

PRONÓSTICO HORARIO DE CAUDALES MEDIANTE FILTRO DE KALMAN DISCRETO EN EL RÍO HUAYNAMOTA, NAYARIT, MÉXICO

HOURLY STREAMFLOW FORECASTING FOR THE HUAYNAMOTA RIVER, NAYARIT, MEXICO, USING THE DISCRETE KALMAN FILTER

Leticia Alvarado-Hernández¹, Laura A. Ibáñez-Castillo^{1*}, Agustín Ruiz-García¹, Fernando González-Leiva², Mario A. Vázquez-Peña¹

¹Posgrado en Ingeniería Agrícola y Uso Integral del Agua, Universidad Autónoma Chapingo, 56230 Chapingo, Estado de México. (libacas@gmail.com). ²Departamento de Ingeniería Hidráulica y Ambiental, Pontificia Universidad Católica de Chile, Santiago Chile.

RESUMEN

Debido a los eventos de precipitación extrema provocada por el cambio climático y a la alteración acelerada de las cuencas por el crecimiento poblacional, es importante pronosticar los caudales que generan las cuencas por los eventos de precipitación. El objetivo de este estudio fue predecir caudales horarios en la cuenca del río Huaynamota usando el Filtro de Kalman Discreto (DKF) junto con un modelo autorregresivo con entrada exógena (ARX). Al inicio los parámetros del filtro de Kalman se definen y después se recalculan por periodos definidos, es decir los valores de los parámetros del modelo se actualizan constantemente. El pronóstico de caudales se realizó en seis pasos hacia adelante (L=1, 2, 3, 4, 5 y 6 horas). La cuenca de estudio es parte del río Huaynamota, delimitada por la estación hidrométrica Chapalagana, aguas arriba de la presa Aguamilpa, en Nayarit, México. La cuenca del río Huaynamota es un tributario del río Santiago. Series de datos horarias se emplearon para precipitación y caudal, de agosto a septiembre del 2017. El modelo de pronóstico DKF-ARX mostró índices de eficiencia de Nash-Sutcliffe entre 0.99 y 0.85 con L=1 y L=6, respectivamente. Se concluye que es factible obtener un buen pronóstico de caudales horarios con filtro de Kalman discreto.

Palabras clave: filtro de Kalman discreto, modelos autorregresivos, predicción caudal.

INTRODUCCIÓN

El conocimiento de la función de respuesta de la cuenca es necesario para planear y manejar los recursos hídricos, sin embargo, por la

ABSTRACT

Because of extreme rainfall events caused by climate change and of accelerated alteration of basins by population growth, it is important to forecast streamflow generated by precipitation events. The objective of this study was to predict hourly flows in the Huaynamota River basin using the Discrete Kalman Filter (DKF), together with the autoregressive exogenous input model (ARX). Initially, the Kalman filter parameters are defined then recalculated for defined periods; that is, the model parameter values are constantly updated. Flows were forecasted six steps ahead (L=1, 2, 3, 4, 5 and 6 hours). The basin studied is part of the Huaynamota River, delimited by the Chapalagana hydrometric station, upstream from the Aguamilpa reservoir, Nayarit, Mexico. The Huaynamota River is a tributary of the Santiago River. Hourly data series were used for precipitation and flow from August to September 2017. The DKF-ARX forecasting model showed Nash-Sutcliffe efficiency indexes between 0.99 and 0.85, with L=1 and L=6, respectively. It is concluded that it is feasible to obtain a good forecast of hourly streamflow with the discrete Kalman filter.

Key words: discrete Kalman filter, autoregressive models, flow prediction.

INTRODUCTION

Knowledge of the basin response function is necessary to plan and manage water resources. However, because of the complexity of the processes that intervene in the water cycle and their superficial and sub-superficial interrelations, abstraction is necessary to determine and control some aspects of its behavior: the object of watershed modeling (Chong-yu Xu, 2002).

* Autor para correspondencia ♦ Author for correspondence.

Recibido: febrero, 2019. Aprobado: noviembre, 2019.

Publicado como ARTÍCULO en *Agrociencia* 54: 295-312. 2020.

complejidad de los procesos que intervienen en el ciclo hidrológico y sus interrelaciones superficiales y subsuperficiales, la abstracción es necesaria para conocer y controlar algunos aspectos de su comportamiento, lo cual es objeto de la modelación de cuencas (Chong-yu Xu, 2002).

Los modelos del proceso lluvia-escurrimiento se desarrollaron desde el siglo XIX, Vargas-Castañeda *et al.* (2015) realizaron una revisión histórica de este tipo de modelos resaltando las tendencias actuales. Los modelos más completos son los llamados distribuidos de base física como los indicados por Devi *et al.* (2015). Los modelos distribuidos son más complejos que los agregados, aunque no necesariamente los más eficaces (Beven, 1989; Mendoza *et al.*, 2002). Estos autores indican que esta complejidad de los modelos distribuidos se debe a las simplificaciones de sus ecuaciones, la no linealidad de los procesos individuales del ciclo hidrológico, así como la falta de información de las condiciones físicas de la cuenca y datos climatológicos. Algunos ejemplos de este tipo de modelos son los desarrollados por Vargas-Castañeda *et al.* (2018) y Zhen-lei *et al.* (2018).

Uno de los mayores inconvenientes para la generación de modelos de lluvia-escurrimiento de cualquier tipo, es la falta de información para alimentarlos, así como para calibrarlos y validarlos. Las bases de datos climatológicos horarios generados en México son principalmente precipitación y caudales de escurrimiento, sobre todo en cuencas con gran inversión económica, como por ejemplo las monitoreadas por la Comisión Federal de Electricidad (CFE) en donde existen presas para la generación de energía eléctrica; éste es el caso de la cuenca del río Santiago, del cual forma parte la zona de estudio en esta investigación. Sobre el río Santiago, en Nayarit, hay tres presas para generación de energía eléctrica: Aguamilpa, El Cajón y la Yesca (Comisión Federal de Electricidad, 2017). Al considerar este tipo de información para predecir la respuesta de la cuenca, se desarrollaron modelos de caja negra como el filtro de Kalman.

El filtro de Kalman es un sistema de ecuaciones matemáticas que implementa un estimador tipo predictor-corrector que es óptimo en el sentido que minimiza el error estimado de la covarianza (Castañeda *et al.*, 2013). Este tipo de algoritmo se utiliza en una amplia gama de temas como edafología, biología, medicina, hidrología, etc. (Valdés *et al.* 1980; Hernández y Medina, 2012; Padilla *et al.*, 2013; Hobson y Kardynal, 2015).

Models of the rainfall-runoff process were developed since the 19th century. Vargas-Castañeda *et al.* (2015) carried out a historic review of this type of model, highlighting the current trends. The most complete models are those called physically based distributed models, such as those described by Devi *et al.* (2015). Distributed models are more complex than aggregated models, though not necessarily more effective (Beven, 1989; Mendoza *et al.*, 2002). These authors indicate that the complexity of distributed models is due to the simplifications of their equations, the non-linearity of the individual processes of the water cycle, as well as the lack of information of the physical conditions of the basin and climatological data. Examples of this type of model were developed by Vargas-Castañeda *et al.* (2018) and Zhen-lei *et al.* (2018).

