Limnetica, 37 (1): 145-158(2018). DOI: 10.23818/limn.37.12

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

Francisco J. Peñas[1,2,*], Jose Barquín[3] and César Álvarez[3]

[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.
[3] Environmental Hydraulics Institute. "IH Cantabria". Universidad de Cantabria. Parque Científico y Tecnológico de Cantabria. C/Isabel Torres, 15, 39011, Santander, Spain.

[*] 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 (MAG), 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 MAG obtuvieron, por norma general, los mejores resultados, sin embrago, consideramos que otros elementos, tales como los recursos computacionales requeridos o la experiencia *necesaria para aplicar la técnica e interpretar los resultados, deben tenerse en muy en cuenta a la hora de seleccionar el método más adecuado.*

Palabras clave: *Régimen natural de caudales, predicción, regresión múltiple, modelos aditivos generalizados, aprendizaje automático*

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-

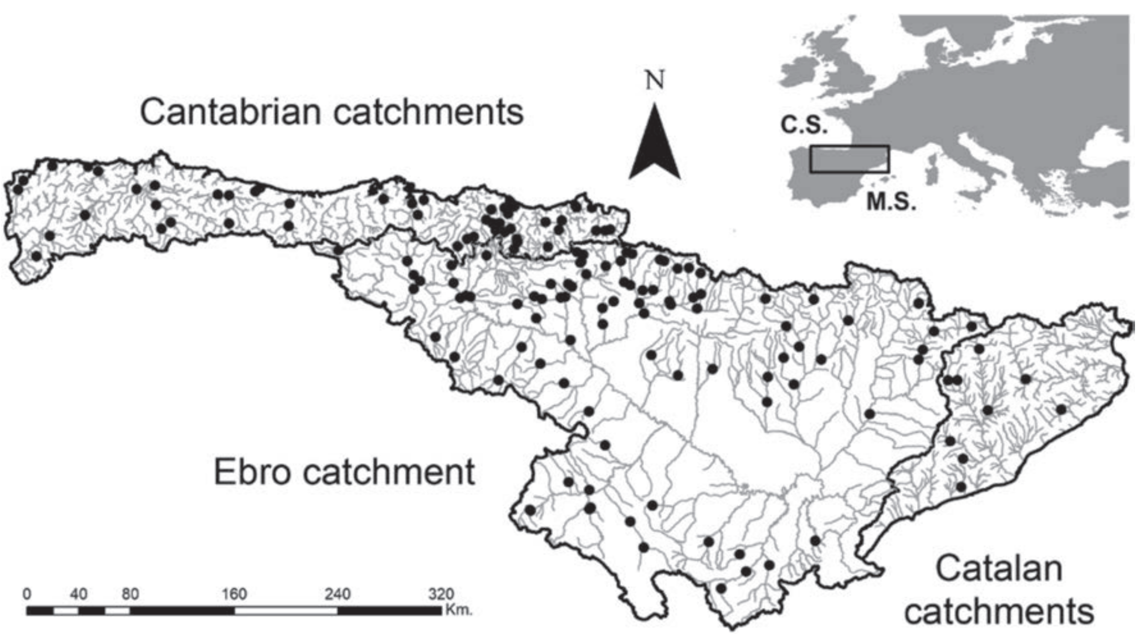

Figure 1. Map of unregulated gauges (●; n=156) in the study area. Black lines divide the Cantabrian, Ebro and Catalan catchments. (CS: Cantabrian Sea; MS: Mediterranean Sea). *Mapa de aforos no regulados en el área de estudio (●; n=156). Las líneas negras dividen las cuencas del Cantábrico, del Ebro y de Cataluña (CS: Mar Cántabro; 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 ungauged 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 km^2. 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 km^2, covering a total area of 22 000 km^2. 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 oro- 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, scrubs 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 km^2. 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 km^2, covering a total area of 16 500 km^2 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, scrubs 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.*

