A comparison of modeling techniques to predict hydrological indices in ungauged rivers

Francisco J. Peñas1,2,*, Jose Barquín3 and César Álvarez3

1 Department of Civil Engineering. Universidad Católica de la Santísima Concepción, Alonso de Ribera 2850, Concepción, Chile.
2 Centro de Investigación en Biodiversidad y Ambientes Sustentables. “CIBAS”. Universidad Católica de la Santísima Concepción, Alonso de Ribera 2850, Concepción, Chile.

* Corresponding author: fpenas@ucsc.cl

Received: 23/06/17 Accepted: 11/12/17

ABSTRACT
A comparison of modeling techniques to predict hydrological indices in ungauged rivers

Predicting the natural flow regime in ungauged rivers is an important challenge in water resource management and ecological research. We developed models to predict 16 hydrological indices in a river network covering the northern third of the Iberian Peninsula. Multiple Linear Regression (MLR), Generalized Additive Models (GAMs), Random Forest (RF) and Adaptive Neuro Fuzzy Inference System (ANFIS) were used and compared according to their prediction accuracy. The results showed that predictive performance varied greatly depending on the modeled hydrological attribute. The magnitude and frequency indices were predicted with excellent accuracy. In contrast, no technique was capable of developing precise models for hydrological indices of timing, duration and rate of change. This is mainly related to the lack of proper environmental databases on the scales on which these flow regime patterns are influenced. In addition, complex modeling techniques did not always outperform linear models and no single approach was optimal for all indices. ANFIS and GAMs provided the best results; however, other issues such as computational cost and the level of knowledge required to apply the method and interpret the results should be taken into account.

Key words: natural flow regime, prediction, linear regression, generalized additive models, machine learning

RESUMEN
Comparación de técnicas de modelado para predecir índices hidrológicos en ríos no aforados

La predicción del régimen natural de caudales en ríos no aforados representa un problema esencial para superar los nuevos retos a los que se enfrenta la gestión de los recursos hídricos y la ecología de los sistemas de agua dulce. En este trabajo hemos desarrollado modelos para predecir 16 índices hidrológicos en la red fluvial que cubre el tercio norte de la Península Ibérica. En concreto se han desarrollado y comparado Regresiones Lineales Múltiples (RLM), Modelos Aditivos Generalizados (GAMs), Bosques Aleatorios (BA) y Sistemas Adaptativos de Inferencia de Lógica Difusa (SAILD). Los resultados han puesto de manifiesto que la capacidad predictiva varía significativamente dependiendo del tipo de índice hidrológico modelado. Los modelos de los índices de magnitud y frecuencia mostraron una capacidad predictiva excelente. Por el contrario, los modelos de los índices hidrológicos relacionados con la temporalidad, la duración de periodos de caudales altos o bajos y la tasa de cambio mostraron una capacidad de predicción limitada. Esto se relaciona, en gran medida, con la falta de bases de datos de variables predictoras con escalas espacio-temporal adecuadas. Por otro lado, las técnicas estadísticas más complejas no siempre mostraron capacidades predictivas mayores que los RLM y, además, no se encontró un método que ofreciese resultados óptimos para todos los índices. SAILD y GAM obtuvieron, por norma general, los mejores resultados, sin embargo, consideramos que otros elementos, tales como los recursos computacionales requeridos o la experiencia...
INTRODUCTION

River flow regime is a key element that structures freshwater ecosystems (Poff et al., 1997). Indeed, the understanding of the bio-physical associations between hydrological variability and stream biological communities is a critical scientific and management challenge (Alvarez-Cabria et al., 2017). However, it is frequently the case that streamflow data are not available at a site of interest such as where biomonitoring is carried out (Poff & Zimmerman, 2010; Sanborn & Bledsoe, 2006). This hinders the exploration of the flow regime influence on stream ecology and ultimately the management of these systems.

Natural flow regime can be described through a collection of ecologically relevant hydrological indices (Olden & Poff, 2003). Hence, interest in the prediction of these hydrological indices in ungauged streams has grown rapidly in recent years (Carlisle et al., 2010; Kennen et al., 2008). Most of the work has been aimed at addressing water yield and flooding issues. Thus, models to predict average flows, flood quantiles, flow duration curves or low-flow parameters dominate the literature (Sanborn & Bledsoe, 2006). In contrast, prediction of ecologically relevant hydrological indices has received limited attention (Carlisle et al., 2010; Knight et al., 2011; Sanborn & Bledsoe, 2006).