One of the greatest inconveniences in generating rainfall-runoff models of any type is the lack of data to feed them, as well as to calibrate and validate them. The hourly climate databases generated in México are mainly rainfall and runoff flows, especially in basins where economic investment is high, such as those monitored by the Federal Electricity Commission (CFE) where there are hydroelectric dams: Aguamilpa, El Cajón and La Yesca (Comisión Federal de Electricidad, 2017). Considering this type of information for forecasting basin response, black box models like the Kalman filter were developed.

The Kalman filter is a system of mathematical equations that implement a predictor-corrector type estimator, which is optimal in the sense that it minimizes the estimated error of the covariance (Castañeda *et al.*, 2013). This type of algorithm is used in a wide variety of fields, such as edaphology, biology, medicine, hydrology, etc. (Valdés *et al.* 1980; Hernández and Medina, 2012; Padilla *et al.*, 2013; Hobson and Kardynal, 2015).

Morales-Velázquez *et al.* (2014) applied the discrete Kalman filter to predict hourly flows in the watershed of the Ángel Albino Corso (Peñitas) dam. The model was evaluated using the flows obtained from the modified inverse routing reservoir, measured at the Sayula hydrometric station and obtained Nash-Sutcliffe efficiency indexes of 0.98 for hourly flow forecast. González-Leiva *et al.* (2015) implemented a discrete Kalman Filter model, autoregressive exogenous input (DKF-ARX) to predict daily mean streamflow in the Turbio River basin, Guanajuato;

Morales-Velázquez *et al.* (2014) aplicaron el filtro de Kalman discreto para predecir caudales horarios en la cuenca de la presa Ángel Albino Corso (Peñitas); el modelo se evaluó a partir de caudales obtenidos del tránsito inverso modificado en vasos o antitránsito medidos en la estación hidrométrica Sayula, y tuvo valores del índice de eficiencia de Nash-Sutcliffe para pronóstico de caudal a cada hora de 0.98. González-Leiva *et al.* (2015) implementaron un modelo de Filtro de Kalman discreto, autorregresivo y entrada exógena (DKF-ARX) para predicción de caudal medio diario en la cuenca del río Turbio, Guanajuato; no se implementó una predicción horaria, porque en la estación de aforo trabajada, Las Adjuntas, no existen datos medidos horarios; la cuenca del río Turbio no es de interés hidroeléctrico para la CFE, por lo que no hay medición continua de caudales. González-Leiva *et al.* (2015) consideraron cuatro pasos de tiempo de anticipación (L1, L2, L3 y L4), es decir los tiempos en que no se ejecutó la fase de corrección del modelo DKF; de los dos periodos analizados el del año 2003 presenta valores del índice de eficiencia de Nash-Sutcliffe de 0.95, 0.87, 0.76 y 0.63 para los distintos pasos de anticipación, y análogamente para la serie de datos de 2004 los valores de Nash-Sutcliffe fueron 0.93, 0.82, 0.72 y 0.62.

Al considerar los cambios actuales de los regímenes de precipitación y por ende de la función de respuesta de la cuenca a fenómenos extremos, se hace necesario el desarrollo de modelos alimentados con información horaria, que puedan generar información fehaciente para utilizar en los sistemas de alerta temprana o en los planes de operación del sistema de presas hidroeléctricas. Así el objetivo de esta investigación fue evaluar, a partir de medidas de bondad de ajuste, el pronóstico de caudales horarios mediante un modelo autorregresivo con entrada exógena (ARX) en conjunto con filtro de Kalman discreto (DKF), pronosticados a 1, 2, 4 y 6 h de anticipación, en la cuenca del río Huaynamota en los estados de Durango, Nayarit, Jalisco y Zacatecas. La hipótesis fue que el filtro de Kalman discreto puede ser una buena herramienta para pronosticar caudales horarios.

MATERIALES Y MÉTODOS

El algoritmo de filtro de Kalman discreto (DKF) junto con un modelo autorregresivo con entrada exógena (ARX) se utilizó

hourly prediction was not implemented because at the gauging station Las Adjuntas there are no measured hourly data. The Turbio River basin is not of hydroelectric interest for the CFE and thus there is no continuous measurement of flows. González-Leiva *et al.* (2015) considered four anticipation times (L1, L2, L3 and L4); that is, the times at which the correction phase of the DKF model was not executed. Of the two periods analyzed, 2003 had Nash-Sutcliffe efficiency indexes of 0.95, 0.87, 0.76 and 0.63 for the different anticipation times, and analogously, for the 2004 data series, the Nash-Sutcliffe values were 0.93, 0.82, 0.72 and 0.62.

Considering the current changes in precipitation regimes and, therefore, in the basin's response function to extreme phenomena, it is necessary to develop models fed hourly information that can generate reliable information to use in early warning systems or in system operation plans of hydroelectric dams. Thus, the objective of this research was to evaluate, using goodness of fit measurements, hourly streamflow forecasts 1, 2, 4 and 6 h in anticipation, of an autoregressive model with exogenous input (ARX) in conjunction with a discrete Kalman filter, in the Huayanamota River basin in the states of Durango, Nayarit, Jalisco and Zacatecas. The hypothesis was that the discrete Kalman filter can be a good tool for forecasting hourly streamflow.

MATERIALS AND METHODS

The discrete Kalman filter (DKF) algorithm, together with an autoregressive model with exogenous input (ARX), was used to predict flows, considering past time series of rainfall and flows; rainfall is in mm and flows are in $\text{m}^3 \text{s}^{-1}$. The current considered, Chapalagana or Atengo River, is one of the main tributaries of the Huaynamota River basin, delimited by the Chapalagana hydrometric station, Nayarit, Mexico. The other tributary of the Huaynamota is the Jesús María River. The Chapalagana River (12 080 km^2) approximately 4 km downstream of the Chapalagana station joins the Jesús María River (5185 km^2) and together form the Huaynamota River; 25 km downstream after the union is the Aguamilpa hydroelectric dam (INEGI, 2018).

Huaynamota River Basin

The Huaynamota River basin in central Mexico (21° 24' 36.82" and 23° 25' 3.26" N and 104° 30' 34.03" and 103° 24' 26.73" W) includes the states of Durango, Jalisco, Zacatecas

para predecir los caudales, considerando series de tiempo pasadas de precipitación y caudales; las precipitaciones están en mm y los caudales están en $\text{m}^3 \text{s}^{-1}$. La corriente considerada, río Chapalagana o río Atengo, es uno de los dos principales tributarios de la cuenca del río Huaynamota, delimitada por la estación hidrométrica Chapalagana en Nayarit, México; el otro tributario de la cuenca Huaynamota es el río Jesús María. De hecho, aproximadamente 4 kilómetros aguas abajo, de la estación Chapalagana, al continuar su rumbo el río Chapalagana ($12\,080 \text{ km}^2$) se une al río Jesús María (5185 km^2) y juntos forman el río Huaynamota. Una vez unidos, 25 kilómetros aguas abajo, está la presa hidroeléctrica Aguamilpa (INEGI, 2018).