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

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.*

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

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 in 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 R^2. 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 R^2 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 R^2 were deemed superior. In addition, we used Kruscal-Wallis to test whether the differences in adjusted R^2 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 R^2 and the RMSD (Root Mean Square Distance). Increases in adjusted R^2 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 R^2 ajustado y la distancia cuadrática media. Aumentos en el valor del R^2 ajustado por encima del 5 % respcto al del RLM se representa en negrita. Todos los modelos presentaron resultados significativos (*p*-valor<0.01) en el test de Fisher.*

	MLR		GAM		RF		ANFIS	
Index	**Adj r^2**	**RMSD**	**Adj r^2**	**RMSD**	**Adj r^2**	**RMSD**	**Adj r^2**	**RMSD**
l1	0.77	2.99	**0.82**	**2.67**	0.78	2.95	**0.87**	**2.21**
l2	0.74	1.79	**0.80**	**1.56**	0.75	1.73	**0.79**	**1.59**
M4	0.74	4.88	**0.80**	**4.32**	0.72	5.02	**0.88**	**3.37**
M9	0.73	1.07	0.76	1.01	0.72	1.09	0.74	1.05
30LF	0.58	0.67	0.59	0.52	0.60	0.66	**0.63**	**0.62**
X95	0.54	0.55	0.54	0.54	0.52	0.58	**0.60**	**0.51**
30HF	0.77	9.05	**0.82**	**8.04**	0.75	9.59	**0.88**	**6.48**
X5	0.75	10.55	**0.80**	**9.27**	0.74	10.73	**0.84**	**70.89**
JMax	0.18	21.94	0.18	22.14	0.15	22.14	0.19	21.73
JMin	0.19	17.63	0.19	17.65	0.19	17.66	**0.25**	**17.01**
Pred	0.16	0.13	**0.33**	**0.12**	**0.32**	**0.12**	**0.41**	**0.11**
FRE3	0.63	1.1	**0.71**	**0.97**	0.65	1.16	**0.69**	**1.02**
dPHigh	0.24	4.94	**0.29**	**4.82**	**0.37**	**4.50**	**0.30**	**4.74**
dPLow	0.30	25.29	0.32	24.72	0.28	25.62	0.27	25.81
nPos	0.46	12.47	**0.51**	**11.94**	**0.53**	**11.72**	**0.54**	**11.50**
nNeg	0.45	12.55	**0.51**	**11.97**	**0.52**	**11.83**	**0.50**	**12.13**

RESULTS

Model performance and predictor variables

The results exhibited a wide range of predictive performance, with adjusted R^2s 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: *dPHigh* and DL: *dPLow*) 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^2s that commonly exceeded 0.8. In contrast, models of *30LF* and *X95* registered lower adjusted R^2s, 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^2s 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 *MPrRn* were selected in all models for *Jmax*. *MPrRn*, *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^2s close to 0.5 (Table 3). *Pre*, *Ele* and *MPrRn* 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^2s 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 (Carlisle *et al.*, 2010; Yadav *et al.*, 2007). Among the magnitude indices, MA and MH outperformed the ML indices. The high predictive performance of the

MA and MH indices is related to their dependence on precipitation events and direct catchment runoff (Tisseuil *et al.*, 2010). Precipitation variables were derived from 1000x1000-m precipitation grids and it was demonstrated that they were precise enough to produce reliable models. In this regard, there are several previous studies that have found strong relationships between hydrological variables and climatic predictor attributes (Carlise *et al.*, 2010: Reidy-Liermann *et al.*, 2011; Sanborn & Bledsoe, 2006). For instance, Reidy-Lierman *et al.* (2011) found that spring precipitation was the most important variable for discriminating rivers dominated by rain, snowmelt or mixed rivers. This agreed with our findings that *Pre4* was among the most important variables for predicting flow magnitude, as all these river types can be found in this study area (Bejarano *et al.*, 2010). Moreover, Solans & Poff (2015) and Bejarano *et al.* (2010) found that the segregation of river types in the Ebro Basin is largely explained by the variability of climatic predictors such as temperature, evapotranspiration and precipitation. In this regard, the high gradient of *EvMx* and *PreSu* values that prevails from the oceanic western part to the eastern Mediterranean sector plays a significant role in the discrimination of flow magnitudes across the study area.