Multiple linear regression has been the most commonly used statistical technique to predict hydrological indices in ungauged sites (Knight et al., 2011). However, the potential improvement in model performance when using other modeling procedures that do not assume specific distribu-

Figure 1. Map of unregulated gauges (●; n=156) in the study area. Black lines divide the Cantabrian, Ebro and Catalan catchments. (CS: Cantabrian Sea; MS: Mediterranean Sea). Mapa de aforos no regulados en el área de estudio (●; n=156). Las líneas negras dividen las cuencas del Cantábrico, del Ebro y de Cataluña (CS: Mar Cantábrico; MS: Mar Mediterráneo).
tion fittings has been pointed out (Sanborn & Bledsoe, 2006). There are many examples of the use of other modeling and machine learning techniques to model many environmental issues (e.g., Alvarez-Cabria et al., 2016; Elith et al., 2006; Manel et al., 1999, Marcé et al., 2004). In contrast, their application in the prediction of hydrological indices has been limited, although they could provide important benefits in this field (Alcázar et al., 2008; Heuvelmans et al., 2006; Snelder et al., 2009).

In this study we concentrated on developing statistical models for 16 hydrological indices covering the 5 ecologically relevant hydrologic attributes (i.e., magnitude, timing, frequency, duration and rate of change; Poff et al., 1997). We used one traditional technique (Multiple Linear Regression (MLR)) and three more complex techniques that apply contrasting rationale to model the distribution of the response variable: Generalized Additive Models (GAMs), Random Forest (RF) and Adaptive Neuro-Fuzzy Inference System (ANFIS). Therefore, the objectives of this study were to 1) explore the ability of models to predict different types of hydrological indices and 2) compare the performance of 5 modeling techniques to predict 16 hydrological indices at ungaged sites.

METHODS

Study Area

The study area comprises the catchments of the northern third of the Iberian Peninsula (Fig. 1), covering a total area greater than 124 000 km². It includes a heterogeneous set of environmental conditions.

The area draining into the Cantabrian Sea encompasses several catchments with drainage areas ranging from 30 to 4907 km², covering a total area of 22 000 km². The rivers are confined by the Cantabrian Cordillera, which reaches up to 2600 m a.s.l. and runs parallel to the coast. Thus, they are characterized by high slopes and short main stream lengths. The climate varies from thermo-temperate Atlantic on the coast to omo-and supra-temperate in the inner regions (Rivas-Martínez et al., 2004). Precipitation is abundant throughout the year with a mean of 1300 mm/year, with maximum rainfall in December (150 mm/month) and minimum in July (50 mm/month). Snowfall is frequent in winter above 1000 m a.s.l. More than 50 % of the surface is covered by deciduous forest, scrub and grasslands, while 10 % is occupied by agriculture.

Meanwhile, the Mediterranean area is mainly covered by the Ebro catchment, along with a set of medium-sized basins in the east coast. The Ebro catchment covers a total area of 85 530 km². It is enclosed by the Cantabrian Mountains and the Pyrenees (3400 m a.s.l.) in the north, the Catalan Coastal Chain (1712 m a.s.l.) in the east, and from the north-west to the south-east by the Iberian massif (2300 m a.s.l.), which creates a dense river network in the catchment boundaries and an extended flat surface in the interior. This area is characterized by a meso-Mediterranean and supra-Mediterranean climate (Rivas-Martínez et al., 2004), with a mean annual precipitation of 650 mm, varying from 300 mm in the central area of the main fluvial axis to 1700 mm in the Pyrenees Mountains, where snow is abundant in winter and early spring (Bejarano et al., 2010). The precipitation regime in the Mediterranean region has its maxima in autumn and spring and minima in winter and summer. Agricultural land accounts for 50 % of this territory.

The Catalan catchments comprises several catchments ranging from 72 to 5000 km², covering a total area of 16 500 km² that drains directly from the Pyrenees or the Catalan Coastal Chain to the sea. This area is dominated by the Mediterranean oceanic climate on the coast and a temperate climate in the mountains. Precipitation declines from an annual mean of 1200 mm/year at the northern river heads to less than 500 mm/year in the southern catchments. Coniferous and broadleaf forest, scrub and grasslands occupy more than 60 % of the surface in the northern catchments, which are progressively replaced by agricultural land in the south.

Hydrologic Data and Hydrological Indices

Several Spanish water agencies and regional governments provided series of daily mean flow measured at 428 gauging stations. Only gauges

Table 1. Hydrological indices for which models were developed and their type of hydrological attribute (MA: Magnitude Average; MH: Magnitude High; ML: Magnitude Low; T: Timing; F: Frequency; DH: Duration of high-flow events; DL: Duration of low-flow periods; RC: Rate of change). Non-compliance with the MLR assumptions is indicated by the following superscripts: n: normality; h: homoscedasticity; i: independence. Índices hidrológicos modelados y el tipo de atributo al que hacen referencia (MA: Magnitud media; MH: Magnitud Altos; ML: Magnitud Bajos; T: Temporalidad; F: Frecuencia; DH: Duración de eventos de caudal alto; DL: Duración de periodos de caudal bajo; RC: Tasa de cambio). Se indica el no cumplimiento de las asunciones de MLR mediante los superíndices: n: normalidad, h: homocedasticidad, i: independencia.