Cuenca Huaynamota

La cuenca del río Huaynamota en el centro de México ($21^\circ 24' 36.82''$ y $23^\circ 25' 3.26''$ N y $104^\circ 30' 34.03''$ y $103^\circ 24' 26.73''$ O), abarca parte de los estados de Durango, Zacatecas y Nayarit (Figura 1). El área de la cuenca es $12\,080 \text{ km}^2$, las elevaciones van de 216 a 3148 msnm, y la elevación mínima corresponde a la salida de la cuenca en la estación hidrométrica Chapalagana. La pendiente media de la cuenca es 31%. La

and Nayarit (Figure 1). The basin covers $12\,080 \text{ km}^2$ at altitudes of 216 to 3148 m; the lowest elevation is the outlet at the Chapalagana hydrometric station. The mean slope of the basin is 31%. The main channel is 321 km and has a slope of 0.68%. The basin concentration time is 44 h.

Climate information

Average annual rainfall in the basin is 600 mm, most between June and October. For this study, we selected a period of analysis of 1360 h, covering 09:00 h 02 August to 00:00 h 28 September 2017. This period was selected because it is one of the periods with the largest streamflow values during the year (two floods in the order of $1171 \text{ m}^3 \text{s}^{-1}$ and $1070 \text{ m}^3 \text{s}^{-1}$). Another criterion was the availability of complete information within the basin; that is, it was selected based on two criteria: 1) extreme rainfall or streamflow events, or both, and 2) access to more recent and complete historic information.

Hydroclimatological information of the study period 2017 was obtained from the network of hydroclimatological stations of the Comisión Federal de Electricidad (2017) located in the Santiago River basin: Jesús María, Chapalagana, Bolaños,

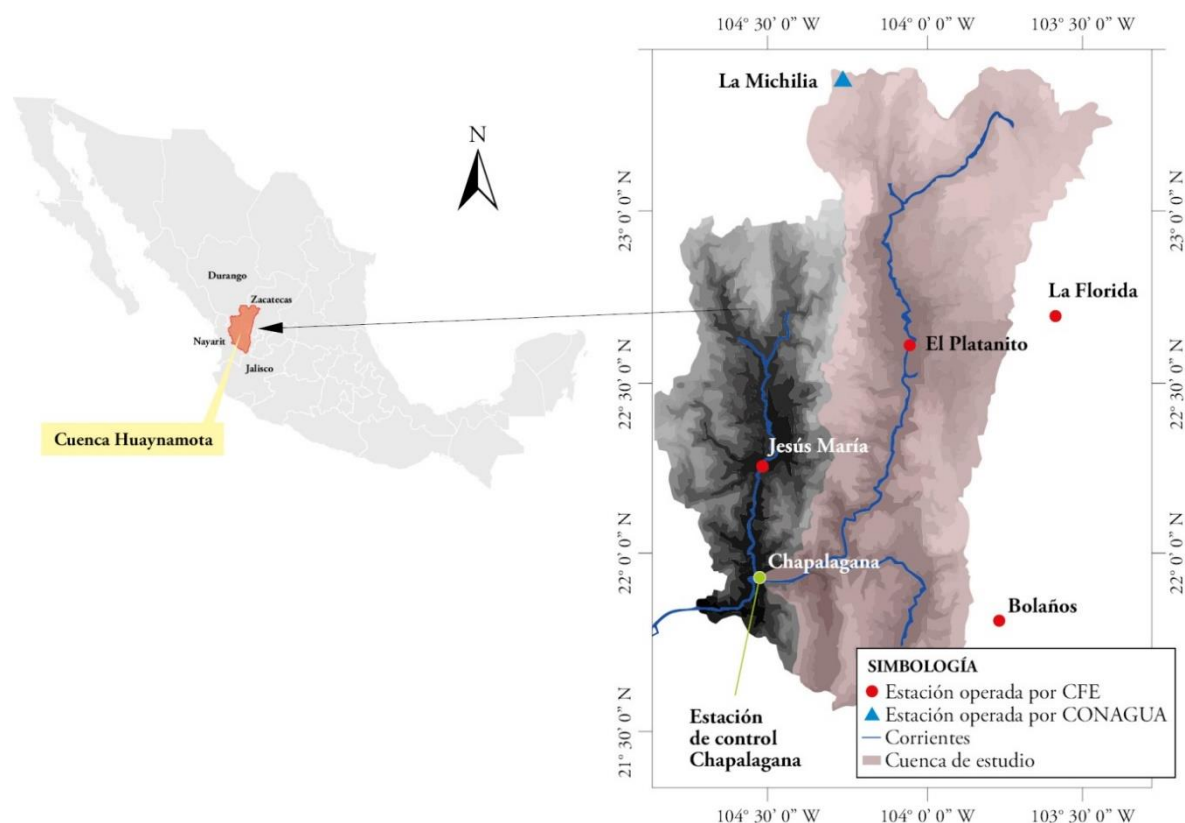


Figura 1. Ubicación de la cuenca del río Huaynamota.
Figure 1. Location of the Huaynamota River.

longitud y pendiente de su cauce principal son 321 km y 0.68%. El tiempo de concentración de la cuenca es 44 h.

Información climatológica

La lluvia promedio anual en la cuenca es 600 mm, la mayor parte entre junio y octubre. Para esta investigación se eligió un periodo de análisis de 1360 h, desglosado a nivel hora y comprendido entre el 02 de agosto a las 09:00 h y el 28 de septiembre a las 00:00 h de 2017. Este periodo se eligió por ser uno de los periodos con valores de caudales más grandes durante el año (dos avenidas máximas de $1171 \text{ m}^3 \text{ s}^{-1}$ y de $1070 \text{ m}^3 \text{ s}^{-1}$). Otro criterio fue la disponibilidad completa de la información dentro de la cuenca, o sea se eligió con base a dos criterios: 1) eventos máximos de lluvia o caudal o ambos, y 2) contar con la información histórica más reciente y completa.

La información hidroclimatológica del periodo de estudio 2017 se obtuvo de la red de estaciones hidroclimatológicas de la Comisión Federal de Electricidad (2017) ubicadas en la cuenca del río Santiago, y las estaciones son: Jesús María, Chapalagana, Bolaños, Platanito y Florida. Además se usó la información de lluvia registrada en la estación meteorológica automática la Michilia, Durango, operada por el Servicio Meteorológico Nacional, SMN (2017). En el Cuadro 1 se muestran las coordenadas de ubicación de las estaciones de registro de lluvia y caudal. El periodo de la información para alimentar al modelo corresponde a la precipitación media de la cuenca calculada mediante el método de polígonos de Thiessen a partir de los datos de precipitación horaria en cada estación; además, se obtuvieron los caudales de escurrimiento en la estación Chapalagana, la cual es la salida de la cuenca.

