JOURNAL OF ENVIRONMENTAL HYDROLOGY

The Open Access Electronic Journal of the International Association for Environmental Hydrology
On the World Wide Web at http://www.hydroweb.com

VOLUME 23 2015

DEVELOPMENT AND APPLICATION OF OPTIMALLY ADAPTIVE LINEAR COMBINATIONS (OALC) IN HYDROLOGICAL FORECASTING (DAILY FORECAST RESOLUTION AND DIFFERENT LEAD TIMES)

Efraín Domínguez[1]
Hector Angarita[1]
Fabián Caicedo[2]
Hebert Rivera[3]

[1]Faculty of Environmental and Rural Studies, Pontificia Universidad Javeriana, Bogota, Colombia
[2]Faculty of Engineering, Pontificia Universidad Javeriana, Bogota, Colombia
[3]Faculty of Engineering, Universidad Militar Nueva Granada, Bogota, Colombia

This paper presents an application of optimally adaptive linear combinations (OALC) to forecast daily discharges and water levels for different forecast horizons (one day to two weeks ahead). To calibrate the OALC, we have used the forecast mean squared error. The optimal length of the calibration period and the optimal number of predictors were established using the performance criteria of the Hydrometeorological Centre of Russia as the goal function (Domínguez et al., 2011). The percentage of successful forecast for error levels lower or equal than 5, 10, and 20% were also determined to find the most efficient OALC. The forecasting technique presented here was implemented in Colombia to predict daily streamflows to the Betania hydropower reservoir located in the upper Magdalena River, as well as to forecast the water levels at the hydrometric station "El Banco" located within the navigable sector of the same river. Finally, we discuss the possibility of using OALC as a deterministic kernel to simulate the evolution of conditioned probability density curves (CPDC) through the multidimensional Fokker-Planck-Kolmogorov (FPK) equation to provide users with probabilistic forecasts of river water levels. Developed forecast method, and its potential probabilistic enhancement, can be useful for forecasting river water levels under climate change conditions, where uncertainty maximizes and extremes flows are expected to occur more frequently. The OALC, being a simple linear method, can be solve by means of Kolmogorov's optimal interpolation or by least square methods as it is done in this paper. OALC proved to be efficient for lead times "T" from 1 to 10 but it is expected to increase computational time requirements when issuing forecasts for lead times from 11 to 24 or more.

INTRODUCTION

The hydrological variability of surface runoff is not only a source of hazards for human settlements located near the river shoreline but also a source of risk for the productive sectors. In Colombia, two sectors are particularly sensitive to hydrological variability: hydropower and fluvial navigation. Recently, extreme hydrological values have become more severe, possibly as a consequence of global climate change (IPCC, 2007). Higher variability in the hydrologic regime of surface streams leads to increased failures among water-dependent economic sectors. The hydropower sector in Colombia covers 75% of the energy demand; therefore, diminishing the hydrological risk for this sector is an ongoing critical task (UPME and Ministerio de Minas y Energía, 2006). At the same time, the sum of the freight transported through the Magdalena River accounts for 4% of the country's total payload, and many efforts are being made to increase this level to 12% by year 2020 (Mintransporte, 2007). To meet this aim, a satellite-based navigation system that allows navigation, even at night, was implemented (Alvarado, 2006; Mintransporte, 2007). For both the hydropower and navigation sectors, short-, medium- and long-term forecasts are an essential planning tool, especially under the conditions of climate change. Due to the disruption of stable global climate conditions, the mentioned economic sectors require hydrologic forecast methods that can predict the time evolution of characteristics for streamflows or stages in unstable conditions. The method presented here is intended to be a component of the tools that could be used to address such a requirement.

Forecasting process is a complex task that needs real time measuring, transmission and assimilation systems, forecast method and forecast broadcasting procedure (NOAA, 1999; WMO, 1994). Requirements for the forecast method are: to be simple, low computing and time demanding, to be of best precision and to enhance process predictability (WMO, 1994). Prediction of high lead times usually demands long computational time but at the same time high lead time forecasts are the most useful information for decision makers, specially for decision makers at the hydropower sector in Colombia (Zevin, 1994). A plethora of mathematical models and methods of different type have being developed but where they fulfill before presented requirements is still a question. Of special importance is the required run time for forecast emission. Previous work showed that methods of artificial intelligence (i.e. neural networks, diffusive logic) and physically based models demand high computational time making it unfeasible their use as forecast operator in real time regime (Domínguez et al., 2010, 2009). At the same time, conceptual models (Biondi et al., 2012; Fenicia et al., 2011; Schumann, 1993; Solomatine and Wagener, 2011) are sufficiently fast for real time operation but, their recalibration and operator restructuring is not feasible on the flight. Finally, not of less importance for conceptual models, their rigorous probabilistic generalization is not well studied yet. Mentioned features are crucial for real time systems due to the possibility o transmission failures in real time. A forecast operator able to reconfigure its structure and recalibrating its parameters itself, in real time, is more than mandatory. The quantitative handling of forecast uncertainty is also a desirable to support decision making process in hydropower reservoir operation (World Commission on Dams, 2000). Classical modeling approaches do not exhibit described properties, leading to ridge forecast operators requiring the measuring, transmission and assimilation systems to work without failures. Present work aims to find a forecast technique able to reconfigure and recalibrate its forecast operator on the flight but also to be theoretically feasible to enhance to probabilistic forecast operator. As remarked OALC techniques postulates as a candidate for such technique and it looks promissory that derived OALC can serve as deterministic kernels for the implementation of the Fokker-Planck-Kolmogorov equation as the theoretic approach for forecast uncertainty handling.

The forecasting technique we present here was developed to become the deterministic kernel for the Fokker-Planck-Kolmogorov equation when dealing with high-order Markov processes (processes with high memory). In the context of changing climate and high human pressure over the river basins, it is expected for the hydrological regime to become unstable, gaining in uncertainty and requiring probabilistic forecast to be a common requirement for early warning systems (Domínguez and Lozano-Báez, 2014). The first application of OALC was presented by Domínguez (Domínguez, 2005), who showed the feasibility of the method without explaining its foundations. Herein, we present the foundations for the establishment of the method and some results and recommendations for its proper application and integration with the Fokker-Planck-Kolmogorov Equation approach to forecast hydrological processes but coping with uncertainty of hydrological inputs and parameters.

A FRAMEWORK FOR MODELING WITH DYNAMIC CALIBRATION

In typical mathematical modeling, the user tends to calibrate the model using all available historical information. We propose a dynamic approach, showing that a on the flight real time calibration of linear combinations could be better than a static calibration that uses all registered information regarding the forecasting state variable. Dynamic approach means the structure of forecast operator and the optimal values of this operator are established at the issuing forecast moment, "on the flight". We suggest, dynamic calibration can be useful when dealing with time series that have cycles of different frequency; it can be used to forecast for a lead time *«T»* with $T \leq \rho$ and $\rho << Ns$, where ρ is the correlation radius and Ns is the total time series length.

Given a time interval $[t-(N-1),\ t] \in \Re$, with a record of N observed values of the variable $Y(t)$, the forecast $Y(t+T)$ can be expressed in the form:

$$Y(t+T)=L[W(t)] \tag{1}$$

In (1), "*L*" is a mathematical operator that functions over the *k*-order polynomial $W(t)$, combining endogenous and exogenous predictors within the following structure:

$$L[W(t)]=\sum_{i=t}^{t-\rho_Y} a_i Y_i^{\alpha_i}+\sum_{k=1}^{m}\sum_{j=t}^{t-\rho_{X_{kj}Y}} b_{kj} X_{kj}^{\beta_{kj}} \tag{2}$$

Here, ρ_Y and $\rho_{X_{kj}Y}$ are the correlation radius for the endogenous variable $Y(t)$ and the cross-correlation radius of $Y(t)$ with the exogenous variables $X(t)_k$; a_i, b_{bj}, α_t; and β_{kj}; are the exponents and coefficients for $W(t)$; and $k = \max(\alpha_i, \beta_j)$ represents the order of the polynomial $W(t)$.