<table>
<thead>
<tr>
<th>Index</th>
<th>Type</th>
<th>Units</th>
<th>Description</th>
<th>MLR transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>MA</td>
<td>m³/s</td>
<td>Linear moment that represents the mean daily annual flow</td>
<td>x₁/₅</td>
</tr>
<tr>
<td>L2</td>
<td>MA</td>
<td>m³/s</td>
<td>Linear moment that represents the variance of the daily annual flow.</td>
<td>x₁/₅ n</td>
</tr>
<tr>
<td>M4</td>
<td>MA</td>
<td>m³/s</td>
<td>Mean daily April flow</td>
<td>x₁/₅ n</td>
</tr>
<tr>
<td>M9</td>
<td>MA</td>
<td>m³/s</td>
<td>Mean daily September flow</td>
<td>x₁/₅</td>
</tr>
<tr>
<td>30LF</td>
<td>ML</td>
<td>m³/s</td>
<td>Magnitude of minimum annual flow of 30-day duration.</td>
<td>x₁/₆ n</td>
</tr>
<tr>
<td>X95</td>
<td>ML</td>
<td>m³/s</td>
<td>Mean magnitude of flow exceeded 95% of the time</td>
<td>x₁/₄</td>
</tr>
<tr>
<td>30HF</td>
<td>MH</td>
<td>m³/s</td>
<td>Magnitude of maxima annual flow of 30-day duration</td>
<td>x₁/₅</td>
</tr>
<tr>
<td>X5</td>
<td>MH</td>
<td>m³/s</td>
<td>Mean magnitude of flow exceeded 5% of the time</td>
<td>x₁/₆ n</td>
</tr>
<tr>
<td>Jmax</td>
<td>T</td>
<td>Day of year</td>
<td>Julian day of annual maximum</td>
<td>None</td>
</tr>
<tr>
<td>Jmin</td>
<td>T</td>
<td>Day of year</td>
<td>Julian day of annual minimum</td>
<td>None</td>
</tr>
<tr>
<td>Pred</td>
<td>T</td>
<td></td>
<td>Predictability</td>
<td>log(x+1) n, h</td>
</tr>
<tr>
<td>FRE3</td>
<td>F</td>
<td>Events/year</td>
<td>Number of high-flow events per year using an upper threshold of 3 times the median flow over all years</td>
<td>None</td>
</tr>
<tr>
<td>dPHigh</td>
<td>DH</td>
<td>Days</td>
<td>Duration of high-flow pulses</td>
<td>log(x+1) n, h</td>
</tr>
<tr>
<td>dPLow</td>
<td>DL</td>
<td>Days</td>
<td>Duration of low-flow pulses</td>
<td>x₁/₆</td>
</tr>
<tr>
<td>nPos</td>
<td>RC</td>
<td>Days</td>
<td>Number of days with increasing flow</td>
<td>log(x+1)</td>
</tr>
<tr>
<td>nNeg</td>
<td>RC</td>
<td>Days</td>
<td>Number of days with decreasing flow</td>
<td>none</td>
</tr>
</tbody>
</table>
Comparison of techniques to model hydrological indices

unaffected by impoundments or significant abstraction upstream were selected for analysis. In addition, we selected gauges with data available for the 1976-2010 period and analyzed the quality of the series (Peñas et al., 2014). Finally, 156 gauges were selected, which accounted for an average length of 17 years of data (Fig. 1).

It was beyond the scope of this study to predict and evaluate all the hydrological indices currently in use (see Olden & Poff, 2003); therefore, we selected one or several indices representing each of the five ecologically relevant aspects of the flow regime, i.e., magnitude, timing, frequency, duration and rate of change (Olden & Poff, 2003; Table 1).