Metodología

Modelo autorregresivo con entrada exógena (ARX)

El modelo ARX relaciona las entradas con las salidas del sistema mediante una ecuación lineal (ecuación 1) en diferencias

Platanito and Florida. We also used the rainfall information recorded in the automated weather station La Michilia, Durango, operated by the Servicio Meteorológico Nacional, SMN (2017). Table 1 shows the coordinates where the rainfall and flow gauging stations are located. The information to feed the model is mean rainfall in the basin during the study period calculated with the Thiessen polygon method using data on hourly precipitation in each station. The runoff flows from the Chapalagana station at the basin outlet were also obtained.

Methodology

Autoregressive model with exogenous input (ARX)

The ARX model relates the system's input with output using a linear equation (equation 1) in differences with constant coefficients; that is, programmed in OFF LINE (Hsu *et al.*, 2009).

$$y_{t+1} = \sum_{i=0}^{na} \alpha_i y_{t-i} + \sum_{j=0}^{nb} \beta_j r_{t-j} + e_{t+1} \quad (1)$$

where:

y_t is the vector that represents the output variable, which is the runoff flow at the Chapalagana station.

r_t is the vector that represents the exogenous input variable, precipitation at time t .

e_{t+1} is the term of error in flow estimation.

α_i and β_j are parameters to determine how, indicated in the following paragraph.

The indexes na and nb indicate the order of the model; na is the number of delays of the variable flow, and nb is the number of delays of the external variable precipitation.

The ARX model was programmed in Matlab® with the Toolbox systems identification (Ljung, 2017) with which the order of the model was determined (na , nb), as well as the

Cuadro 1. Estaciones de registro horario de lluvia y caudal.
Table 1. Gauging stations that register hourly rainfall and flow.

Estación	Tipo	Latitud	Longitud	Operador
Chapalagana, Nayarit	hidroclimatológica	21.945	-104.508	CFE
Bolaños, Jalisco	hidroclimatológica	21.825	-103.783	CFE
La Florida, Zacatecas	hidroclimatológica	22.686	-103.602	CFE
Jesús María, Nayarit	hidroclimatológica	22.2552	-104.516	CFE
El Platanito, Zacatecas	hidroclimatológica	22.6113	-104.050	CFE
La Michilia, Durango	Meteorológica	23.3875	-104.247	SMN

con coeficientes constantes, es decir programado en OFF LINE (Hsu *et al.*, 2009).

$$y_{t+1} = \sum_{i=0}^{na} \alpha_i y_{t-i} + \sum_{j=0}^{nb} \beta_j r_{t-j} + e_{t+1} \quad (1)$$

donde:

y_t es el vector que representa la variable de salida que es el caudal de escurrimiento en la estación Chapalagana.

r_t es el vector que representa la variable de entrada exógena y es la precipitación en el tiempo t ,

e_{t+1} es el término de error en la estimación del caudal.

α_i y β_j son parámetros para determinar como se indica en el siguiente párrafo.

Los índices na y nb indican el orden del modelo; na es el número de retrasos de la variable caudal y nb el número de retrasos de la variable externa precipitación.

El modelo ARX se programó en Matlab® con el Toolbox identificación de sistemas (Ljung, 2017) con lo cual se determinó el orden del modelo (na , nb), así como los parámetros a y b , en el caso de nk para estudios hidrológicos se considera uno, a no ser que existan presas de almacenamiento que retrasan la respuesta del sistema lluvia-escurrimiento. Ljung (2017) indica que debido al uso del mismo conjunto de datos para la estimación y validación, se deben utilizar los criterios de Rissanen MDL e información de Akaike (AIC) para obtener las órdenes del modelo.

Filtro de Kalman Discreto (DKF)

El Filtro de Kalman es una técnica de asimilación de datos que opera en dos fases: predicción y corrección. Solera (2003) indica que el algoritmo pronostica el estado del sistema en el tiempo t a partir de información en $t-1$ y añade un término de corrección proporcional al error de predicción; este último se minimiza estadísticamente. La Figura 2 muestra la idea simplificada del filtro de Kalman Discreto (Welch y Bishop, 2006); además, ellos indican que el Filtro de Kalman discreto estima el estado $X \in R_n$ de un proceso controlado en un tiempo discreto a partir de una ecuación lineal estocástica (ecuación 2):

$$x_t = Ax_{t-1} + BU_t + w_t \quad (2)$$

Con una medida del estado $z \in R_m$, que es la ecuación 3:

$$z_t = Hx_t + v_t \quad (3)$$

parameters a and b . In the case of nk for hydrological studies, it is considered 1, unless there are reservoirs that delay the response of the rain-runoff system. Ljung (2017) indicates that, due to the use of the same data set to estimate and validate, the Rissanen MDL and Akaike information criteria should be used to obtain the orders of the model.

Discrete Kalman Filter (DKF)

The Kalman filter is a data assimilation technique that operates in two phases: prediction and correction. Solera (2003) indicates that the algorithm forecasts the state of the system in time t from information in $t-1$ and adds a correction term proportional to the prediction error, which is minimized statistically. Figure 2 shows Welch and Bishop's (2006) simplified idea of the discrete Kalman filter. Moreover, they indicate that the discrete Kalman filter estimates the $X \in R_n$ state of a controlled process in a discrete time using a linear stochastic equation (equation 2):

$$x_t = Ax_{t-1} + BU_t + w_t \quad (2)$$

With a measure of the $z \in R_m$ status, which is equation 3:

$$z_t = Hx_t + v_t \quad (3)$$

Matrix A ($n \times n$) relates the state in the previous step $t-1$ to the state of the current step. Matrix B ($n \times l$) relates the control input, which in this case is the exogenous variable. Matrix H ($m \times n$) relates the state to measurement.

The variables w_t and v_t represent the process and measurement errors, respectively, which are assumed as independent of each other; they are white noise and fit a Gaussian probability distribution function with a mean of zero and variance Q and R for the noise of the process and measurements, respectively.

The covariance matrixes of disturbance of the Q process and R measurement can change over time, but it is assumed that they are constant. Valdés *et al.* (1980) defines the R matrix as $\propto Q_{t-1}$, that is, the variance corresponds to a value that is proportional to flow, and González-Leiva *et al.* (2015) indicate that the value of a is 5%, which is the error likely committed in measurement of the flow at the gauging station.

As indicated above, the Kalman filter operates in two phases, and the equations used at both times are presented in the Figure 2 algorithm.