Let the difference be set as $Y(t+T)^{Observed} - Y(t+T)^{Forecasted} = \delta$; then, we can denominate L as an optimal operator if it minimizes some function of δ. For instance, the minimization of the mean squared error may be defined as follows:

$$\min \rightarrow E[\delta^2]=E\{[Y(t+T)-L[W(t)]]^2\} \tag{3}$$

If there is a known finite data vector of observed values for $Y(t)$, and if the autocorrelation and cross-correlation radius for this vector and the exogenous variable are ρ_Y and $\rho_{X_{kj}Y}$, then the forecast $Y(t+T)$ can be expressed as a combination of previous values of $Y(t)$ and $X(t)$ between the time interval $[t, t\text{-}\rho_Y]$ for $Y(t)$ and the time interval $[t, t\text{-}\rho_{XY}]$ for the cross-correlation component, given the following combination:

$$Y(t+T)=\sum_{i=t}^{t-\rho_Y} a_i Y_i^{\alpha_i}+\sum_{k=1}^{m}\sum_{j=t}^{t-\rho_{X_{kj}Y}} b_{kj} X_{kj}^{\beta_{kj}} \tag{4}$$

Hence, a_i, b_i, α_i and β_i are the exponents and coefficients that are used to minimize the following expression:

$$\sigma^2(\mathbf{a}, \mathbf{b}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = E\left\{\left[Y(t+T) - \left(\sum_{i=t}^{t-\rho_Y} b_i Y_i^{\alpha_i} + \sum_{k=1}^{m} \sum_{j=t}^{t-\rho_{X_{kj}Y}} b_{kj} X_{kj}^{\beta_{kj}}\right)\right]^2\right\} \quad (5)$$

in a time window of length θ. From this fact, we understand that in this time domain:

$$\mathbf{a} = \{a_t, a_{t-1}, \dots, a_{t-\rho}\}$$

$$\mathbf{b} = \{b_{k,t}, b_{k,(t-1)}, \dots, b_{k,(t-\rho_{X_{kj}Y})}\}$$

$$\boldsymbol{\alpha} = \{\alpha_t, \alpha_{t-1}, \dots, \alpha_{t-\rho_Y}\}$$

$$\boldsymbol{\beta} = \{\beta_{k,t}, \beta_{k,(t-1)}, \dots, \beta_{k,(t-\rho_{X_{kj}Y})}\}$$

are the solution vectors for the following systems of equations:

$$\begin{gathered}\frac{\partial \sigma^2}{\partial a_i} = 0, \quad i = t, \quad t\text{-}1, \dots, t\text{-}\rho_Y; \quad \frac{\partial \sigma^2}{\partial b_j} = 0, \quad i = t, \quad t\text{-}1, \dots, t\text{-}\rho_{X_{kj}Y}; \\ \frac{\partial \sigma^2}{\partial \alpha_i} = 0, \quad k = t, \quad t\text{-}1, \dots, t\text{-}\rho_Y; \quad \frac{\partial \sigma^2}{\partial \beta_j} = 0, \quad l = t, \quad t\text{-}1, \dots, t\text{-}\rho_{X_{kj}Y}.\end{gathered} \quad (6)$$

For the time interval θ, we could also require, even for nonlinear combinations, that the quotient between the standard deviation of the forecast squared error $(S = \sigma(\mathbf{a}, \mathbf{b}, \boldsymbol{\alpha}, \boldsymbol{\beta}))$ and the standard deviation for the increments of $Y(t)$ and (σ_Δ) be less than 0.8, thus fulfilling the model performance criterion of Russian Hydrometeorological Center (S/σ_Δ) as it is applied in (Domínguez et al., 2011).

From the above, it follows that the implementation of OALC requires four parameters to be established: i) the autocorrelation and cross-correlation radiuses $\boldsymbol{\rho}_Y$ and $\boldsymbol{\rho}_{XY}$ (obtained from the autocorrelation and cross-correlation functions); ii) the forecast horizon (lead time), T; and iii) the length of the parameterization time window, θ. The last parameter can be found by means of optimization, and in general, $\theta << N$ (here N is the total time series length for $Y(t)$.

Selection of Predictors

As shown above, information about the predictors is obtained using correlation analysis. As predictors, we can use lagged values of $Y(t)$ and actual (where available) and lagged values for $X_i(t)$. For instance, the lag for each predictor is a function of time in the form $t\text{-}m\Delta t$, where m=0...ρ.

Let us consider that

$$\alpha_t = \alpha_{t-1} = \dots = \alpha_{t-\rho_Y} = \beta_t = \beta_{t-1} = \dots = \beta_{t-\rho_{X_kY_{kj}}} = 1 \quad (7)$$

This reduces the search domain of the optimal operator to the field of first-order polynomials. To avoid the selection of redundant predictors, we must choose only those predictors with coefficients $\{a_t, a_{t-1}, ..., a_{t-\rho}\}$ or $\{b_t, b_{t-1}, ..., b_{t-\rho_Y}\}$ that fulfill the conditions $\left(a_k/\sigma_{a_k} \geq 2\right)$ or $\left(b_k/\sigma_{b_k} \geq 2\right)$, where σ_{a_k} and σ_{b_k} are the squared definition error for each coefficient.

As is common in hydrological forecasting, we can use actual and previous rainfall values and upstream river inflow data. Usually, an exhaustive search algorithm, using the criteria S/σ_{Δ} as the goal function, can be used to select the optimal predictors in a reasonable time. To do so, a three-level embedded search cycle must be programmed. The first cycle searches for the optimal predictor number within the interval from 1 to $[2t-(\rho_Y+\rho_{XY})]$ predictors. The second cycle has to find the optimal parameterization window length θ, from $[2t-(\rho_Y+\rho_{XY})]$ to N days. The last cycle selects the optimal forecast lead time T from 1 to $k[\max(\rho_Y,\rho_{XY})]$ days, where $k \in \Re$ and takes values from 0 to 1. Using k >1 seems illogical, but nevertheless, experience has shown that using k >1 sometimes gives good results. This could be explained by the fact that the peer correlation coefficients explain the peer relationship level, so that radiuses ρ_Y and ρ_{XY} are not sufficient to explain more complex relationships.

USING OALC AS A DETERMINISTIC KERNEL FOR STOCHASTIC SIMULATION THROUGH THE MULTIDIMENSIONAL FOKKER-PLANCK-KOLMOGOROV EQUATION

Although the operator $L[W(t)]$ and its adaptive parameterization procedure show good forecasting performance when predicting $Y(t+T)$ values, the stochastic decision-making procedures for the hydropower and navigation sectors require a probabilistic forecast of stages and streamflows (Domínguez, 2005). The probabilistic forecast of streamflows and stages has been implemented for a first-order Markov process to simulate the evolution of monthly and annual conditioned probability density curves (CPDC) (Domínguez and Kovalenko, 2009; Domínguez, 2004; Khaustov and Kovalenko, 1998). For daily streamflows and daily stages, the autocorrelation radius is too high to consider these processes as first-order Markov processes. High autocorrelation suggests that the processes must be described with high-order ordinary differential equations (ODE). A first-order ODE is equivalent to a first-order autoregressive model (Kazakievich, 1989), so the ODE:

$$\alpha \frac{dx}{dt} + x(t) = Z(t) \tag{8}$$

expressed in finite differences corresponds to an autoregressive process of type:

$$\overset{\circ}{X}(t) = \alpha_1 \overset{\circ}{X}(t-1) + Z(t) \tag{9}$$

where Z(t) is a white noise process.