Table 2. Environmental variables used in the models of the 16 hydrological indices. Variables ambientales utilizadas para modelar los 16 índices hidrológicos.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Units</th>
<th>Description</th>
<th>MLR Transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre</td>
<td>mm/year</td>
<td>Mean annual precipitation</td>
<td>none</td>
</tr>
<tr>
<td>Pre4</td>
<td>mm/month</td>
<td>Mean April precipitation</td>
<td>none</td>
</tr>
<tr>
<td>PreSu</td>
<td>mm/3month</td>
<td>Mean summer precipitation</td>
<td>$x^{1/3}$</td>
</tr>
<tr>
<td>PreMx</td>
<td>mm/month</td>
<td>Maximum monthly precipitation</td>
<td>none</td>
</tr>
<tr>
<td>PreMn</td>
<td>mm/month</td>
<td>Minimum monthly precipitation</td>
<td>$x^{1/2}$</td>
</tr>
<tr>
<td>MPrMn</td>
<td>month</td>
<td>Month of minimum precipitation</td>
<td>None</td>
</tr>
<tr>
<td>MPrRn</td>
<td></td>
<td>Monthly precipitation range</td>
<td>$x^{1/6}$</td>
</tr>
<tr>
<td>QPrRn</td>
<td></td>
<td>Quarterly precipitation range</td>
<td>$x^{1/3}$</td>
</tr>
<tr>
<td>Tem</td>
<td>°C</td>
<td>Mean annual temperature</td>
<td>none</td>
</tr>
<tr>
<td>TemSu</td>
<td>°C</td>
<td>Mean summer temperature</td>
<td>none</td>
</tr>
<tr>
<td>Eva</td>
<td>mm/year</td>
<td>Mean annual evapotranspiration</td>
<td>none</td>
</tr>
<tr>
<td>EvMx</td>
<td>mm/month</td>
<td>Maximum monthly evapotranspiration</td>
<td>log(x+1)</td>
</tr>
<tr>
<td>Are</td>
<td>Km$^2$</td>
<td>Total catchment area</td>
<td>$x^{1/5}$</td>
</tr>
<tr>
<td>Gra</td>
<td>%</td>
<td>Mean stream gradient</td>
<td>$x^{1/2}$</td>
</tr>
<tr>
<td>Ele</td>
<td>m</td>
<td>Reach elevation</td>
<td>$x^{1/2}$</td>
</tr>
<tr>
<td>Agr</td>
<td>%</td>
<td>Surface covered by agricultural land upstream of the river reach</td>
<td>arcsin(x)</td>
</tr>
<tr>
<td>For</td>
<td>%</td>
<td>Surface covered by forest upstream of the river reach</td>
<td>arcsin(x)</td>
</tr>
<tr>
<td>Perm</td>
<td>-</td>
<td>Soil permeability</td>
<td>none</td>
</tr>
<tr>
<td>Hard</td>
<td>-</td>
<td>Rock resistance to erosion</td>
<td>none</td>
</tr>
</tbody>
</table>
Environmental variables (predictors)

Several studies have highlighted the importance of climate, topography, land cover and geology on the hydrological regime regardless of geographic location (Kennard et al., 2010; Lane et al., 2017; Sanborn & Bledsoe, 2006). Thus, environmental variables were used to explain the hydrological character of the recorded flow series and predict this character in the entire river network. A synthetic river network (SRN) was delineated using a 25-m digital elevation model (DEM) with the NestStream program (Benda et al., 2007). The SRN comprised 667,406 segments with lengths ranging from 16 to 800 m and was used as a spatial network to integrate the hydrological and environmental information. Predictor variables were extracted from existing databases provided by several national and regional institutions. The predictor variables for each segment represented the mean value of the variables in the upstream catchment. A set of 19 variables was selected (Table 2); detailed information regarding the units, scale and sources of information can be found in Peñas et al. (2014).

In addition, according to the maximum number of degrees of freedom allowed by the different techniques, a maximum of 6 predictor variables was established for the models. The selection of these 6 variables was based on the combination of scatter plots (hydrological indices versus environmental variables) and parametric correlations to identify the environmental variables that were most meaningful for the prediction of each dependent variable (Knight et al., 2011). In this regard, the Pearson correlation values between the hydrological indices and the predictor variables were used as the main screening criterion. Hence, for each hydrological index, we selected the 6 predictor variables with the highest correlation values.

Modeling Techniques

The predictive performance of 4 distinct techniques to model hydrological indices was compared in this study. Modeling and statistical analysis were performed with R statistical software using the stats (v.3.3.2), gam (v.1.14) and randomForest (v.4.6) packages, except in the case of ANFIS models, which were developed using functions from the Mathwork’s MATLAB Fuzzy Logic Toolbox (FLT) included in a MATLAB code programmed by Marcé et al. (2004) and adapted by the authors to carry out the specific analyses performed in this study. The following section briefly describes each of the five modeling techniques.

Multiple Linear Regression (MLR)

MLR assumes a linear relationship between the predictor and the response variables through the estimation of parameters for each predictor. Specific transformations (Tables 1 and 2) were applied to meet the assumptions (normality, independence and homoscedasticity) for applying MLR. If data did not meet the assumptions through any transformation, that which was closest to meeting these requirements was used. The relative importance of each variable was established based on the comparison of the regression test statistic $T$ value.

Generalized Additive Models (GAMs)

GAMs are semi-parametric models (Hastie & Tibshirani, 1986) that relate the predictor and dependent variable through a link function and estimate a non-parametric function for each predictor in order to adapt it to the local behavior of the regression function in several regions (Venables & Dichmont, 2004). The identity link function of the Gaussian family was applied to the transformed variables using the same transformations as in MLR, given that they were assumed to be normally distributed. Thin plate regression splines were used with a maximum of 3 degrees of freedom. Parallel to MLR, the relative importance of each variable was established based on the comparison of the regression test statistic $T$ value.

