Potential predictability of an Iberian river flow based on its relationship with previous winter global SST

S.R. Gámiz-Fortis a, b, M.J. Esteban-Parra a, R.M. Trigo a, b, c, Y. Castro-Díez a

a Applied Physics Department, University of Granada, Granada, Spain
b University of Lisbon, CCGIL, IDI, Lisbon, Portugal
c Departamento de Engenharias, Universidade Lusófona, Lisboa, Portugal

INTRODUCTION

In recent years, interest in seasonal predictability of river discharge variability over Europe has increased markedly (Trigo et al., 2004; Rimbu et al., 2004, 2005; Gámiz-Fortis et al., 2008a,b; Ionita et al., 2008). The hydrological system acts as a spatial and temporal integrator of precipitation (rain and snow), temperature, and related evapotranspiration over a specific region. Therefore, seasonal to interannual streamflow variability in many large river-basins can be controlled by corresponding changes in large scale oceanic and atmospheric circulation patterns (Rimbu et al., 2004, 2005). Characterisation of hydrological variability on climatic timescales and identification of connections to climate forcings may provide potential improvement for building hydrological models. Such preliminary steps are required if one aims to produce useful seasonal forecasts of river flow amounts which have enormous relevance in water resources management such as in hydroelectric power production and agriculture crop yields. Additionally, in the climate change context, some studies for the Iberian Peninsula points to a general increase in the risk of winter and summer droughts for the 21st century with increasing uncertainty in the reliability of water supplies (IPCC, 2007; Giorgi and Bi, 2005). The provision of seasonal forecast for hydrological climate variables could help to alleviate some of the negative impact of climate change on water infrastructure, namely through increased preparedness (Wedgebrow et al., 2002).

The availability of water is mostly influenced by climate conditions that vary on seasonal, interannual, and decadal time scales. On seasonal timescales, anomalous atmospheric conditions are often linked with seasonal variations in the rivers streamflow and reservoir storages, via variations in precipitation and temperature (e.g. Dettinger and Diaz, 2000; Cullen et al., 2002; Trigo et al., 2004; Karabörk et al., 2005; López-Moreno et al., 2007). Usually, the skill of these long-range forecasts is associated with the introduction of predictors that represent the slow varying components of the climate system such as sea ice, snow cover, soil moisture, SSTs and major oceanic and atmospheric circulation patterns such as the ENSO, the NAO in Europe or the PNA in North America (Güetter and Georgakakos, 1996; Wedgebrow et al., 2002). However, the relationship between the climatic regimes and its hydrological response, through its streamflow, presents different
grades of complexity according to the physical characteristics of the basin.

In a previous work the authors showed that the river flow regime in the major Iberian Peninsula rivers is controlled by the highly seasonal precipitation regime (Trigo et al., 2004). Similar to other Mediterranean regions winter- and spring-time river flow account for the majority of runoff, this being followed by a relatively long and dry summer period (Daveau, 1988; INAG, 2001). Because the NAO impact in Iberian Peninsula precipitation is particularly strong in winter, most studies have been focused to predict the river flow in this season (Trigo et al., 2004; Gámiz-Fortis et al., 2008a,b). However studies developed for other regions of Europe clearly show potential predictability of rainfall and temperature during other seasons of the year, namely spring and fall (Shongwe et al., 2007; Ionita et al., 2008). The most relevant predictors responsible for such links have range from anomalous conditions of sea surface temperature in the North Atlantic (Colman, 1997; Colman and Davey, 1999; Wilby, 2001) and/or the tropical Pacific (Kladis and Diaz, 1989; Lloyd-Hughes and Saunders, 2002; van Oldenborgh et al., 2000; Rimbu et al., 2005; Ionita et al., 2008). In this context, the need for spring river flow modelling in Iberia is particularly acute, because it could provide useful insight on the water availability and managers, possibly related to agriculture and hydropower generation, prior to extremely dry summer months.

This paper presents a modelling scheme for spring Douro streamflow anomalies based on the combination of two methodologies: (1) the identification of stable teleconnections between oceanic SST anomalies and river flow and (2) the use of ARMA models to predict quasi-oscillatory modes of the flow.