At the same time, a process described through a high-order ODE can be explained as a high-order autoregressive process. It is also known that any ODE of order n of the type:

$$u^n = f(x, u, u', \ldots, u^{n-1}) \tag{10}$$

can be expressed as a system of first order ODEs. In fact, by introducing new known functions $y_1 = u$, $y_2 = u'$,..., $y_n = u^{n-1}$, equation (10) becomes an ODE system with n first order ODEs as follows:

$$\begin{aligned} & y'_1 = y_2; \\ & y'_2 = y_3; \\ & \vdots \\ & \vdots \\ & y'_{n-1} = y_n; \\ & y'_n = f(x, y_1, y_2, \ldots, y_n). \end{aligned} \tag{11}$$

The notion of a Markov process can be generalized for a set of random processes (Sveshnikov, 1968). Given a set of n random processes $U_1(t)$, $U_2(t)$,…,$U_n(t)$ describing the vector $\boldsymbol{R}(t)$ and referencing the ordinates of $\boldsymbol{R}(t)$ at times τ and t ($\tau > t$) as $Y_1, Y_2,...,Y_n$ and $X_1, X_2,...,X_n$, respectively, a multidimensional random process of Markov type results if the conditioned distribution law for the moment τ only depends on the distribution law at time t (t >1). In this case, by analogy with the one-dimensional process, the system can be described with a multidimensional CPDC of type $p(t, x_1, x_2,..., x_n; \tau, y_1, y_2,..., y_n)$. For such a CPDC, it is possible to deduce the corresponding backward and forward Kolmogorov equations:

$$\frac{\partial p}{\partial t} + \sum_{l=1}^{m} A_l \frac{\partial p}{\partial x_l} + \frac{1}{2} \sum_{\substack{l=1 \\ j}}^{m} B_{lj} \frac{\partial^2 p}{\partial x_l \partial x_j} = 0 \tag{12}$$

$$\frac{\partial p}{\partial t} + \sum_{l=1}^{m} \frac{\partial (A_l p)}{\partial y_l} - \frac{1}{2} \sum_{\substack{l=1 \\ j}}^{m} \frac{\partial^2 (B_{lj} p)}{\partial y_l \partial y_j} = 0 \tag{13}$$

Equation (13) is also known as the multidimensional Fokker-Planck-Kolmogorov equation. Hence, in (12) and (13), A_l and B_{lj} are the drift and diffusion coefficients, respectively, and they have the same physical meaning as in the one-dimensional case (Domínguez, 2004; 2007; Gardiner, 1985). According to Sveshnikov (Sveshnikov, 1968), a stochastic process $U(t)$ with a spectral density $S_u(\omega)$ of the type:

$$S_u(\omega) = \frac{|P_m(i\omega)|^2}{|Q_n(i\omega)|^2}, \quad \text{m} < n \tag{14}$$

where the polynomials P_m and Q_m are defined with the coefficients $\alpha_1, \alpha_2,..., \alpha_n$ and $\beta_1, \beta_2,..., \beta_n$, is the stationary solution for the differential equation:

$$\frac{d^n U(t)}{dt^n} + \alpha_1 \frac{d^{n-1} U(t)}{dt^{n-1}} + ... + \alpha_n U(t) = \beta_0 \frac{d^m \xi(t)}{dt^m} + ... + \beta_m \xi(t) \tag{15}$$

where $\xi(t)$ is a random function with independent ordinates (white noise) and with the autocorrelation function:

$$K_\xi(\tau) = \delta(\tau) \tag{16}$$

Equation (15) can be reduced to a system of first-order ODEs as presented above (see equations (10) and (11)), and the respective A_l and establish a deterministic kernel based on OALC, the best-performance OALC has to be transformed into an ODE of order n and subsequently into a system of first-order ODEs with n equations. An example of this may be found in Xie et al (Xie et al., 2005), where a second-order ODE that describes a Rayleigh oscillator is reduced to a system of first-order ODEs with two equations, which can be described in terms of the Fokker-Planck-Kolmogorov Equation. It is clear that OALC will produce an ODE system of order greater than 2, but the algorithm used to build the multidimensional Fokker-Planck-Kolmogorov Equation remains the same as presented in Xie et al. (Xie et al., 2005), Sveshnikov (Sveshnikov, 1968) and Domínguez (Domínguez et al., 2007).

OALC APPLICATION

Data and area of application

To evaluate the OALC performance, operator (4) with assumption (7) was applied in the upper basin and in the navigable part of the Magdalena River.

In the upper part of the river, the forecast of streamflows to the Betania hydropower reservoir was examined. For this goal, the water levels registered by the hydrological stations Paicol (2105706) and Puente Balseadero (2104701); and the rainfall data from meteorological stations Esc. Riosucio N2 (2105031), Esc. Agr. La Plata (2105502), La Escalereta (2104007), Macagual (403503), Bajo Frutal (2101013), La Candela (2101014), San Adolfo (2103006), Resina (2103502), Loma Redonda (2601005), Laguna San Rafael (2601007) and Santa Rosa (4401010) were used. In the navigable sector of the Magdalena River, we used river stage data collected at the station El Banco (2502501). All mentioned data were provided by the Instituto de Hidrología, Meteorología y Estudios Ambientales – IDEAM, which manages the hydrometeorological network and maintains the hydrometeorological information system using the cited station identification codes. For the Betania reservoir, the streamflow registered by the Paicol and Puente Balseadero stations accounts for 93% of the total streamflow to the reservoir (Figure 1). A correlation analysis between hydrological data and meteorological stations indicated the sum of the rainfall recorded by the abovementioned stations as a good streamflow predictor. An additional analysis of available satellite data led to the conclusion that the potential precipitation from MODIS sensors and from the system PERSIANN can also be used as a predictor (Capacci and Conway, 2005; Kou-Lin et al., n.d.; Soroosh et al., 2000).

Other sources of data used to delineate the basins and to define the hydrological stations' catchment areas included a digital elevation model with 90 m spatial resolution from the SRTM Project(CGIAR-CSI, 2012), IDEAM's station cadaster and technical information about the Betania reservoir provided by the reservoir operator. All numerical experiments were coded within the Python and Delphi programming environments, and calibrating and validation were performed using 700 forecast tests for each forecasting point and using exhaustive search loops to find the optimal parameters (θ and *n)*. The optimization time was not more than 10 minutes on a PC with a 3.4 Ghz Pentium processor and 1 GB of RAM.

A total of 90% of the streamflow to the Betania reservoir is accounted for by the sum of the inflows measured at the Paicol and Puente Balseadero Stations. To forecast this total streamflow, we built OALCs with endogenous predictors only and with both endogenous and exogenous predictors. A special OALC was also built to forecast water levels at Paicol station. The exogenous predictor used was the total rainfall registered by the meteorological stations 2101013, 2101014, 2103006, 2103502, 2601005, 2601007 and 4401010. To define the calibration and validation periods, the time interval from January 1984 to May 1994 was used. For stage forecasting at the Paicol station, we used precipitation information from stations 2105031 and 2105502 and potential rainfall data from the satellite-based system PERSIANN, both from the period from 2000 to 2005. According to hydrological records, the annual average inflow to the Betania reservoir is 430 m^3/s, the maximal registered inflow is 630 m^3/s and the catchment area covers 13600 km^2. The catchment area for the Paicol station is 4840 km^2. To show the ability of OALC to work with satellite rainfall data as an exogenous predictor, the Paicol station was selected, due to the common observation period shared by Paicol's ground-based records and the information of the PERSIANN system.