Random Forest (RF)

RF (Breiman, 2001) comprises an ensemble of individual Classification and Regression Trees (CARTs). CARTs split the dimensional space defined by the predictors into groups that are as
homogeneous as possible based on series of binary rules. RF introduces random variation to CARTs by growing a defined number of trees with a bootstrap sample of the training data and a random sample of the predictors. The importance of the predictor variables is evaluated by randomly permuting each predictor variable in turn and predicting the response of the bootstrap sample observations. The decrease in prediction performance is the measure of importance of the original variable. Non-transformed response or predictor variables were used in the RF models.

**Adaptive Neuro-Fuzzy Inference System (ANFIS)**

ANFIS combines qualitative aspects of human knowledge from Fuzzy Inference Systems (FISs) with an effective, advanced machine learning method (neural networks) to adjust and tune these rules (Jang, 1993).

A FIS is based on fuzzy decision rules and the fuzzy reasoning unit (Jang, 1993). The fuzzy decision rules (if-then rules) are rules expressed in the form “if X (input variable) is A then Y (output variable) is B”, where A (premise) and B (consequence) are linguistic values (e.g., high and low). Fuzzy logic allows, within these decision rules, any judgment state to take values between 0 and 1 according to its probability. In this regard, Membership Functions (MFs) are the functions that relate a variable to the probabilities associated with the judgment states.

Fuzzy reasoning is an inference procedure used to derive conclusions from a set of fuzzy decision rules. The steps of fuzzy reasoning performed by a FIS are (Jang, 1993; Marce et al., 2004):

1. Compare the input variables with the MFs in the premise part of the fuzzy rules to obtain the probability of each linguistic label (fuzzification).
2. Combine (through logic operators) the probability in the premise part to get the weight of each rule.
3. Generate the qualified consequent of each rule depending on its weight.
4. Aggregate the qualified consequents to produce a crisp output (defuzzification).

Given an input-output problem, the construction of a FIS has two fundamental steps: the specification of an appropriate number and type of input and output MFs (structure identification) and the specification of the shape of the MFs (parameter estimation). The structure identification was solved by applying a trial-and-error procedure and a conservative criterion (i.e., minimum number of parameters in the best fit). Moreover, since the maximum number of parameters to be fitted increases exponentially with the number of variables and MFs and the total number of parameters should not exceed 1/6 the number of cases (Marce et al., 2004), a maximum of 3 MFs was established. Once the model structure, i.e., number of MFs, was defined, we estimated the parameters corresponding to each MF through the use of a numerical method called the Hybrid Learning Method (Marce et al., 2004). Specifically, these parameters were defined using adaptive neural networks algorithms. To avoid overfitting problems during the estimation of these parameters, the data set was randomly split into a training set (2/3 of the data set) used to fit the values and a trial set (1/3 of the data), which was not used by the hybrid learning algorithm. The splitting procedure was repeated 200 times and each time the parameters were adjusted individually. The hydrological indices were converted to the range (0 1), while the environmental variables were converted to z-scores (i.e., mean=0, standard deviation=1) according to ANFIS requirements. Finally, to obtain the importance of the predictors in each model, environmental variables were removed from the model one at a time while holding all other predictor variables. Then, for each model we calculated the predictive performance through the adjusted R². The larger the decrease of predictive performance, the greater the assumed importance of that variable.

**Validation and evaluation of model performance**

A jackknife cross-validation procedure was performed with R statistical software to test the predictive performance of each modeling technique for the 16 hydrological indices. This cross-validation procedure was applied by

leaving out one gauge at a time, developing a new model based on the remaining 155 observations and finally estimating the hydrological index for the left-out gauge. The results from this procedure produced estimates of each hydrological metric as if the gauging station were an ungauged site. The variation between observed and predicted values represents the uncertainty with which the model would be applied to predict index values at ungauged sites (Carlisle et al., 2010) and allows an assessment of the robustness of each method for estimating hydrological indices.

We employed the root-mean-square-deviance (RMSD) and the adjusted R² to assess the correspondence between observed and estimated values as a relative performance of each model, following other authors (Carlisle et al., 2010; Sanborn & Bledsoe, 2006; Van Sickle et al., 2006). Hence, models producing the lowest RMSD and the highest adjusted R² were deemed superior. In addition, we used Kruskal-Wallis to test whether the differences in adjusted R² found between the modeled indices and modeling techniques were significant.