Data

The Douro river presents the most extensive basin within the Iberian Peninsula with an area of 97,290 km$^2$ (Fig. 1). Moreover with 897 km in length it corresponds to one of the longest rivers in Iberia (second only to Tagus), running from the Iberian Mountains in north-eastern Spain down to the Atlantic Ocean across Portugal. There is a lack of proportionality between area, runoff and volume in the two countries due to the fact that the Douro Basin is a complex region from both the topographical and meteorological perspectives. This region includes prominent mountains, several mountain ranges, hilly areas, valleys and tributary streams. The analysis of temperature, precipitation and insolation fields in the global Douro Basin reveals the existence of four main climatic regions (Leite and Peixoto, 1995). A western region dominated by oceanic air masses with heavy precipitation, on contrast with a central and eastern region drier and warmer. Two more peripheral regions were identified, the Cantabric Cordillera in the North and Central System, where the relief is the dominant factor. Based on these topographical considerations it is expectable that the influence of anomalous SST on the river streamflow to be more important over the western sector of the Douro Basin. Additionally, the Douro river is subjected to large regulation in the Spanish part of the basin, which can interfere the SST/river discharges relationship. Note that the expected effect of large dams is an increased time between precipitation episodes and the arrival of the corresponding flow to lower sections of the river basin. Trigo et al. (2004) shown that there is a significant correlation between the winter NAO index and the winter Douro river flow, especially high since the 1970s, and that the correlation increase is larger at 1-month lag than it is for simultaneous correlations. The strength of this correlation is not just related to the observed strengthening of the NAO relationship with the winter precipitation in the western part of the Iberian Peninsula during the last part of the 20th century, but also is related to the increase in water storage volume associated with the construction of major dams in the 1950s and 1960s (Daveau, 1988; Melo and Gomes, 1992).

Based on these considerations it is understandable that the contribution from the SST to the river streamflow to be more important for the western part of the Douro Basin. For this reason the monthly time series of Douro discharge, used in this paper, were recorded at Pocinho (37.18°N, 7.55°W), which is situated in the Portuguese side of the border, in the lower part of the Douro catchment area (Fig. 1). Data were kindly provided by the Portuguese National Electrical Supply Company (REN) and is restricted to the period 1956–2006, that was intensively checked for inconsistencies. Fig. 2 presents the seasonal variability of the mean monthly flow for the Douro river along the entire period of available data. As expected for this region, winter and spring-time river flow account for the majority of runoff, being followed by a relatively long and dry summer period from July to September (Trigo et al., 2004). Increased precipitation in winter causes the increase in water flow with the peak value being observed between January and February. When the summer drier conditions arrive at the beginning of June, the stream flow decreases rapidly towards July, August and September where it reaches the annual minimum value.

Hydrological series commonly do not follow a normal distribution, being highly biased, often requiring some preliminary transformation in order to adjust the records to a more appropriate probability distribution. Previous studies on precipitation found gamma distributions suitable for the larger part of Europe (Lloyd-Hughes and Saunders, 2002). For data from Northwest Europe, Zaidman et al. (2001) showed a relatively good adjustment of the discharge series to the log-normal distribution, while the Pearson III distribution has been found more appropriate over some
parts of the Iberian Peninsula (Vicente-Serrano, 2006; López-Moreno et al., 2009). Using these considerations we tested three different theoretical distributions for modelling the monthly stream river flow of May and June: the Pearson III, the log-normal and the normal distribution. Furthermore, the goodness of fit was evaluated using three different tests, the Kolmogorov–Smirnov test, the Anderson–Darling test and the Chi-squared. We find that, although the three theoretical distributions present a good result for spring months, the Pearson III distribution offers better results (based on the Kolmogorov–Smirnov and Anderson–Darling tests), while the log-normal distribution is better according to the Chi-squared test. Thus, taking into account these results, the original monthly time series were standardized using the Pearson III distribution and also averaged on a seasonal basis to generate the spring flow anomaly time series.