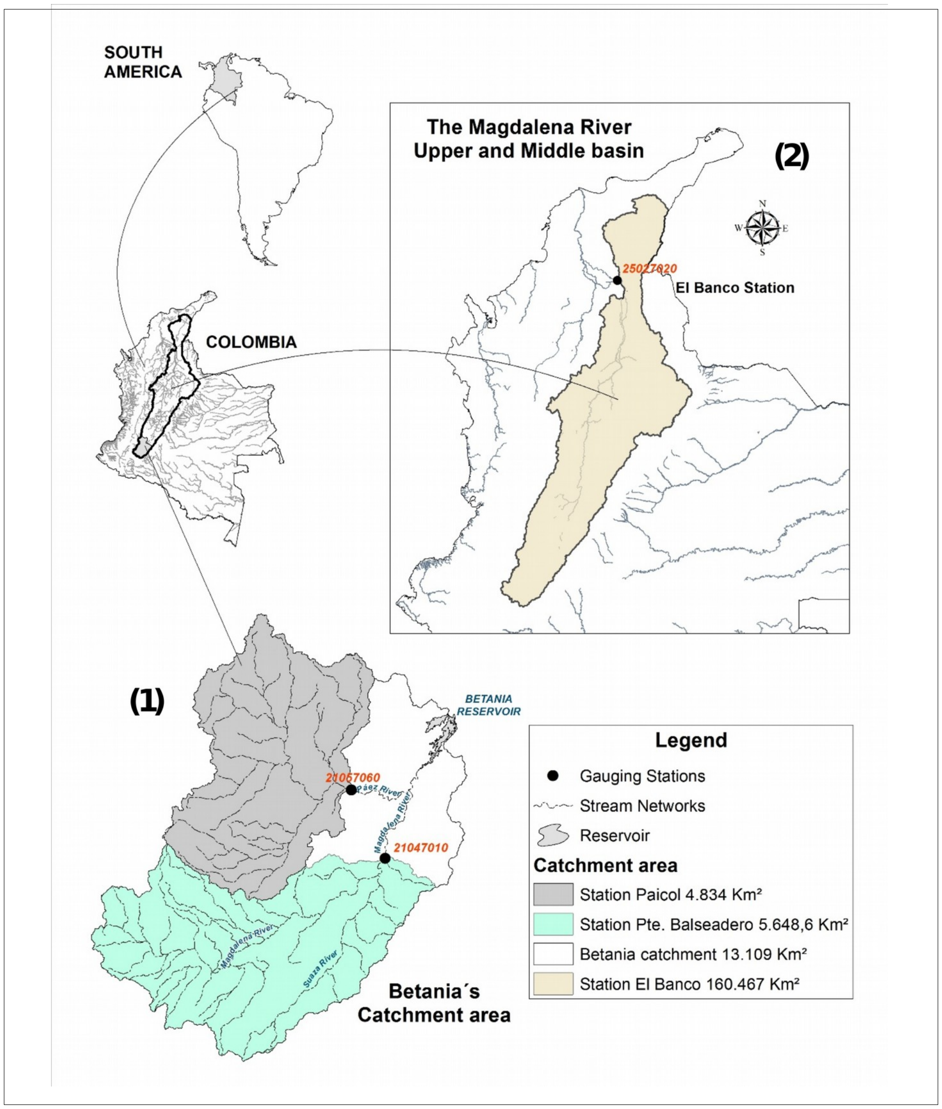

Figure 1. The study area: 1) Betania's hydropower reservoir and 2) El Banco hydrometric station (Middle part of the Magdalena river basin)

RESULTS AND DISCUSSION

OALC streamflow forecast for Betania reservoir

Autocorrelation functions for Betania's daily streamflows show that the runoff process has a two-day memory. Cross-correlation between daily streamflows and predictor rainfall (the sum of daily precipitation for meteorological stations 2101013, 2101014, 2103006, 2103502, 2601005, 2601007 and 4401010) shows that the concentration time is less or equal to one day (Figure 2). The best OALC can be found by optimizing θ and the number of predictors n used for equation (4). The number of predictors to be included can be defined according to the correlation radius for the given rainfall – runoff process. Nevertheless, in this case, an exhaustive search algorithm was implemented to check OALC performance when using between 1 and 30 predictors. The parameter θ was defined from among integers from 25 to 300. We built both, an OALC that uses only endogenous predictors and one that combines endogenous and exogenous predictors. The optimal OALC that we could find using endogenous predictors is as follows:

$$Y(t+T)=\sum_{i=t}^{t-\rho_Y} a_i Y_t \tag{17}$$

Equation (17) can be interpreted as an autoregressive model of order n. For this model, the lead time was T=1 , so the next day's streamflow was forecasted for Betania's reservoir.

For admissible values of S/σ_Δ, an optimal OALC is defined by a combination that yields a general performance of $(S/\sigma_\Delta) \leq 0.80$ (Appolov *et al.*, 1974). The OALC with a different number of endogenous predictors and without exogenous predictors showed the performance presented in Figure 3. The percentage of correctly predicted forecasts is presented in Figure 3 as well.

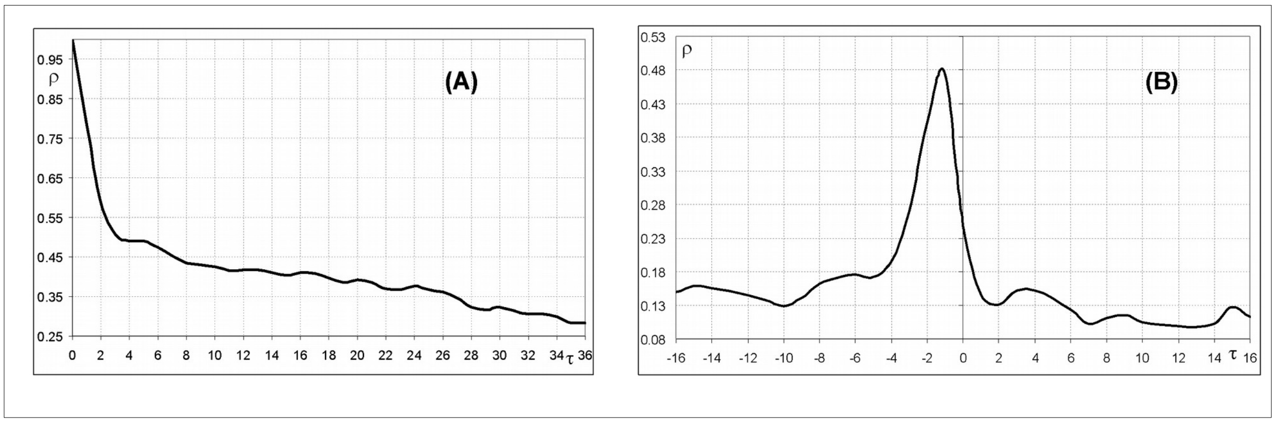

Figure 2**.** Streamflow autocorrelation (A) and rainfall/runoff cross-correlation (B) functions

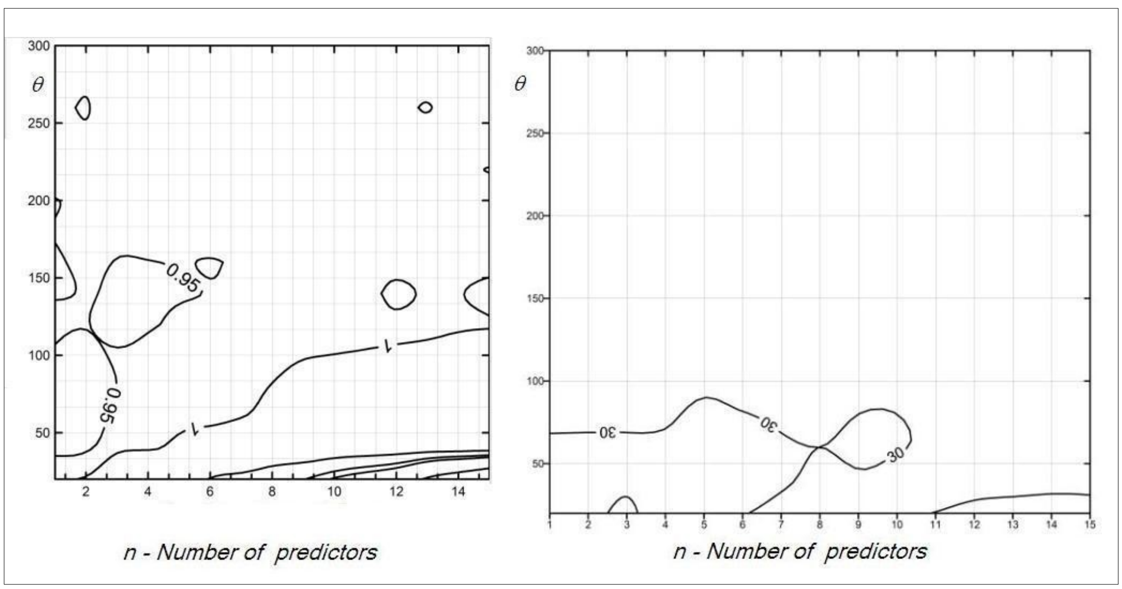

Figure 3. S/σ_Δ behavior (left) and percentage of forecasts with error level ≤ 5% (right) using equation (17) for Betania reservoir daily streamflows.