Table 3. Predictive accuracy of the 16 hydrological indices using 4 different modeling techniques. The accuracy is compared according to the adjusted R² and the RMSD (Root Mean Square Distance). Increases in adjusted R² beyond 5% with respect to the MLR are represented by bold letters. All models presented significant results (p-value<0.01) in Fisher statistical testing. Capacidad predictiva de los 16 índices hidrológicos utilizando 4 técnicas de modelado diferentes. La comparación de la capacidad predictiva se ha realizado mediante el R² ajustado y la distancia cuadrática media. Aumentos en el valor del R² ajustado por encima del 5% respecto al del RLM se representa en negrita. Todos los modelos presentaron resultados significativos (p-valor<0.01) en el test de Fisher.

<table>
<thead>
<tr>
<th>Index</th>
<th>MLR</th>
<th>GAM</th>
<th>RF</th>
<th>ANFIS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adj r²</td>
<td>RMSD</td>
<td>Adj r²</td>
<td>RMSD</td>
</tr>
<tr>
<td>M1</td>
<td>0.77</td>
<td>2.99</td>
<td>0.82</td>
<td>2.67</td>
</tr>
<tr>
<td>M2</td>
<td>0.74</td>
<td>1.79</td>
<td>0.80</td>
<td>1.56</td>
</tr>
<tr>
<td>M4</td>
<td>0.74</td>
<td>4.88</td>
<td>0.80</td>
<td>4.32</td>
</tr>
<tr>
<td>M9</td>
<td>0.73</td>
<td>1.07</td>
<td>0.76</td>
<td>1.01</td>
</tr>
<tr>
<td>30LF</td>
<td>0.58</td>
<td>0.67</td>
<td>0.59</td>
<td>0.52</td>
</tr>
<tr>
<td>30HF</td>
<td>0.77</td>
<td>9.05</td>
<td>0.82</td>
<td>8.04</td>
</tr>
<tr>
<td>X5</td>
<td>0.75</td>
<td>10.55</td>
<td>0.80</td>
<td>9.27</td>
</tr>
<tr>
<td>JMax</td>
<td>0.18</td>
<td>21.94</td>
<td>0.18</td>
<td>22.14</td>
</tr>
<tr>
<td>JMin</td>
<td>0.19</td>
<td>17.63</td>
<td>0.19</td>
<td>17.65</td>
</tr>
<tr>
<td>Pred</td>
<td>0.16</td>
<td>0.13</td>
<td>0.33</td>
<td>0.12</td>
</tr>
<tr>
<td>FRE3</td>
<td>0.63</td>
<td>1.1</td>
<td>0.71</td>
<td>0.97</td>
</tr>
<tr>
<td>dPHigh</td>
<td>0.24</td>
<td>4.94</td>
<td>0.29</td>
<td>4.82</td>
</tr>
<tr>
<td>dPLow</td>
<td>0.30</td>
<td>25.29</td>
<td>0.32</td>
<td>24.72</td>
</tr>
<tr>
<td>nPos</td>
<td>0.46</td>
<td>12.47</td>
<td>0.51</td>
<td>11.94</td>
</tr>
<tr>
<td>nNeg</td>
<td>0.45</td>
<td>12.55</td>
<td>0.51</td>
<td>11.97</td>
</tr>
</tbody>
</table>

RESULTS

Model performance and predictor variables

The results exhibited a wide range of predictive performance, with adjusted $R^2$ ranging from 0.16 (MLR-Pred) to 0.88 (ANFIS-M4 and ANFIS-30HF; Table 3). However, all the models presented a level of significance $p$-value$<$0.01 when tested against the F statistic hypothesis. Model performance was higher when predicting the flow magnitude (MA, MH and ML) and frequency indices (FH: FRE7) than when predicting the timing (T: JMax, JMin and Pred), duration (DH: dHigh and DL: dLow) and rate-of-change (RC: nPos and nNeg) indices (K-W chi-squared$=$57.9, df$=$7, $p$-value$<$0.001; Table 3). In addition, for each hydrological index, the predictor variables kept the order of importance regardless of the modeling technique used. In this regard, it must be highlighted that, according to the Person correlation values, neither of the geology variables, Perm and Hard, were selected within the set of 6 initial predictor variables. The MA (except for $M9$) and MH indices were predicted with excellent accuracy, showing adjusted $R^2$s that commonly exceeded 0.8. In contrast, models of 30LF and $X95$ registered lower adjusted $R^2$s, which ranged from 15 to 25 % below those of the MA and MH indices. $Are$, annual (Pre), summer (PreSu) and April precipitation (Pre4) were the most important variables in practically all the flow magnitude models, especially those developed for MA and MH. On the other hand, when predicting the $M9$ and ML indices, other environmental variables such as gra, EvMx and QPrRn presented high contribution rates to the models. The timing index models presented the lowest predictive performances (Table 3). In general, adjusted $R^2$s for the JMax and JMin indices were not greater than 0.2, while the best model for Pred reached 0.4 (Table 3). Pre and MPPrRn were selected in all models for Jmax. MPPrRn, Eva and Ele were commonly included in the models for JMin. Pred was related mainly to PreMx and Gra. FRE3 was predicted with a maximum adjusted $R^2$ of 0.71 (Table 3), and the most influential variables were Ele, PrMx and QPrRn. Models for predicting dPHigh and dPlow rarely reached an adjusted $R^2$ over 0.3, and PrMx and PrMn were the most contributing variables, respectively (Table 3). Finally, models for nPos and nNeg showed adjusted $R^2$s close to 0.5 (Table 3). Pre, Ele and MPPrRn were the most influential variables in all of these models.