Valdés *et al.* (1980) and Morales-Velázquez *et al.* (2014) indicate that the considerations for the initial state referent to the covariance matrix can assume the value of $P_0=KI$, where K is a scale large enough to reflect the uncertainty of the assumed

La matriz A ($n \times n$) relaciona el estado en el paso anterior $t-1$ al estado del paso actual, la matriz B ($n \times l$) relaciona la entrada de control, en este caso corresponde a la variable exógena, y la matriz H ($m \times n$) relaciona al estado con la medición.

Las variables w_t y v_t representan el error del proceso y de las mediciones, respectivamente, las cuales se asumen como independientes entre ellas, son ruido blanco y se ajustan a una función de distribución de probabilidad Gaussiana con media cero y varianza Q y R para el ruido del proceso y mediciones, respectivamente.

Las matrices de covarianza de perturbación del proceso Q y medición R pueden cambiar en el tiempo, pero se asume que son constantes. Valdés *et al.* (1980) definen la matriz R como $\propto Q_{t-1}$, es decir la varianza corresponde a un valor proporcional al caudal, y González-Leiva *et al.* (2015) indican que el valor de α es 5% que representa el error que probablemente se haya cometido en la medición del caudal en la estación hidrométrica.

Como ya se indicó, el filtro de Kalman opera en dos fases, y las ecuaciones que se emplean en ambos tiempos se presentan en el algoritmo de la Figura 2.

Valdés *et al.* (1980) y Morales-Velázquez *et al.* (2014) indican que las consideraciones para el estado inicial, referentes a la matriz de covarianza, puede asumir el valor de $P_0=KI$, donde K es un escalar lo suficientemente grande como para reflejar la incertidumbre de los valores supuestos en el estado inicial, se asume un valor de 1000, e I es la matriz identidad.

values in the initial state; a value of 1000 is assumed and I is the identity matrix.

Model ARX+DKF

To determine the possible orders of the ARX model with the system identification Toolbox, three subsets of data were analyzed to estimate parameters and validate the ARX model: (1) 10:90, (2) 25:75, and (3) 50:50 (%;%). For example, the case of 25%:75%, means that 25% of the total data were used to estimate parameters and 75% to validate the model. Of course, the AIC and MDL criteria are considered in all cases.

The model ARX+DKF initially uses a period for estimating the parameters a and b of the ARX model; then the DKF model is used to forecast streamflow. Because of climate variability and changes in physical conditions of the basin that could change its response function, we recalculated the a and b parameters of the ARX model every certain time (Figure 3).

To determine the orders of the more adequate ARX model (na and nb), in the combined ARX+DKF model the periods for parameter estimation were evaluated 10% (136 h), 25% (340 h) and 50% (680 h), and the orders of the model determined with the system identification Toolbox were evaluated to forecast at one step ahead of $L=1$. Once the period and the orders of the more adequate model were selected, the flow forecasts were evaluated at different forward passages of time (L).

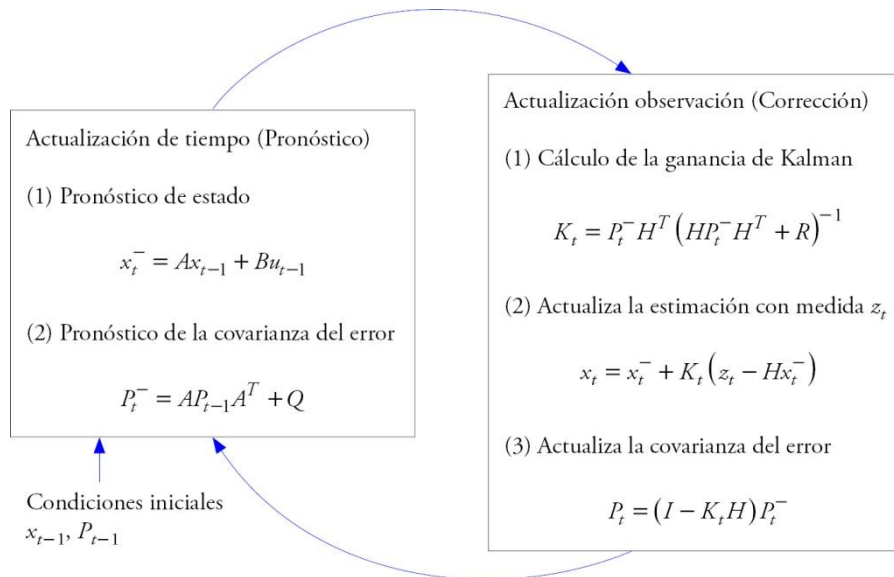


Figura 2. Algoritmo del filtro de Kalman Discreto (Welch y Bishop, 2006).
Figure 2. Discrete Kalman filter algorithm (Welch and Bishop, 2006).

Modelo ARX+DKF

Para determinar las posibles órdenes del modelo ARX con el Toolbox identificación de sistemas se analizaron tres subconjuntos de datos para estimar parámetros y validar el modelo ARX: (1) 10:90, (2) 25:75, y (3) 50:50 (%; %). Por ejemplo, en el caso 25%:75%, significa que se usó el 25% del total de los datos para estimar parámetros y el 75% para validar el modelo. Desde luego se consideran los criterios de AIC y MDL en todos los casos.

El modelo en conjunto ARX+DKF utiliza al inicio un periodo para la estimación de los parámetros a y b del modelo ARX, y después se utiliza el modelo DKF para el pronóstico de caudales. Debido a la variabilidad climática y al cambio de las condiciones físicas de la cuenca que pudieran cambiar la función de respuesta de la misma, se planteó recalcular los parámetros a y b del modelo ARX cada cierto periodo (Figura 3).

Para determinar las órdenes del modelo ARX más adecuado (na y nb) se evaluaron en el modelo en conjunto ARX+DKF los periodos para estimación de parámetros; 10% (136 h), 25% (340 h) y 50% (680 h), y las órdenes del modelo determinadas con el Toolbox identificación de sistemas, se evaluaron para el pronóstico a un paso hacia adelante $L=1$. Una vez elegido el periodo y órdenes del modelo más adecuados se realizó la evaluación de pronóstico de caudales a distintos pasos de tiempo (L) hacia adelante.

En la aplicación del modelo DKF la etapa de pronóstico se realizó 6 pasos (L) hacia adelante, es decir 1, 2, 3, 4, 5 y 6 h, después se realizó la etapa de actualización para el caudal pronosticado en el tiempo $L=1$ y a partir de éste se pronosticó de nuevo seis pasos hacia adelante, como se muestra en la Figura 4.

When applying the DKF model, the forecasting stage carried out 6 forward steps (L), that is 1, 2, 3, 4, 5, and 6 h. The updating stage was then carried out for the predicted flow at time $L=1$ and from this forecast, six steps ahead were again forecasted, as shown in Figure 4.

Statistical analysis

Krause *et al.* (2005) suggest Nash-Sutcliffe efficiency criteria (E) and the root mean square of the error ($RMSE$) as efficiency criteria for evaluating hydrological models.

RESULTS AND DISCUSSION

To feed the model, 1360 data were introduced every hour (Figure 5). These data reflect rainfall and response functions of several events in the Huaynamota River Basin.