The best OALC for (17) was found using θ = 60 and n=1. This OALC has the structure:

$$Y(t+T)=aY(t) \tag{18}$$

With (18), we reached a S/σ_Δ value of 0.93. The percentage of forecast values with an error less than 5% was 38%, and 90% of the simulated values yielded an error smaller than 20%. To improve this performance, we built an OALC using rainfall as an exogenous predictor. Including this exogenous variable leads to a lower S/σ_{Δ} value of 0.68 and a total of 48% of the forecast values with a prediction error of less than 5%. For 95% of the predictions, the forecast error was less than 20%. The optimal parameters for such performance are θ = 250 and n=22. In this case, we can find more than 4000 OALCs that offer similar performance (Figure 4). For instance, we can use:

$$Y(t+T)=\sum_{i=t}^{t-2} a_i Y_i + \sum_{k=1}^{1}\sum_{j=t}^{t-2} b_{kj} X_{kj}; \qquad \theta = 40. \tag{19}$$

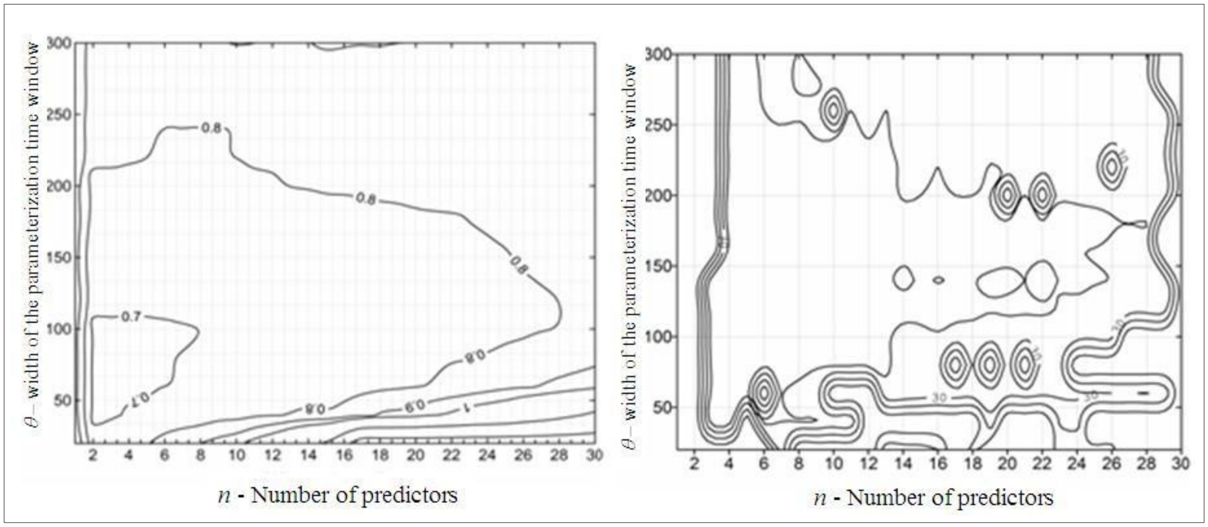

Figure 4. S/σ_{Δ} behavior (left) and percentage of forecasts with error level ≤ 5% (right) using equation (19) to forecast Betania reservoir daily streamflows.

Use of OALC to forecast water levels at Paicol station using PERSIANN potential precipitation estimates as an exogenous predictor

To evaluate the benefit from using satellite information as an exogenous predictor at the Paicol station, we implemented OALC using the potential precipitation estimates from the PERSIANN satellite-based system (Soroosh et al., 2000). To assess this possibility, we have selected the hydrologic station Paicol, which is located within the Betania reservoir catchment area and has available hydrological and meteorological data within the working period of the PERSIANN system. Two types of OALCs were built: the first OALC uses ground-based rainfall measurements from stations 2105031 and 2105502; the second OALC applies the potential precipitation estimates from the PERSIANN satellite-based system. After calculations, we compared the performance for both OALCs and found that for the first, the best S/σ_{Δ} value was 0.94. The percentage of forecasts with a 5% or less error was 56%. The number of predictions with 20% or less error included 99% of the total evaluated forecasts. The structure for this OALC is as follows:

$$Y(t+T)=\sum_{i=t}^{t-5} a_i Y_i + \sum_{k=1}^{1}\sum_{j=t}^{t-1} b_{kj} X_{kj}; \qquad \theta = 280 \tag{20}$$

For the second OALC, using satellite rainfall estimates as exogenous predictors (X_{kj}), we obtained S/σ_{Δ} = 0.84, with 58% of forecasts with an error ≤ 5% and 99.6% of forecasts with an error <20% (Figures 5 and 6). The OALC equation for this case is:

$$Y(t+T)=\sum_{i=t}^{t-9} a_i Y_i + \sum_{k=1}^{1}\sum_{j=t}^{t-6} b_{kj} X_{kj}; \qquad \theta = 240 \tag{21}$$

The presented results not only show that satellite-based information can be used as predictor information but also that such rainfall estimates can even produce a better forecasting performance. However, this should still be tested for different geographical and hydrological conditions. The simultaneous use of ground-based and satellite rainfall information was not evaluated, but we can expect that such a combination of multiple predictors could lead to improved forecast performance.

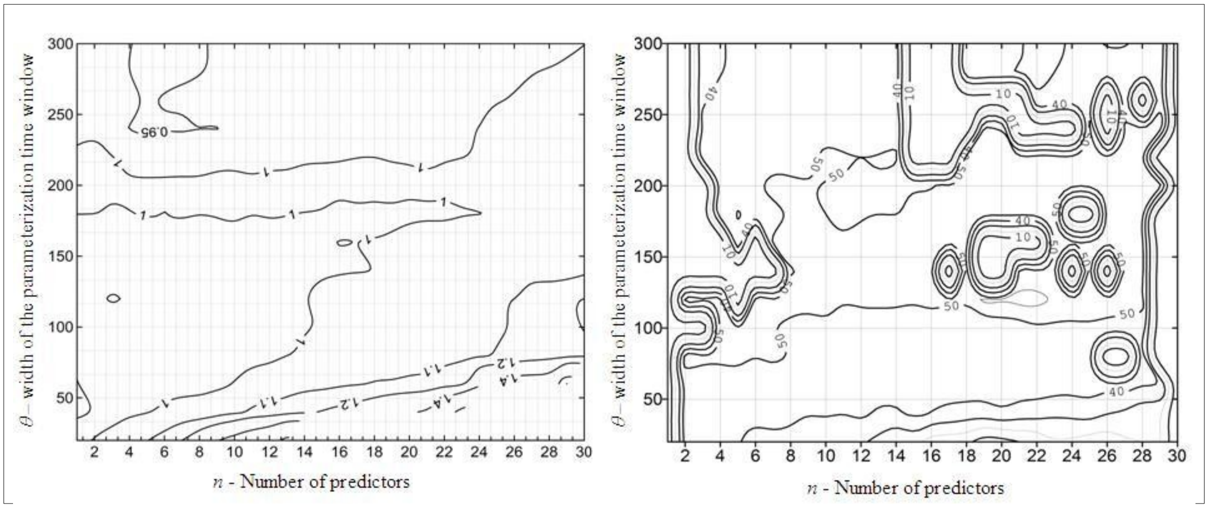

Figure 5. S/σ_Δ behavior (left) and percentage of forecasts with error level ≤ 5% (right) using equation (21) and ground-based rainfall measurements to forecast daily stages at Paicol station.