On the other hand, the errors of ML index models have been relatively high. In most instances, significant correlations between ML indices and soil and geology characteristics have been observed in previous works (Clausen & Pearson, 1995; Kroll *et al.*, 2004, Lane *et al.*, 2017). The inclusion of these variables allowed prediction performances comparable to those showed by MA and MH indices to be obtained (Knight *et al.*, 2011; Sanborn & Bledsoe, 2006). It is likely that the small contribution of these variables in the present study was due to the low precision of the geology and soil data rather than the lack of causal links. The most detailed soil and geology maps in the study area have a 1:200 000 scale, which contrasts with the accuracy of the topography (25x25-m DEM), climatic (1000x1000-m grid) and land-use (1:25 000) data sources. Thus, we believe that improving soil and subsurface geology information should lead to improvements in modeling ML indices.

Regressions carried out elsewhere (Knight *et al.*, 2011; Sanborn & Bledsoe, 2006) have encountered difficulties in accurately predicting frequency indices, while we were able to predict *FRE3* with a reasonable accuracy. The most important predictor variables for *FRE3* 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. to 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 *FRE3* 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 *MPrRn*) 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; Marmion *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

ALCÁZAR, J., A. PALAU & C. VEGA-GARCIA. 2008. A neural net model for environmental flow estimation at the Ebro River Basin, Spain. *Journal of Hydrology*, 349, 44-55. DOI: 10.1016/j.jhydrol.2007.10.024

ALVAREZ-CABRIA, M, F.J. PEÑAS & J. BARQUÍN. 2016. Modelling the spatial and seasonal variability of water quality for entire river networks: Relationships with natural and anthropogenic factors. *Science of the Total Environment*, 545-546, 152-162. DOI: 10.1016/j.scitotenv.2015.12.109

ALVAREZ-CABRIA, M., A.M. GONZÁLEZ-FERRERAS, F.J. PEÑAS & J. BARQUÍN. 2017. Modelling macroinvertebrate and fish biotic indices: From reaches to entire river networks. *Science of the Total Environment*, 577, 308-318. DOI: 10.1016/j.scitotenv.2016.10.186

BEJARANO, M. D., M. MARCHAMALO, D. GARCÍA DE JALÓN AND M. GONZÁLEZ DEL TÁNAGO. 2010. Flow regime patterns and their controlling factors in the Ebro basin (Spain), *Journal of Hydrology*, 385, 323–335, DOI: 10.1016/j.jhydrol.2010.03.001, 2010

BOOKER, D. J. & T.H. SNELDER. 2012. Comparing methods for estimating flow duration curves at ungauged sites. *Journal of Hydrology*, 434, 78-94. DOI: 10.1016/j.jhydrol.2012.02.031

BREIMAN, L. 2001. Random Forest, *Machine Learning*, 45, 5-32. DOI: 10.1023/A:1010933404324

CARLISLE, D. M., J. FALCONE, D.M. WOLOCK, M.R. MEADOR & R.H. NORRIS. 2010. Predicting the Natural Flow Regime: Models for Assessing Hydrological Alteration in Streams. *River Research and Application*, 26, 118-136. DOI: 10.1002/rra.1247

CLAUSEN, B., & C.P. PEARSON, 1995. Regional Frequency-Analysis of Annual Maximum Streamflow Drought. *Journal of Hydrology*, 173, 111-130. DOI: 10.1016/0022-1694(95)02713-Y

HEUVELMANS, G., B. MUYS & J. FEYEN. 2006, Regionalisation of the parameters of a hydrological model: Comparison of linear regression models with artificial neural nets. *Journal of Hydrology*, 319, 245-265. DOI: 10.1016/j.jhydrol.2005.07.030

ELITH, J., C.H. GRAHAM, R. P. ANDERSON, M. DUDIK, S. FERRIER, A. GUISAN, R.J. HIJMANS, F. HUETTMANN, J.R. LEATHWICK, A. LEHMANN, J. LI, L.G. LOHMANN, B.A. LOISELLE, G. MANION, C. MORITZ, M. NAKAMURA, Y. NAKAZAWA, J.M. OVERTON, A.T. PETERSON, S.J. PHILLIPS, K. RICHARDSON, R. SCACHETTI-PEREIRA, R.E. SCHAPIRE, J. SOBERON, S. WILLIAMS, M.S. WISZ, & N.E. ZIMMERMANN. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29, 129-151. DOI: 10.1111/j.2006.0906-7590.04596.x

HASTIE, T. & R. TIBSHIRANI. 1986. Generalized additive models. *Statistical Science*, 1, 297–318.