Comparison of modeling techniques

Differences in prediction accuracy among the different modeling techniques were not large (K-W chi-squared$=$1.44, df$=$3, $p$-value$=$0.7; Table 3). However, although differences were not significant, it must be remarked that the GAMs and ANFIS techniques outperformed MLR by more than 5 % of the adjusted $R^2$ in 10 and 13 hydrological indices, respectively. The greatest improvement in the predictive performance of these two techniques with respect to MLR was observed for the magnitude indices. ANFIS presented a mean increase of 7 % in the adjusted $R^2$ compared to MLR in all the magnitude indices, but this only resulted in marginal differences (K-W chi-squared$=$2.8487, df$=$1, $p$-value$=$0.09). If only the MA and MH indices were considered, the differences in performance between MLR and ANFIS reached up to more than 10 %, and significant differences were observed (K-W chi-squared$=$5.13, df$=$1, $p$-value$=$0.02). Differences between MLR and GAMs that resulted in improvements in adjusted $R^2$s beyond 5 % were found only for the MA and MH indices (WK-W chi-squared$=$6.72, df$=$1, $p$-value$=$0.01). In addition, ANFIS and GAMs outperformed MLR in one or several of the other index types (T, F and RC). On the other hand, RF did not show significant enhancements in relation to MLR (K-W chi-squared$=$0.017439, df$=$1, $p$-value$=$0.8949).

DISCUSSION

Model performance and predictor variables

This study confirms the findings of other works that not all the hydrological indices present the same potential to be predicted (Carlise et al., 2010; Yadav et al., 2007). Among the magnitude indices, MA and MH outperformed the ML indices. The high predictive performance of the
Regression carried out elsewhere (Knight et al., 2011; Sanborn & Bledsoe, 2006) have encountered difficulties in accurately predicting frequency indices, while we were able to predict FRES with a reasonable accuracy. The most important predictor variables for FRES were ele and pre, which agreed with results highlighted in previous studies (Carlisle et al., 2010; Knight et al., 2011; Ourada et al., 2001). This result is not surprising, as the combination of peak flows of nonsynchronous tributaries in the travel of flows downstream, i.e., river segments showing lower elev, has been observed to attenuate and dampen flow peaks, reducing the number of times a flow overcomes a threshold (Naiman et al., 1998; Poff et al., 1997).

However, it must be also pointed out that FRES takes into account moderate-high flow events that usually last several days. The duration of these events contrasts with the time scale of the commonly available climate database. For instance, in our study area only mean monthly precipitation series were available, which presumably lacked the proper time scale to characterize these events. Hence, the availability of daily precipitation data and its inclusion in models similar to those used in this study could be assumed to be highly beneficial for predicting these indices.

In addition, the lack of proper predictor variables has probably been the critical element hindering the development of more accurate models for the duration, rate-of-change and timing indices. For these three groups of hydrological indices, predictor variables derived from precipitation series (PreMx, PreMn and MPrn) were the most contributing variables. Given these results, it could be speculated that that the wettest areas presented longer high-flow and shorter low-flow events, along with a higher rate of flow rise and fall, than zones where precipitation is scarce. However, even if these relationships may seem obvious and expected, they cannot be assured with certainty due to the low accuracy of the obtained models.

**Comparison of modeling techniques**

Our analysis demonstrated that there was not an optimal technique to predict all hydrological indices. Several works focused on modeling a
variety of ecological and earth science variables have also highlighted that alternative complex techniques did not exhibit great differences in their prediction accuracy relative to traditional modeling approaches (Manel et al., 1999; Marmon et al., 2008). In contrast, other authors have found that complex modeling techniques outperformed linear approaches for predicting hydrological attributes (Booker & Snelder, 2012; Tisseuil et al., 2010), fluvial nutrient load (Marce et al., 2004) or species distribution (Elith et al., 2006).

Most of these authors emphasized the high flexibility of non-linear techniques in capturing complex relationships between predictor and response variables (Elith et al., 2006). However, when the underlying data structure and assumptions are met for a particular modeling method (e.g., linearity for MLR), the application of complex techniques does not necessarily produce significant improvements in model performance (Olden & Jackson, 2002a). This is the case for the hydrological indices of magnitude. However, GAMs, RF and ANFIS usually outperformed MLR in indices in which linearity was rarely achieved, e.g., pred, FRE3, dPhigh, nPos and nNeg (Table 3). It must be also stressed that GAMs and ANFIS outperformed MLR (>5%) in five and seven out of eight magnitude indices, respectively. GAMs allow for both linear and non-linear additive response shapes (Hastie & Tibshirani, 1986; Wood & Augustin, 2002). Hence, despite the linearity of several relationships, GAMs were able to tune the response more finely in specific sections where relationships were not linear.