A preliminary inspection of the monthly data shows that for April riverflow anomalies, the correlation is much higher with March flow anomalies ($r = 0.70$) than with May ($r = 0.40$). Additionally, the atmospheric patterns associated to March and April streamflow time series (not shown) are very similar, while they change in May. However, time series of May and June show very high correlation ($r = 0.81$). For this reason we averaged the monthly anomalies over the months May and June (MJ) to generate the spring flow anomaly time series.

As predictor we use the global SST taken from the HadISSTv1.1 data set (Rayner et al., 2003) derived from the Hadley Centre for Climate Prediction and Research (UK Meteorological Office). We generate the winter, spring, summer and autumn SST fields by averaging the monthly SST anomalies (using the mean and standard deviation for the period 1961–1990) for DJF, MAM, JJA and SON, respectively.

### Methodology

Following the approach adopted by other authors (e.g. Lohmann et al., 2005; Rimbu et al., 2005; Ionita et al., 2008) the first step requires to identify sectors of oceanic SST anomalies that can be used as predictors for Douro’s river flow. In order to do that we have evaluated the point linear correlation between the spring streamflow anomalies and the global SST anomalies from previous seasons. Regions showing significant correlations are identified as potential predictors. The second step is to identify, among these regions, those that can be classified as stable predictors. This is achieved through the analysis of the variability of the correlation between spring Douro flow anomalies and SST anomalies from potential predictor regions using a moving window of 15 years. Additional analysis using window lengths of 10 and 20 years were carried out in order to examine the stability of the connections in time. Following the criterion of Ionita et al. (2008) the correlation is considered to be stable for those regions where spring streamflow and SST anomalies are significantly correlated at 90% level ($r = 0.44$) for more than 80% of the 15-year windows covering the period 1956–2006 and, furthermore, that the sign of the correlation does not change with time. Regions verifying this criterion are considered as robust predictors which will be used in a multiple linear regression model (hereafter SST\_model) to simulate the spring Douro river flow anomalies. Additionally, in order to detect some other kind of influence, we study the residual time series, which is obtained by subtracting the SST\_model from the raw streamflow series. Monte Carlo Singular Spectral Analysis (MCSSA) is applied to this remainder in order to detect quasi-oscillatory modes that could be associated with other parts of the Ocean (Allen and Smith, 1996). Finally an ARMA\_model is fitted to the residual time series filtered by MCSSA (named residual_filter) and the improvement obtained by the combination of SST\_model and ARMA\_model is evaluated.

The separation of calibration and validation periods is fundamental for reliable skill assessment (Wilks, 1995). We employ data from 1956 to 1989 to calibrate the Douro model, while data from 1990 to 2006 are used for validation purposes. We would like also to apply cross-validation using development data set of size n – 1 and verification data set containing the remainder single observation of the predictand, leading to n partitions of the dataset. The model is then calculated for each of these partitions, resulting in n similar forecast equations, each one computed without one of the observations of the predictand. Unfortunately, this procedure cannot be applied in our case for a number of reasons, namely because when fitting ARMA models, the temporal location of each data cares: i.e. the “history” of the series is very important. When removing one single data in the middle of the data set, the remaining data are not useful to fit the model, because the temporal structure of the data is then broken (Gámiz-Fortis et al., 2002). For model evaluation we employ two appropriate skill measures, namely the percentage improvement in the root-mean-square error over a climatological forecast (RMSEcl) and over persistence (RMSEper). Climatological is taken as the standardized long-term average prior to each year being forecasted, while persistence is taken as winter (JFM) Douro streamflow standardized anomalies. Additionally, the Willmott agreement index ($D$) is also computed (Willmott, 1982). $D$ is a dimensionless measure of relative error in model estimates which ranges from 0 (complete disagreement between estimated and observed values) to 1 (complete agreement). This combined methodology (SST\_model and ARMA\_model) has already been used by the authors in a previous study concerning the predictability of Iberian Peninsula winter river flows (Gámiz-Fortis et al., 2008a,b) obtaining very useful results.