JANG, J. S. R. 1993. Anfis-Adaptive-Network-Based Fuzzy Inference System. *IEEE Transactions on Systems Man and Cybernetics*, 23, 665-685.

KAMPICHLER, C., R. WIELAND, S. CALME, H. WEISSENBERGER & S. ARRIAGA-WEISS. 2010. Classification in conservation biology: A comparison of five machine-learning methods. *Ecological Informatics*, 5, 441-450. DOI: 10.1016/j.ecoinf.2010.06.003

KENNARD, M. J., B.J. PUSEY, J.D. OLDEN, S.J. MACKAY, J.L. STEIN & N. MARSH 2010. Classification of natural flow regimes in Australia to support environmental flow management. *Freshwater Biology*, 55, 171–193. DOI: 10.1111/j.1365-2427.2009.02307.x

KENNEN, J. G., L.J. KAUFFMAN, M.A. AYERS, D.M. WOLOCK & S.J. COLARULLO. 2008. Use of an integrated flow model to estimate ecologically relevant hydrologic characteristics at stream biomonitoring sites. *Ecological Modelling*, 211, 57-76. DOI: 10.1016/j.ecolmodel.2007.08.014

KNIGHT, R. R., W.S. GAIN & W.J. WOLFE. 2011. Modelling ecological flow regime: an example from the Tennessee and Cumberland River basins. *Ecohydrology*, 5, 613-627. DOI: 10.1002/eco.246

KROLL, C. N., J. LUZ, B. ALLEN & R.M. VOGEL. 2004. Developing a watershed characteristics database to improve low streamflow prediction. *Journal of Hydrologic Engineering*, 9(2): 116-125 DOI: 10.1061/(ASCE) 1084-0699(2004)9:2(116)

LANE, B. A, H.E. DAHLKE, G.B. PASTERNACK & S. SANDOVAL-SOLIS. 2017. Revealing the diversity of Natural Hydrologic Regimes in California with Relevance for Environmental Flows Applications. *Journal of the American Water Resources Association*, 53 (2), 411-430. DOI: 10.1111/1752-1688.12504

MANEL, S., J.M. DIAS, S.T. BUCKTON, & S.J. ORMEROD. 1999. Alternative methods for predicting species distribution: an illustration with Himalayan river birds. *Journal of Applied Ecology*, 36, 734-747. DOI: 10.1046/j.1365-2664.1999.00440.x

MARCE, R., M. COMERMA, J.C. GARCIA, & J. ARMENGOL. 2004. A neuro-fuzzy modeling tool to estimate fluvial nutrient loads in watersheds under time-varying human impact. *Limnology and Oceanography: Methods*, 2, 342-355. DOI: 10.4319/lom.2004.2.342

MARMION, M., J. HJORT, W. THUILLER, & M. LUOTO. 2008. A comparison of predictive methods in modelling the distribution of periglacial landforms in Finnish Lapland. *Earth Surface Processes and Landforms*, 33, 2241-2254. DOI: 10.1002/esp.1695

NAIMAN, R.J., T.J. BEECHIE, L.E. BENDA, D.R. BERG, P.A. BISSON, L.H. MACDONALD, M.D. O'CONNOR, P.L. OLSON, & E.A. STEEL. 1992. Fundamental elements of ecologically healthy watersheds in the pacific northwest coastal ecoregion. In: *Watershed Management: Balancing Sustainability and Environmental Change*. R.J. Naiman (ed):