The small gains in predictive performance of complex modeling, i.e., machine learning, techniques can be attributed to the low number of training sites (Kampichler et al., 2010). Since machine learning techniques are viewed as data-intensive methods and the spatial availability of hydrological data sets is typically small, their application is limited. In this sense, studies in which complex modeling methods outperformed linear approaches have presented a number of sites on a scale of thousands (e.g., Prasad et al., 2006), which contrasts with the 156 sites used in this work. Therefore, the application of these kinds of methods is promising where spatial coverage of hydrological data is substantial.

Beyond the predictive performance of the models, other characteristics such as the statistical skills needed to develop them and interpret the results must be taken into account when selecting the optimal modeling technique. For instance, ANFIS required the definition of the number and shape of MF, and it is recommended (Marcé et al., 2004) that these processes be carried out through an independent cross-validation process, as achieved in this work. On the other hand, the application of MLR involves complying with the assumptions of normality, homoscedasticity, independence and linearity, which was accomplished through different transformations (Tables 1 and 2). Given the disparity in the nature of hydrological indices and environmental data, no single transformation could be applied systematically and, as shown here, transformation does not always assure compliance with assumptions. In contrast, RF was the only fully automated technique, in which the distribution of the variables does not have to comply with any assumption (Breiman, 2001), which reduces the time needed and facilitates its application by users who are not specialists in statistics. Lastly, the ability of each technique to identify the actual relationships between the hydrological indices and the environmental variables must be taken into account. The four techniques agreed in the identification of the most important predictors for most of the models. However, MLR and GAMs allow straightforward relationships between predictors and response variables to be set (Manel et al., 1999). In contrast, machine learning methods have been largely seen as “black boxes.” For instance, the development of ANFIS models and the understanding of results require substantial time and knowledge, although enormous progress has been made in understanding the relationships underlying this technique (Marce et al., 2004; Olden & Jackson, 2002b). On the other hand, RF results form an ensemble of regression trees and may also become a black box when interpreting the results (Prasad et al., 2006). Nonetheless, the ‘randomForest’ package of R statistical software incorporates specific functions to numerically and graphically visualize the marginal effect of each predictor variable on the response (e.g., Alvarez-Cabria et al., 2016). These features definitively facilitate the application and
understanding of this technique over other machine learning approaches.

CONCLUSION

The application of four modeling techniques to predict 16 environmentally meaningful hydrological indices evidenced that all techniques might be suitable, since they showed similar prediction ability. Nonetheless, the accuracy of complex modeling techniques equal to that of more classical methods may be associated with the low number of unaltered gauges used to fit the models. Expanding this comparison to larger areas with a higher number of unaltered gauges will allow the actual potential of the most sophisticated methods to be analyzed. ANFIS represented a slight improvement over MLR, although the computational cost and level of knowledge required to apply the method and interpret the results may limit its application. It is widely accepted that machine learning techniques are capable of dealing with linear and non-linear relationships. Hence, we believe that machine learning techniques must be considered when they do not entail a significant increase in the required resources and the links between hydrological indices and predictors can be clearly understood.

On the other hand, not all hydrological indices were predicted with the same accuracy, resulting in critical implications and limitations depending on the further uses of these predictions. Magnitude and frequency indices were generally predicted with excellent accuracy, which opens a promising window to address several freshwater management and ecological issues. In contrast, none of the employed techniques allowed precise models for timing, duration and rate-of-change indices to be developed. Therefore, a major effort should be made to improve environmental databases in order to provide this climatic, geological, edaphological and groundwater information on the spatio-temporal scales on which flow regime patterns are influenced.

ACKNOWLEDGEMENTS

This study was partly funded by the Spanish Ministry of Economy and Competitiveness as part of the HYDRA (Ref. BIA2015-71197) and RIVERLANDS (Ref. BIA2012-33572) projects. Francisco J. Peñas is supported by FONDECYT - Concurso postdoctorado etapa 2017 (#3170313). We would also like to thank the Confederación Hidrográfica del Cantábrico, Confederación Hidrográfica del Ebro, Agencia Vasca del Agua, Agencia Catalana del Agua and Gobierno de Navarra for providing flow series data.

REFERENCES


127-188. Springer-Verlag. New York, INC, New York, USA.


REIDY LIERMANN, C.A., J.D. OLDEN, T.J. BEECHIE, M.J. KENNARD, P.B.M SKID-