### Results

The spatial correlation maps between spring flow and the preceding seasons for SST data are shown in Fig. 3. According to these maps we can say that low correlation values (not significant) began to appear during the previous spring in the north-western Atlantic and during the previous autumn in the south-western Atlantic. Some propagation occurs until reaching maximum correlation values in two regions for winter SST, one close to Brazil in the south-western part of Atlantic Ocean, and another situated in the North Atlantic between Iberia and North America. Two smaller regions located to the north of the Indian Ocean and in the middle of the southern Pacific Ocean also show significant correlations. Based on this correlation map we define three SST indices by averaging the normalized SST anomalies over the regions showing maximum correlation values: SST1 ($35^\circ$W–$25^\circ$W; $15^\circ$S–$10^\circ$S), SST2 ($45^\circ$W–$25^\circ$W; $38^\circ$N–$42^\circ$N) and SST3 ($85^\circ$E–$95^\circ$E; $10^\circ$N–$15^\circ$N). Fig. 4 shows the correlation between the spring streamflow anomalies and the SST1, SST2 and SST3 indices using a moving window of 15 years. Stable correlations according to the criterion adopted in the methodology are found only for the two first indices, positive for the SST1 and negative for the SST2. Similar results (not shown), showing the good stability of these connections, are found for the additional 10-years and 20-years running correlation analysis. The SST3 index fails the stability test and is dismissed from the rest of the analysis. No stable correlations are found between the significant areas associated with previous spring, summer and autumn SST, for what they were excluded from the analysis. It should be noted that a similar study using the land surface temperature and precipitation variables (instead of the SST) does not reveal additional regions showing stable correlations with the spring Douro river flow (maps not shown).
Using the significant and stable indices SST1 and SST2 as predictors for the spring Douro river flow we developed a model based on multiple linear regression, using the calibration period (1956–1989). The optimal model for explaining spring streamflow can be written as:

\[ \text{SST}\_\text{model} = -0.60 \times \text{SST2} + 0.43 \times \text{SST1} - 0.45 \]

The observed and modelled spring Douro streamflow series are shown in Fig. 5 for calibration and validation periods. It should be noticed that the predictors adopted in this study (SST1 and SST2) are independent of each other as shown by their non-significant correlation \((r = -0.11)\). Results obtained reveal a considerable skill achieved by the SST\_model (see Table 1), with a good correlation coefficient between the raw series and the model for the validation period \((r = 0.63)\) and coefficient of multiple determination \(R^2 = 0.50\). Moreover the model presents a relatively low MSE = 0.48, MAE = 0.59 and a relatively high Willmott’s index \((D = 0.79)\). Finally an interesting skill against climatology (persistence) of 30% (64%) can be observed. Skill score values achieved by the model during the calibration period are very similar to those attained during validation.

An additional analysis is carried out using the residual time series (residual = flow \(-\text{SST}\_\text{model}\)) in order to improve the modelling of the spring Douro river flow variability. MESSA applied to the residual shows three significant quasi-oscillatory modes with periods around 5, 3 and 2.4 years. The residual filter time series computed like the sum of the reconstructed components of these oscillatory modes explains 60% of the total variance of the residual time series.

As we expected the correlation map between the residual filter and the previous winter SST field does not show significant values.
(not shown). However, significant correlation values are found between the individual quasi-oscillatory mode with period around 3 years and the previous winter SST in the region of El Niño3 (90°W–150°W; 5°S–5°N), and the individual quasi-oscillatory mode with period around 5 years and the previous spring SST in the region of El Niño3.4 (120°W–170°W; 5°S–5°N), (see Fig. 6). Correlation coefficient between the 3 years oscillation and the previous winter SST anomalies in the region of El Niño3 is 0.40, while it is 0.3 between the 5 years oscillation and the previous spring SST anomalies in the region of El Niño3.4. An additional MCSSA study to the winter SST in the El Niño3 region shows two significant oscillatory modes with periods around 5.3 and 3.4 years explaining a high fraction of the total variance of the series (41.2%). However, no significant correlation values are found between the 2.4 years oscillatory mode and previous seasonal SST.

Using the sample Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) (not shown) we find that while the raw and residual data behave like a white noise process, the residual filter series shows a strong autocorrelation pattern. This feature implies a higher predictability of the residual filter when compared to the unfiltered time series.