Table 2 presents the orders of the ARX model obtained from the subset of percent estimation-validation data, as well as the results of the indexes of efficiency necessary to evaluate the results of the ARX+DKF model, considering an $L=1$ forward step in time. Table 2 will serve as a base to decide which is the best model in a later analysis.

The graphs of the predicted versus observed flows, evaluated with the model orders under the AIC criterion, show 10%:90% and 25%:75% for the first two subsets, large variations in forecasted values at the end of the period of analysis. These differences in

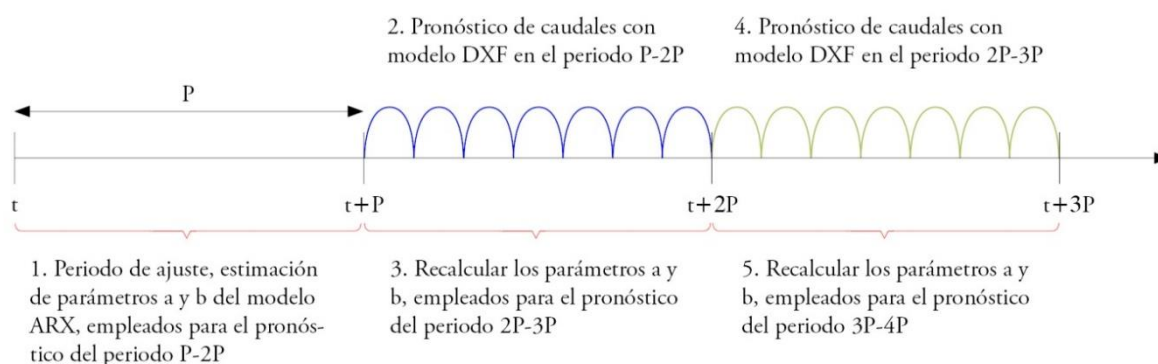


Figura 3. Proceso de estimación de parámetros y pronóstico de caudales.
Figure 3. Process of estimating parameters and forecasting stream flows.

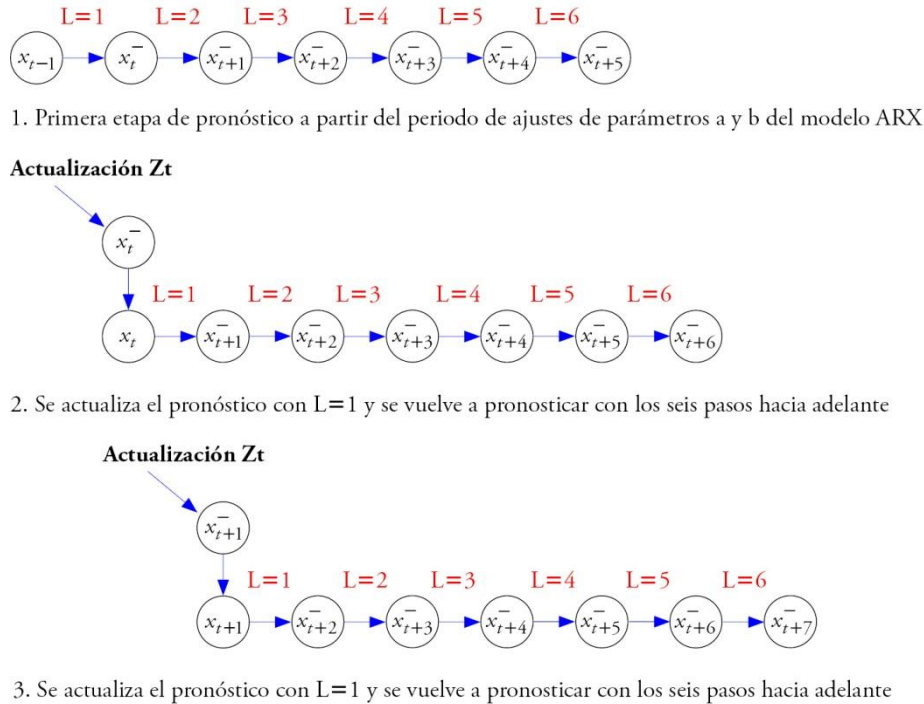


Figura 4. Pronóstico de caudales L pasos hacia adelante.

Figure 4. Forecasting flows L steps ahead.

Análisis estadístico

Krause *et al.* (2005) sugieren como criterios de eficiencia para evaluar modelos hidrológicos los criterios de eficiencia de Nash-Sutcliffe (E) y la raíz del error cuadrático medio ($RMSE$).

RESULTADOS Y DISCUSIÓN

Cada hora se introdujeron 1360 datos para alimentar al modelo (Figura 5), los cuales reflejan las precipitaciones y las funciones de respuesta de varios eventos en la cuenca del río Huaynamota.

El Cuadro 2 presenta las órdenes del modelo ARX obtenidas a partir del subconjunto de datos estimación-validación porcentual, así como los resultados de los índices de eficiencia necesarios para evaluar los resultados del modelo ARX+DKF, considerando un paso de tiempo hacia adelante $L=1$. El Cuadro 2, en un análisis posterior, servirá de base para decidir cual es el mejor modelo.

Las gráficas de los caudales pronosticados versus los observados, evaluados con las órdenes del modelo bajo el criterio AIC, muestra para los primeros dos

the forecasts decrease in the third subset 50%:50%. Like the flow forecasts obtained with the AIC criterion, the first subset evaluated with the MDL 10%:90% criterion show these variations at the end of the period of analysis; the variations decrease for subsets two and three.

The Nash-Sutcliffe efficiency index is not conclusive for evaluating the subsets, relative to the two criteria. However, when considering the analysis of Figures 6 and 7, better results can be seen with the MDL criterion model orders. Moreover, these orders are lower than those of the AIC criterion. The orders of the model that give better results in comparing observed and forecasted flows (Figure 6 and Figure 7) are those of the subset 25%:75% under the Rissanen criterion (MDL) and the $RMSE$ value is lower. In this sense and considering the six steps forward in the flow forecast, the period in which the parameters were recalculated was 338 h and $na=2$ and $nb=1$ model orders

Figure 8 shows the best forecast model (DKF-ARX), considering the previous analysis based on Table 2 and Figures 6 and 7. The general shapes of the