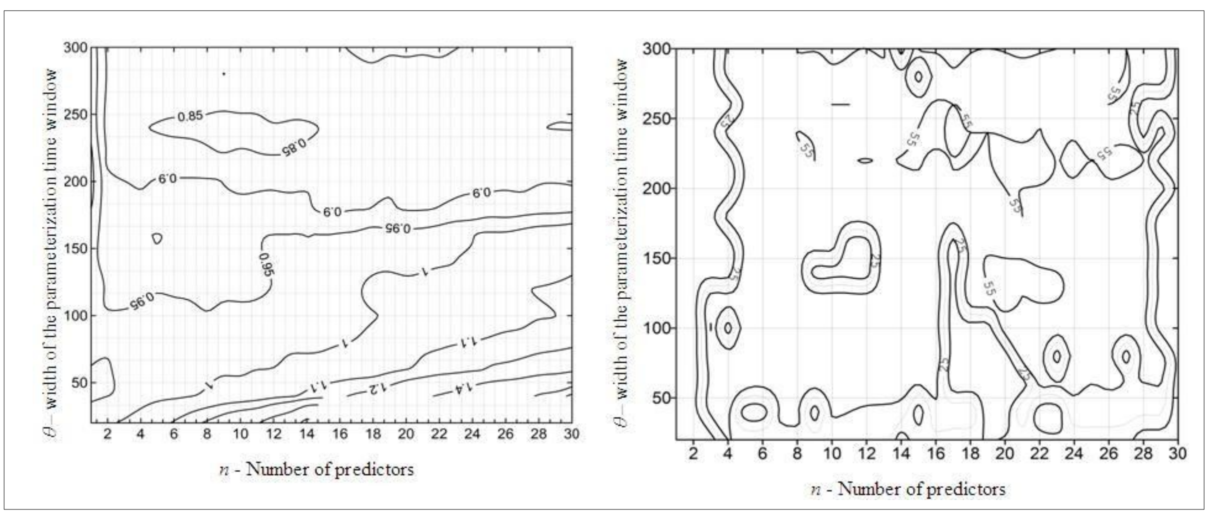

Figure 6. S/σ_Δ behavior (left) and percentage of forecasts with error level ≤ 5% (right) using equation (20) and PERSIANN-based rainfall estimates to forecast daily stages at Paicol station

Use of OALC to forecast daily water levels for El Banco station

Betania reservoir streamflows can be represented as a short-memory Markov process with a two-day autocorrelation radius. To evaluate the performance of OALC when dealing with long memory processes, we established two operators for different lead times (T = 1 and T = 14) at the El Banco hydrological station. The stage autocorrelation function shows an autocorrelation radius of 40 days, which is a pattern of strong inertia. Processes with a high autocorrelation are highly predictable, so the forecasting models for such processes can have very high S/σ_Δ values (in some cases $S/\sigma_\Delta >> 1.0$). Such increased S/σ_Δ values show that the process prehistory by itself contains enough information to forecast the near future of the process. In colloquial terms, this means that in a process with a high autocorrelation radius, the best forecast for a short lead time is the current value, so any implemented model has to be good enough to implement new information about the process dynamics to explain the process beyond the range of influence. Finally, the smaller T is, the smaller σ_Δ becomes (the standard deviation of the increment Δ for the forecasted variable within the lead time). In this case, the standard square error S obtained when forecasting the lead time T with OALC has to be small enough to produce an acceptable value of $S/\sigma_\Delta <= 0.8$. The greater the lead time is, the greater σ_Δ becomes, allowing S to increase while still maintaining good performance values for OALC operators. That means that S/σ_Δ values can decrease when forecasting for lead times $T >> 1$. To check the above statements, we built forecast operators for lead times $T = 1$ and $T = 14$. Exogenous predictors were not used in this implementation of OALC operators. The impact of using rainfall and upstream streamflows as exogenous predictors to improve the forecast performance is a matter of further research.

For the case of a forecasting horizon $T = 1$, the best S/σ_Δ was 0.76, with 99% of forecasts presenting a prediction error ≤ 5% (Figure 7). This performance was reached using $\theta = 200$ and $n = 3$ within the OALC:

$$Y(t+T) = \sum_{i=t}^{t-3} a_i Y_i \tag{22}$$

The forecast performance for a lead time $T = 14$ was also satisfactory, as we obtained $S/\sigma_\Delta = 0.2$ and 90% of forecast values with prediction error <5%. These results were reached with the parameters $\theta = 100$ and $n = 18$ and using the equation:

$$Y(t+T) = \sum_{i=t}^{t-18} a_i Y_i; \qquad \theta = 100 \tag{23}$$

In general, for $T = 14$, there are more than 1200 possible OALCs with $S/\sigma_\Delta = 0.75$ (Figure 8).

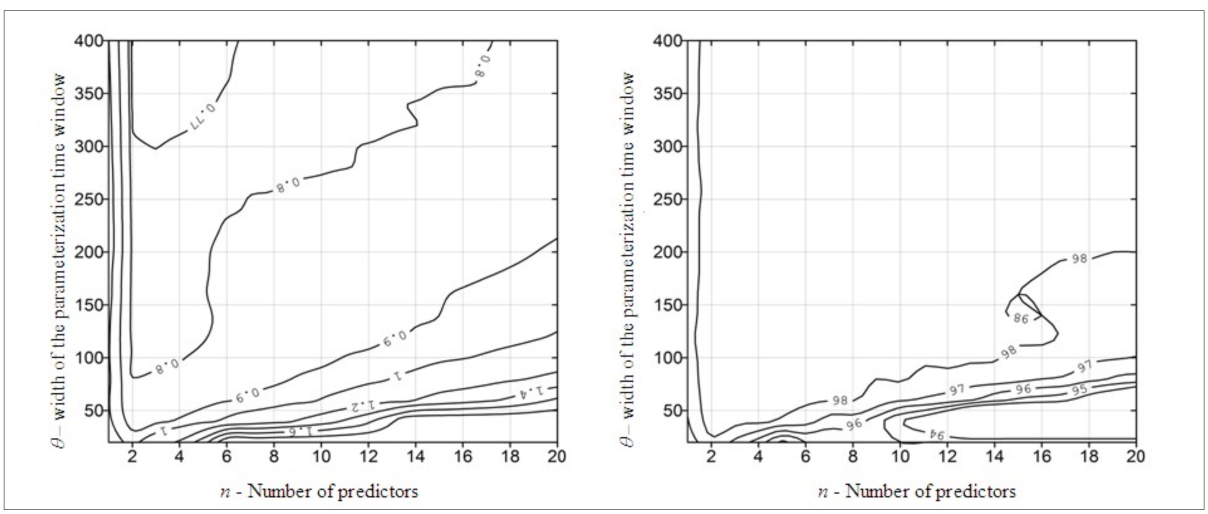

Figure 7. S/σ_Δ behavior (left) and percentage of forecasts with error level ≤ 5% (right) using equation (22) to forecast El Banco daily water levels with $T = 1$.