Based on these analyses, we used the Akaike Information Criterion (AIC) to select an ARMA(4,3) model for the residual filter, containing the following parameters:

\[
AR = (\phi_1 = -0.21, \phi_2 = -0.32, \phi_3 = -0.33, \phi_4 = -0.34)
\]

\[
MA = (\theta_1 = 0.76, \theta_2 = 0.85, \theta_3 = -0.87)
\]

Significance of the parameters was computed using approximate t-values, derived from the parameter standard errors. Parameters highlighted with “*” are statistically significant at 5% significance level. We can see that the overall quality of the model has increased significantly for both calibration and validation periods (see Table 1). For the validation period the correlation coefficient between the raw series and the new model of 0.79 and the variance explained by the combined models rising to 76%. Furthermore, the MSE and MAE values are considerably lower (0.31 and 0.48, respectively) than those obtained for the SST model. Finally this combined model presents better skill scores against climatology (55%) and against persistence (77%), and the Willmott’s index has been improved until 0.88.


**Discussion and conclusions**

We have investigated the predictability of the spring Douro flow anomalies using as predictor the sea surface temperatures from the preceding seasons. Only two key regions located in the north and south halves of the Atlantic Ocean are found to hold winter SST anomalies that are significantly (and consistently) correlated with spring Douro flow anomalies. The corresponding indices computed averaging the SST anomalies in these regions are used as explanatory variables in a multiple linear regression model explaining a significant proportion of the variance (R² = 0.50) of the spring Douro river flow. The modelling of Douro flow anomalies based on our statistical scheme outperforms significantly the modelling based on climatology and persistence, with most of the contribution coming from the middle latitudes in the North Atlantic Ocean. Similar studies using the SLP, land surface temperature and precipitation fields (not shown) were also performed but without providing meaningful additional predictors. We have found a centre of significant negative correlations values between the spring Douro river flow anomalies and the preceding winter SLP anomalies over the Iberian Peninsula, placed to the east of the SST2 region. The correlation between this SLP anomaly centre and the winter NAO index is 0.87. This associated SST–SLP pattern is in agreement with the idea that the atmosphere is inducing the SST anomalies (Bjerknes, 1962; Luksch et al., 1990; Cayan, 1996 and 2001.)

**Table 1**

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SST_model</td>
<td>SST_model + ARMA (4,3)</td>
</tr>
<tr>
<td>MSE</td>
<td>0.54</td>
</tr>
<tr>
<td>MAE</td>
<td>0.55</td>
</tr>
<tr>
<td>Correlation coeff.</td>
<td>0.70</td>
</tr>
<tr>
<td>Willmott index (D)</td>
<td>0.78</td>
</tr>
<tr>
<td>MSE Visibility</td>
<td>0.91</td>
</tr>
<tr>
<td>MSE Visibility</td>
<td>1.24</td>
</tr>
<tr>
<td>% SMSE Visibility</td>
<td>41</td>
</tr>
<tr>
<td>% Phase accordance</td>
<td>70</td>
</tr>
</tbody>
</table>

**Fig. 6.** Correlations maps between (a) the quasi-oscillatory mode with period around 5 years of the residual filter and previous spring SST anomalies and (b) the quasi-oscillatory mode with period around 3 years of the residual filter and previous winter SST anomalies. Contours indicate significant values at the 95% confidence level.
In the presence of a westerly background flow, an anomalous cyclonic flow in the atmosphere will be connected with anomalous cooling of the SST at the cyclone’s western and southern flank and warming at its eastern flank, due to anomalous latent and sensible heat fluxes.