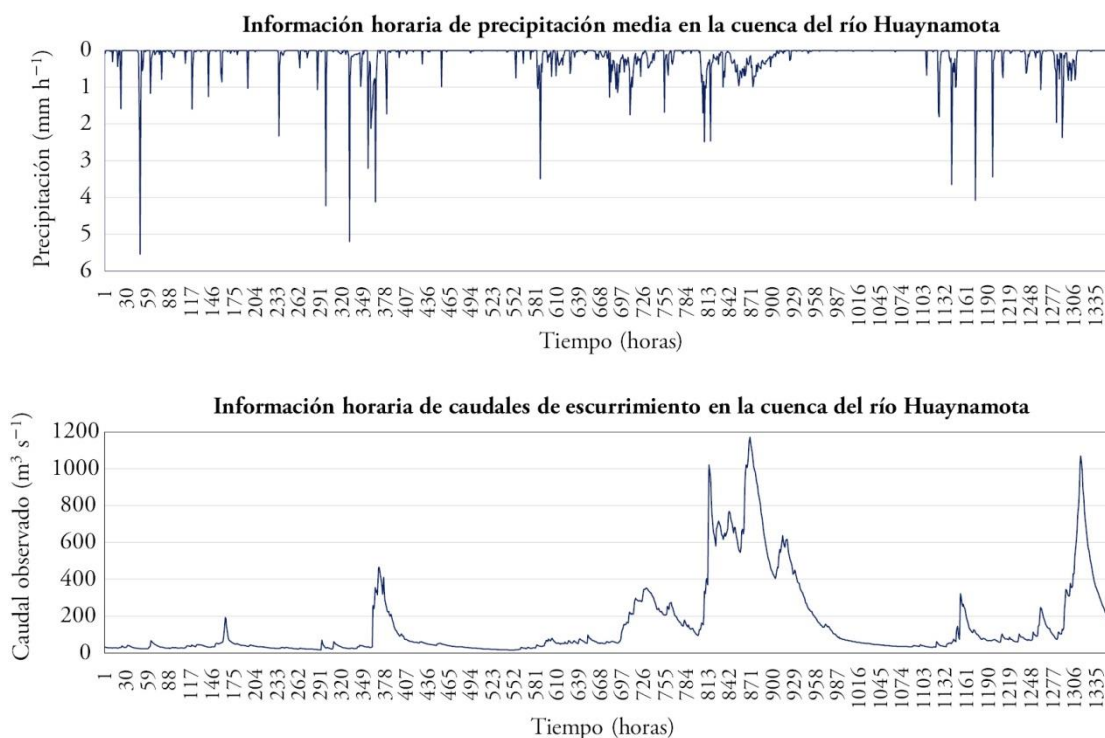


Figura 5. Datos de precipitación y caudales de escurrimiento introducidos al modelo ARX-DKF.
Figure 5. Rainfall and streamflow data introduced into the ARX-DKF model.

subconjuntos 10%:90% y 25%:75% grandes variaciones en los valores pronosticados al final del periodo de análisis; estas diferencias en los pronósticos decrecen en el tercer subconjunto 50%:50%. Al igual que los pronósticos de los caudales obtenidos con el criterio AIC, el primer subconjunto evaluado con el criterio MDL 10%:90% muestra estas variaciones al final del periodo de análisis, las cuales se reducen para los subconjuntos dos y tres.

forecasted hydrograph and the observed hydrograph are similar, with underestimated flow forecasts, out of phase L steps forward.

The maximum flow observed in the analyzed period was $1171 \text{ m}^3 \text{ s}^{-1}$, which occurred September 7, 2017, at 13:00 h. A second maximum flow of $1070 \text{ m}^3 \text{ s}^{-1}$ occurred September 26 at 02:00 h. The mean and standard deviation of the observed flows are 190.27 and $231.78 \text{ m}^3 \text{ s}^{-1}$; for flow forecasted one

Cuadro 2. Órdenes del modelo ARX para un paso hacia adelante.
Table 2. ARX model orders for one step ahead.

Subconjunto	Horas para estimación de parámetros	na	nb	Nash-Sutcliffe	RMSE
Considerando criterio AIC					
10% : 90%	136	2	1	0.98	27.52
25% : 75%	340	3	2	0.98	29.12
50% : 50%	680	3	2	0.98	32.94
Considerando Criterio MDL					
10% : 90%	136	1	1	0.98	28.55
25% : 75%	340	2	1	0.98	28.29
50% : 50%	680	1	2	0.98	32.99

El índice de eficiencia Nash-Sutcliffe no es concluyente para evaluar los subconjuntos respecto a los dos criterios, sin embargo, al considerar el análisis de las Figuras 6 y 7 se aprecian mejores resultados con las órdenes del modelo del criterio MDL, además que estas órdenes resultan menores que las del criterio AIC. Las órdenes del modelo que dan mejores resultados al comparar caudales observados y pronosticados (Figura 6 y Figura 7) son los del subconjunto

step forward $L=1$ they are 187.16 and $227.75 \text{ m}^3 \text{ s}^{-1}$. The Nash-Sutcliffe coefficient for one step forward is 0.99 and RMSE is 28.37 . The forecast of the first peak flow occurred on September 7, 2017, at 14:00 h with a value of $1135.84 \text{ m}^3 \text{ s}^{-1}$; the second peak was on September 26 at 3:00 h with a value of $1087.61 \text{ m}^3 \text{ s}^{-1}$. The percent difference is less than 3.1 in both cases, and the forecasts are out of phase one step forward in time (Figure 8), relative to observed