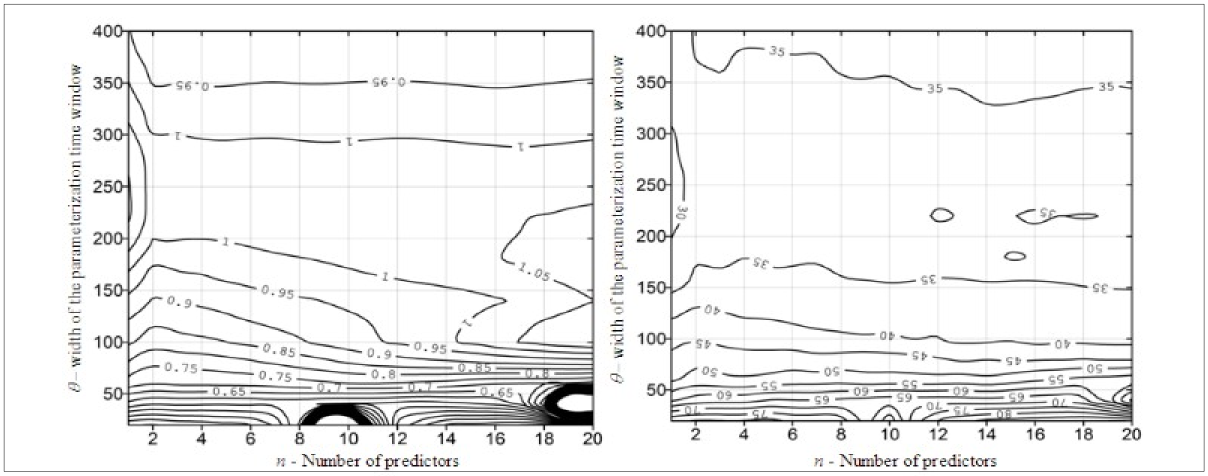

Figure 8. S/σ_Δ behavior (left) and percentage of forecasts with error level ≤ 5% (right) using equation (23) to forecast El Banco daily water levels with T = 14

CONCLUSIONS

The OALC operators have demonstrated satisfactory forecasting performance for all the conditions studied here. In forecasting streamflows for Betania's reservoir and predicting water levels for the El Banco station, we obtained S/σ_Δ values less than 0.8 (Figures 4, 5, 6, 7, 8), as is required to accept a forecast technique as satisfactory (Appolov*, et al.*, 1974). Visually observed and forecasted values show the same time evolution patterns. Analysis of the maximum and minimum observed and forecast values shows no lags in timing and good agreement in terms of magnitude. These facts are confirmed by high coefficients of determination for the line of forecast versus observed values (Figures 9, 10, 11). The forecast performance at Betania's forecast points was lower than for the El Banco station. This decrease in performance is related to the complexity of mountain region where Betania reservoir is located and most likely to the lag of ground-based measurements of precipitation within the catchment area, so the total rainfall input to the system is not available. This may explain the better performance found when forecasting with PERSIANN satellite rainfall estimates as exogenous predictors for the OALC (21).

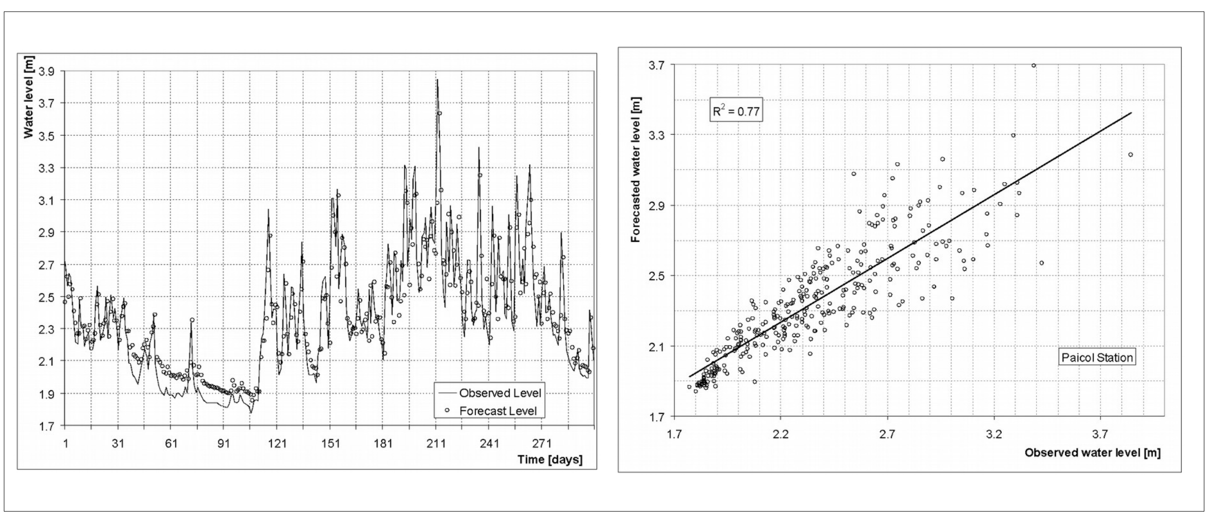

Figure 9. Comparison of forecast and observed water levels (Equation 12) for the Paicol station with lead time T = 1 and using PERSIANN satellite-based precipitation estimates as exogenous predictors.

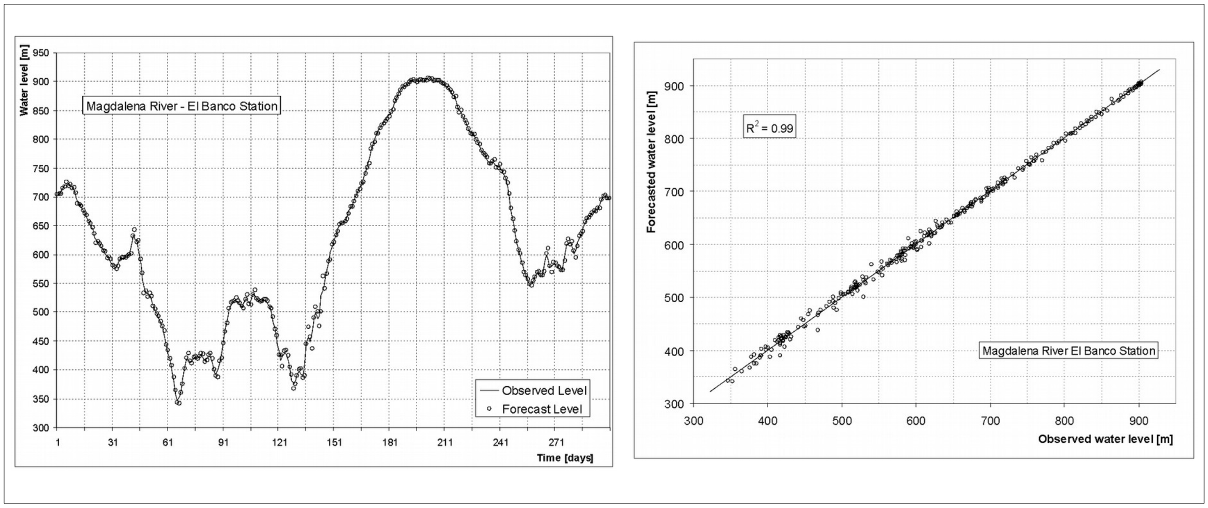

Figure 10. Comparison of forecast and observed water level (Equation 13) for the El Banco station with lead time $T = 1$.

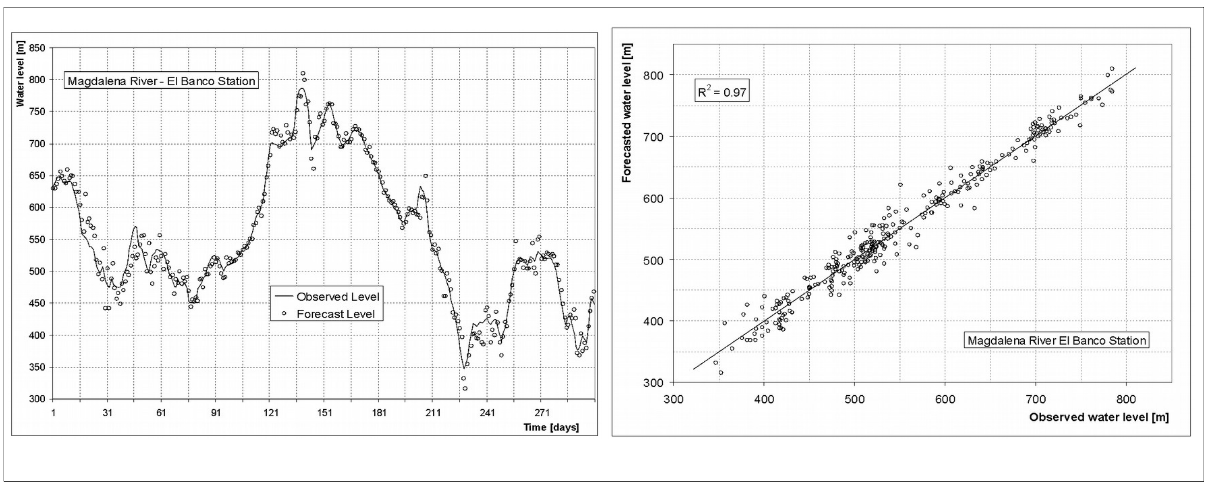

Figure 11. Comparison of forecast and observed water level (Equation 13) for the El Banco station with lead time $T = 14$.