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

OLDEN, J. D., & D.A. JACKSON. 2002a. A comparison of statistical approaches for modelling fish species distributions. *Freshwater Biology*, 47, 1976-1995. DOI: 10.1046/j.1365-2427.2002.00945.x

OLDEN, J. D., & D.A. JACKSON. 2002b. Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*, 154, 135-150. DOI: 10.1016/S0304-3800(02)00064-9

OLDEN, J. D., & N.L. POFF. 2003. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. *River Research and Applications*, 19, 101-121. DOI: 10.1002/rra.700

OUARDA, T., C. GIRARD, G.S. CAVADIAS, & B. BOBEE. 2001. Regional flood frequency estimation with canonical correlation analysis. *Journal of Hydrology*, 254, 157-173. DOI: 10.1016/S0022-1694(01)00488-7

PEÑAS, F. J., J. BARQUIN & C. ÁLVAREZ. 2014. The influence of methodological procedures on hydrological classification performance. *Hydrology and Earth System Sciences*, 18, 3393-3409. DOI: 10.5194/hess-18-3393-2014

POFF, N. L., J.D. ALLAN, M.B. BAIN, J.R. KARR, K.L. PRESTEGAARD, B.D. RICHTER, B. D., R.E. SPARKS & J.C. STROMBERG. 1997. The natural flow regime. A paradigm for river conservation and restoration. *BioScience*, 47, 769-784. DOI: 10.2307/1313099

POFF, N. L. & J.K.H. ZIMMERMAN. 2010. Ecological responses to altered flow regimes: a literature review to inform the science and management of environmental flows. *Freshwater Biology*, 55, 194-205. DOI: 10.1111/j.1365-2427.2009.02272.x

PRASAD, A. M., L.R. IVERSON & A. LIAW. 2006. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems*, 9, 181-199. DOI: 10.1007/s10021-005-0054-1

REIDY LIERMANN, C.A., J.D. OLDEN, T.J. BEECHIE, M.J. KENNARD, P.B.M SKIDMORE, C.P. KONRAD, & H. IMAKI. 2011. Hydrogeomorphic classification of Washington state rivers to support emerging environmental flow management strategies. *River Research and Applications*, 28 (9), 1340–1358. DOI: 10.1002/rra.1541

SANBORN, S. C., & B.P. BLEDSOE. 2006. Predicting streamflow regime metrics for ungauged streams in Colorado, Washington, and Oregon. *Journal of Hydrology*, 325, 241-261. DOI: 10.1016/j.jhydrol.2005.10.018

SNELDER, T. H., N. LAMOUROUX, J.R. LEATHWICK, H. PELLA, E. SAUQUET & SHANKAR, U. 2009. Predictive mapping of the natural flow regimes of France. *Journal of Hydrology*, 373, 57-67. DOI: 10.1016/j.jhydrol.2009.04.011

SOLANS, M.A. & A. MELLADO-DÍAZ 2015. A landscape-based regionalization of natural flow regimes in the Ebro river basin and its biological validation. *River Research and Applications*, 31: 457–469. DOI: 10.1002/rra.2860

TISSEUIL, C., M. VRAC, S. LEK & A.J. WADE. 2010. Statistical downscaling of river flows. *Journal of Hydrology*, 385, 279-291. DOI: 10.1016/j.jhydrol.2010.02.030

VAN SICKLE, J., D.D. HUFF & C.P. HAWKINS. 2006. Selecting discriminant function models for predicting the expected richness of aquatic macroinvertebrates. *Freshwater Biology*, 51, 359-372. DOI: 10.1111/j.1365-2427.2005.01487.x

VENABLES, W. N. & C.M. DICHMONT. 2004. GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research. *Fisheries Research*, 70, 319-337. DOI: 10.1016/j.fishres.2004.08.011

WOOD, S. N. & N.H. AUGUSTIN. 2002. GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. *Ecological Modelling*, 157, 157-177. DOI: 10.1016/S0304-3800(02)00193-X

YADAV, M., T. WAGENER & H. GUPTA. 2007. Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. *Advances in Water Resources*, 30, 1756-1774. DOI: 10.1016/j.advwatres.2007.01.005