Additionally, the residual time series shows some quasi-oscillatory modes with periods around 5 and 3 years that seem to be associated with the SST from the ENSO region. However, the 2.4 years oscillatory mode does not present significant correlations with seasonal SST. This result is suggesting that other influences, besides the SST, must be considered. In a previous work the author has identified a significant oscillatory mode with period around 2.4 years in the winter NAO index series (Gámiz-Fortis et al., 2002). The ARMA modelling applied to the filtered residual is able to provide the interannual linearly predictable signal contained in the history of the time series which is not related to the Atlantic SST. It could be argued that the ARMA model provides part of the low frequency (interannual) useful information for the modelling ($R^2 = 0.17$) that may result from the low frequency relationship between the Pacific SST and the precipitation/streamflow. The effect of ENSO on climate in Europe has been studied intensively during the past few years using both models (Merkel and Latif, 2002; Mathieu et al., 2004) and observational data (Ropelewski and Halpert, 1987; Fraedrich et al., 1992; Gouirand and Moron, 2003; Pozo-Vázquez et al., 2005; Vicente-Serrano, 2005; Mariotti et al., 2005; Brönnimann et al., 2004, 2007). Most studies suggest a weak but significant ENSO response over the North Atlantic–European sector, but there remain considerable uncertainties regarding the regional details and quantitative aspects of the response. Previous studies have shown that there is an asymmetrical impact associated to El Niño and La Niña events on the European climate, and that such behaviour is most probably induced by non-linear mechanisms (Fraedrich, 1990; Wilby, 1993; Pozo-Vázquez et al., 2001). Strong La Niña events during autumn, which usually persists during the following winter, give a detectable signal in winter precipitation over Europe (Pozo-Vázquez et al., 2005) with negative anomalies in southern Europe, resembling the precipitation pattern associated with the positive phase of the North Atlantic oscillation. Particularly, during the final months of La Niña year, and during the initial months of the following year, wide areas of the Iberian Peninsula show significant negative precipitation anomalies (Vicente-Serrano, 2005). Such winter negative precipitation anomalies could be recorded as Douro flow anomalies in the following spring, and could be the basis for the idea of a standing wave train that propagates downstream to the North Atlantic area and, particularly, can give rise to the stable SLP anomaly pattern resembling the positive phase of the NAO found to be associated with the cold phase of the ENSO. Evidence for this hypothesis is found in the work by Wu and Hsieh (2004), which showed the existence of a non-linear response to the ENSO events on the Northern Hemisphere circulation; the response consists of the excitation of the positive PNA and NAO patterns. Nevertheless, this hypothesis must be considered cautiously, since other works, such as Strauss and Shukla (2002), argued that the effect of the ENSO events of the extratropical atmospheric circulation is to force distinct mid-latitude patterns rather than to modify the probability of the internal variability patterns (such as the PNA). It appears, therefore, that the response of the circulation in the North Atlantic region to ENSO events results from mixed influences, particularly of tropical Atlantic forcing related to the tropical Pacific SST anomalies and the mid-latitude atmospheric forcing through the PNA teleconnection. In summary, there is considerable uncertainty regarding the details of the ENSO influence on the North Atlantic-European region, both in terms of the spatial structure of anomalies and the magnitude of the signal. An additional issue to take account is that every ENSO event is unique and different events will have different impacts. This is not simply because of internal atmospheric variability but rather because the precise ocean conditions differ between events (Mathieu et al., 2004).

Results for spring Douro river flow are essentially different from those found in winter (Gámiz-Fortis et al., 2008a,b). Although negative correlation is found between the SST from previous seasons around 40°N and winter and spring Douro river flows, an additional influence is found for spring flow associated with SST anomalies in the south-western part of the Atlantic Ocean. This could be related with the non-linear influence, found for the winter Douro river flow, with summer SST in the tropical south Atlantic. Another important difference is that, contrary to winter, the seasonal linear predictability is considerable larger than the interannual linear predictability (Gámiz-Fortis et al., 2008b).

Our analysis shows that contrary to other variables such as land surface temperature or precipitation, winter SST anomalies from several Atlantic regions provide a significant source of predictability for the following spring Douro flow variability. Also, these regions contribute with an important part of the explained variance. Because the SST from these regions is readily available this represents an important source of predictability that can be further exploited at present aiming to develop forecasting approaches useful to water managers in the Douro river basin.

Acknowledgements

The Spanish Ministry of Science and Innovation, with additional support from the European Community Funds (FEDER), Project CGL2007-61151/CLI, and the Regional Government of Andalusia Project P06-RNM-01622, financed this study. SR Gámiz-Fortis is supported by the University of Granada under a postdoctoral contract.

References