For El Banco's forecast, lead times of $T = 1$ and $T = 14$ were used to compare the performance of short- and long-term predictions for long memory processes. A two-week lead time is more than enough for navigation planning in the Magdalena River, where transit time through the entire navigable sector is less than 1 week; therefore, the forecast technique we present is ready to be used as a real-time planning tool, provided that broadcasting through the internet can be established for the entire navigable sector. The same can be said about streamflow forecasts in Betania's reservoir; probabilistic decision-making procedures may require a stochastic forecasting approach instead of a deterministic one. To develop this type of forecast, this paper shows that OALC can be used as a deterministic kernel for the Fokker-Planck-Kolmogorov equation. Any OALC can be expressed as a system of first-order ODEs (with n equations), allowing the analytical definition of drift (A_i) and diffusion (B_{ij}) coefficients for the multidimensional Fokker -Planck-Kolmogorov equation (13). Upstream inflow data were not used as exogenous predictors, but we can expect that such information will improve the OALC forecast performance. To avoid high computing time when using several exogenous predictors, it is necessary to implement optimization methods such as the conjugate gradient. It is expected that parallel calculations will be mandatory for real time forecasting.

ACKNOWLEDGEMENTS

The present work greatly benefits from the freely available data provided by the "Instituto de Hidrología, Meteorología y Estudios Ambientales – IDEAM" (http:\\www.ideam.gov.co) and also from data obtained from the PERSIANN satellite system, which is supported by the Center for Hydrometeorology & Remote Sensing of the University of California – Irving. Special thanks to Professors Emel Vega (Universidad Nacional de Colombia) and Valentina Shveikina (Water Problems Institute, Russian Academy of Sciences), who kindly reviewed the paper and provided us with valuable comments.

REFERENCES

Alvarado, M., 2006. Río Magdalena – Colombia. Desarrollo de proyecto piloto de navegación satelital – SNS, entre Puerto Berrio (K783) y Regidor (K454), in: Memorias XXII Congreso Latinoamericano de Hidráulica. Ciudad Guayana - Venezuela, p. 10.

Appolov, B., Kalinin, G., and Komarov, V. 1974. Hydrological Forecasting Course. Guidrometeoizdat, Leningrad.

Biondi, D., Freni, G., Iacobellis, V., Mascaro, G., Montanari, A., 2012. Validation of hydrological models: Conceptual basis, methodological approaches and a proposal for a code of practice. Phys. Chem. Earth 42-44, 70–76. doi:10.1016/j.pce.2011.07.037

Capacci, D., Conway, B.J., 2005. Delineation of precipitation areas from MODIS visible and infrared imagery with artificial neural networks. Meteorol. Appl. 291–305.

CGIAR-CSI, 2012. SRTM 90m Digital Elevation Data [WWW Document]. URL http://srtm.csi.cgiar.org (accessed 6.6.12).

Domínguez, E., 2004. Stochastic forecasting of streamflow to Colombian hydropower reservoirs. Hydrophisycs and Hydroforecasting. Russian State Hydrometeorological University, San Petersburg.

Domínguez, E., 2005. Pronóstico probabilístico de afluencias para la evaluación de riesgos en embalses hidroeléctricos. Av. en Recur. Hidráulicos 12, 25.

Domínguez, E., Angarita, H., Ardila, F., Caicedo, F., 2009. Hydrological Risk Modeling using Adaptive Operators: Overview and applications, in: 8th International Conference on Hydroinformatics. IAHR, Concepción, Chile, p. 12.

Domínguez, E., Angarita, H., Rivera, H., 2010. The Feasibility of daily, weekly and ten day water-level forecasting in Colombia. Ing. e Investig. 30, 178–187.

Domínguez, E., Dawson, C.W., Ramírez, A., Abrahart, R.J., 2011. The search for orthogonal hydrological modelling metrics: a case study of 20 monitoring stations in Colombia. J. Hydroinformatics 13, 429. doi:10.2166/hydro.2010.116

Domínguez, E., Kovalenko, V., 2009. Mecanismos de adaptación y bifurcación en la evolución de procesos hidrológicos, in: Complejidad: Revolucion Cientifica y Teoria. Editorial Universidad del Rosario, Bogotá, pp. 59–69.

Domínguez, E., Kovalenko, V., Rivera, H., 2007. The use of conditioned probability density curves for hydropower planning at long, medium and short term time framework : A Colombian case study . Geophys. Res. Abstr. 9, 1–2.

Domínguez, E., Lozano-Báez, S., 2014. Estado del arte de los sistemas de alerta temprana en Colombia. Rev. Acad. Colomb. Cienc. 38, 321–332.

Fenicia, F., Kavetski, D., Savenije, H.H.G., 2011. Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development. Water Resour. Res. 47. doi:10.1029/2010WR010174

IPCC, I.P. on C.C., 2007. Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, United Kingdom and New York, NY, USA.

Kazakievich, D.I., 1989. Concepts from theory of random functions with aplications to hydrometeorological problems. Guidrometeoizdat, Leningrad.

Khaustov, B.A., Kovalenko, V. V, 1998. Stable development criteria of hydrological processes and mapping of expected anomaly for the characteristics of annual runoff of river of the former Soviet Union. Meteorol. i Guid. 12, 96–102.

Kou-Lin, H., Xiaogang, G., Soroosh, S., Hoshin, V.G., n.d. PERSIANN: Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks J. Appl. Meteorol. 1176 - 119.

Mintransporte, 2007. Proyecto YUMA [WWW Document]. URL: http://www.mintransporte.gov.co/Portal_Servicios/proyecto_yuma/proyecto_yuma.asp

NOAA, 1999. The modernized end-to-end forecast process for quantitative precipitation information: hydrometeorological requirements, scientific issues, and service concepts. NOAA, Silver Spring, MD.

Schumann, A.H., 1993. Development of conceptual semi-distributed hydrological models and estimation of their parameters with the aid of GIS. Hydrol. Sci. J. 38, 519–528. doi:10.1080/02626669309492702

Solomatine, D.P., Wagener, T., 2011. 2.16 - Hydrological Modeling, in: Treatise on Water Science. pp. 435–457. doi:http://dx.doi.org/10.1016/B978-0-444-53199-5.00044-0

Soroosh, S., Kuo-Lin, H., Xiaogang, G., Hoshin, V.G., Bisher, I., Braithwaite, D., 2000. Evaluation of PERSIANN System Satellite-Based Estimates of Tropical Rainfall. Bull. Am. Meteorol. Soc. 2035–2045.

Sveshnikov, A.A., 1968. Applied methods of the theory of random functions. Nauka, Moskva.

UPME, Ministerio de Minas y Energía, 2006. Boletín Estadístico de Minas y Energía,. Milenio Editores, Bogotá.

WMO, 1994. Guide to hydrological practice -- Data acquisition and processing, analysis, forecasting and other applications. World Meteorological Organization.

World Commission on Dams, 2000. Dams and Development: A New Framework for Descision-Making.

Xie, W.X., Xu, W., Cai, L., 2005. Path integration of the Duffing - Rayleight oscillator subject to harmonic and stochastic excitations. Appl. Math. Comput. 171, 870 – 884.

Zevin, S.F., 1994. Steps toward an integrated approach to hydrometeorological forecasting services. Bull. Am. Meteorol. Soc. 75, 1267–1276.

ADDRESS FOR CORRESPONDENCE
Efrain A. Dominguez C.
Faculty of Environmental and Rural Studies
Pontificia Universidad Javeriana
Bogota, Colombia

Email: e.dominguez@javeriana.edu.co