over China. *J. Geophys. Res. Atmos.* 119 (12), 7116–7135.
- Piani, C., Haerter, J.O., Coppola, E., 2010. Statistical bias correction for daily precipitation in regional climate models over Europe. *Theor. Appl. Climatol.* 99 (1–2), 187–192.
- Quan, X., Hoerling, M.P., Whitaker, J.S., Bates, G., Xu, T., 2006. Diagnosing sources of U.S. seasonal forecast skill. *J. Climate* 19 (13), 3279–3293.
- Quintana-Seguí, P., Habets, F., Martin, E., 2011. Comparison of past and future Mediterranean high and low extremes of precipitation and river flow projected using different statistical downscaling methods. *Nat. Hazards Earth Syst. Sci.* 11 (5), 1411–1432.
- Rayner, S., Lach, D., Ingram, H., 2005. Weather forecasts are for wimps: why water resource managers do not use climate forecasts. *Clim. Change* 69 (2–3), 197–227.
- Richardson, D.S., 2000. Skill and relative economic value of the ECMWF ensemble prediction system. *Q. J. Roy. Meteorol. Soc.* 126 (563), 649–667.

- Scaife, A.A., Arribas, A., Blockey, E., Brookshaw, A., Clark, R.T., Dunstone, N., Eade, R., Fereday, D.R., Folland, C.K., Gordon, M., Hermanson, L., Knight, J.R., Lea, D.J., MacLachlan, C., Maidens, A., Martin, M., Peterson, A.K., Smith, D., Vellinga, M., Wallace, E., Waters, J., Williams, A., 2014. Skillful long range prediction of European and North American winters. *Geophys. Res. Lett.* 5, 2514–2519.
- Sene, K., 2010. *Hydrometeorology: Forecasting and Applications*. Springer.
- Stockdale, T.N., 2000. An overview of techniques for seasonal forecasting. *Stoch. Env. Res. Risk A.* 14 (4), 0305–0318.
- Tall, A., Mason, S.J., Van Aalst, M., Suarez, P., Ait-Chellouche, Y., Diallo, A.A., Braman, L., 2012. Using seasonal climate forecasts to guide disaster management: the Red Cross experience during the 2008 West Africa floods. *Int. J. Geophys.* 2012, 1–12.
- Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res. Atmos.* 106 (D7), 7183–7192.
- Turco, M., Llasat, M.C., 2011. Trends in indices of daily precipitation extremes in Catalonia (NE Spain), 1951–2003. *Nat. Hazards Earth Syst. Sci.* 11 (12), 3213–3226.
- Turco, M., Llasat, M.C., Herrera, S., Guti, J.M., 2016. Bias correction and downscaling future RCM precipitation projections using a MOS-Analog technique. *J. Geophys. Res. Atmos.* 122 (5), 1–35.
- Turco, M., Quintana-Seguí, P., Llasat, M.C., Herrera, S., Gutiérrez, J.M., 2011. Testing MOS precipitation downscaling for ENSEMBLES regional climate models over Spain. *J. Geophys. Res. Atmos.* 116 (D18), 1–14.
- Watkins, D.W., Wei, W., 2008. The value of seasonal climate forecasts and why water managers don't use them. *World Environmental and Water Resources Congress 2008* 316. pp. 1–5.
- Weigel, A.P., Liniger, M. a., Appenzeller, C., 2007. Generalization of the discrete Brier and ranked probability skill scores for weighted multimodel ensemble forecasts. *Mon. Weather. Rev.* 135 (7), 2778–2785.
- Weigel, A.P., Liniger, M. a., Appenzeller, C., 2009. Seasonal ensemble forecasts: are recalibrated single models better than multimodels? *Mon. Weather. Rev.* 137 (4), 1460–1479.
- Wilhite, D.A., Hayes, M.J., Knutson, C., Smith, K.H., 2000. Planning for drought: moving from crisis to risk management. *J. Am. Water Resour. Assoc.* 36 (4), 697–710.
- Wilks, D.S., 2006. *Statistical Methods in the Atmospheric Sciences*. Academic Press, London.
- GFCS, 2016. Use of climate predictions to manage risks. *Tech. Rep.* 1174. World Meteorological Organization.,