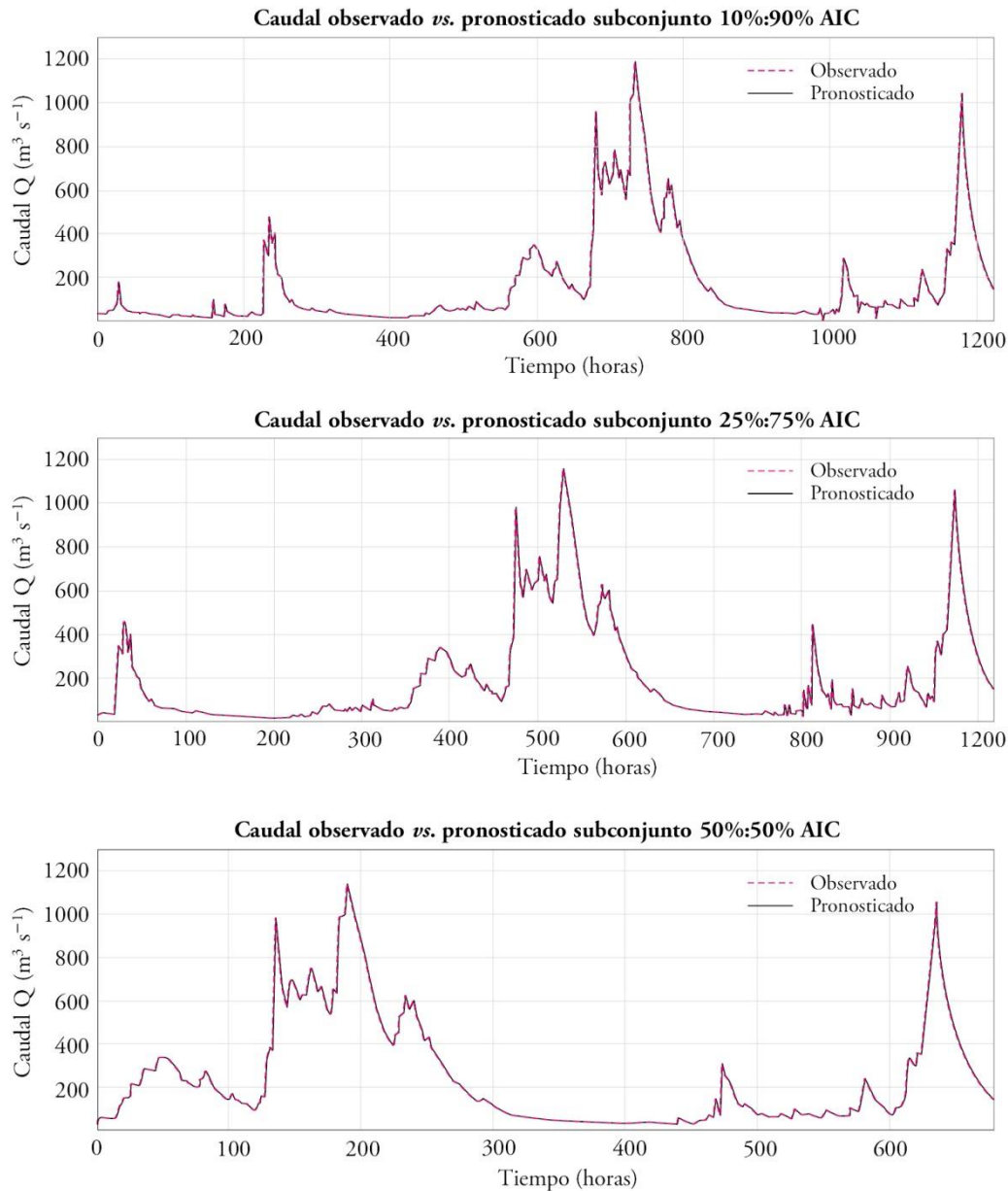


Figura 6. Caudales pronosticados vs. observados para órdenes de modelo bajo el criterio AIC.
Figure 6. Forecasted vs. observed flows for model orders under the AIC criterion.

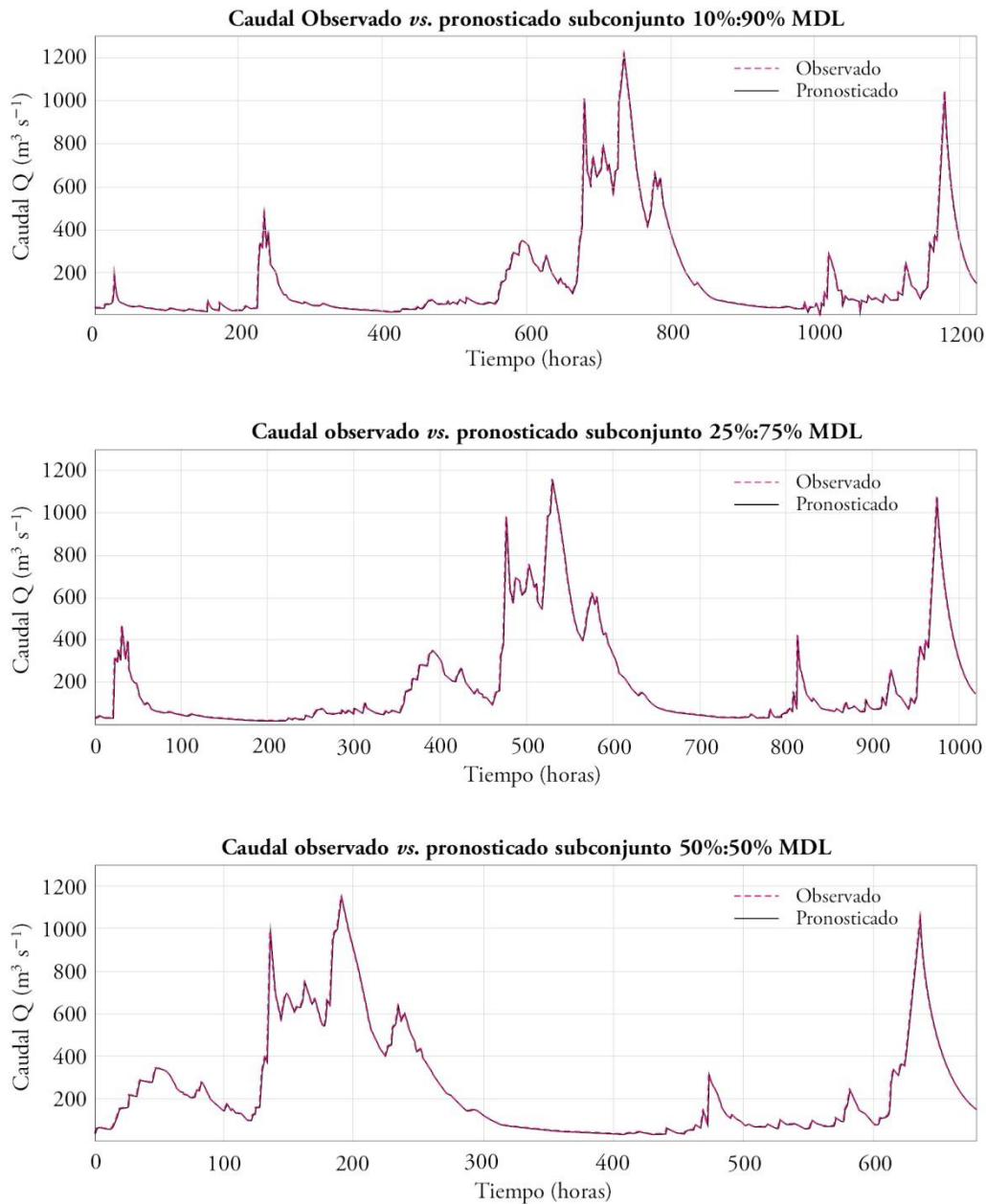


Figura 7. Caudales pronosticados vs. observados para órdenes de modelo bajo el criterio MDL.

Figure 7. Forecasted vs. observed flows for model orders under the MDL criterion.

2%:75%, bajo el criterio de Rissanen (MDL) y se observa que su valor de *RMSE* es menor. En este sentido y al considerar los seis pasos hacia adelante en el pronóstico de caudales, el periodo en el cual se recalcularon los parámetros fue 338 h y órdenes del modelo $na=2$ y $nb=1$.

La Figura 8 muestra el mejor modelo de pronóstico (DKF-ARX) al considerar el análisis anterior basado en el Cuadro 2 y las Figuras 6 y 7. La forma

data. Morales-Velázquez *et al.* (2014) determined a Nash-Sutcliffe coefficient of 0.98 for the short-term forecast of flows with DFK, considering the unit hydrograph as a response function of the basin. For this reason, the model implemented in our study notably improved this result.

As the flow forecasts increased *L* steps forward, the Nash-Sutcliffe coefficients decrease, while *RMSE* increase (Table 3). This suggests that the accuracy of

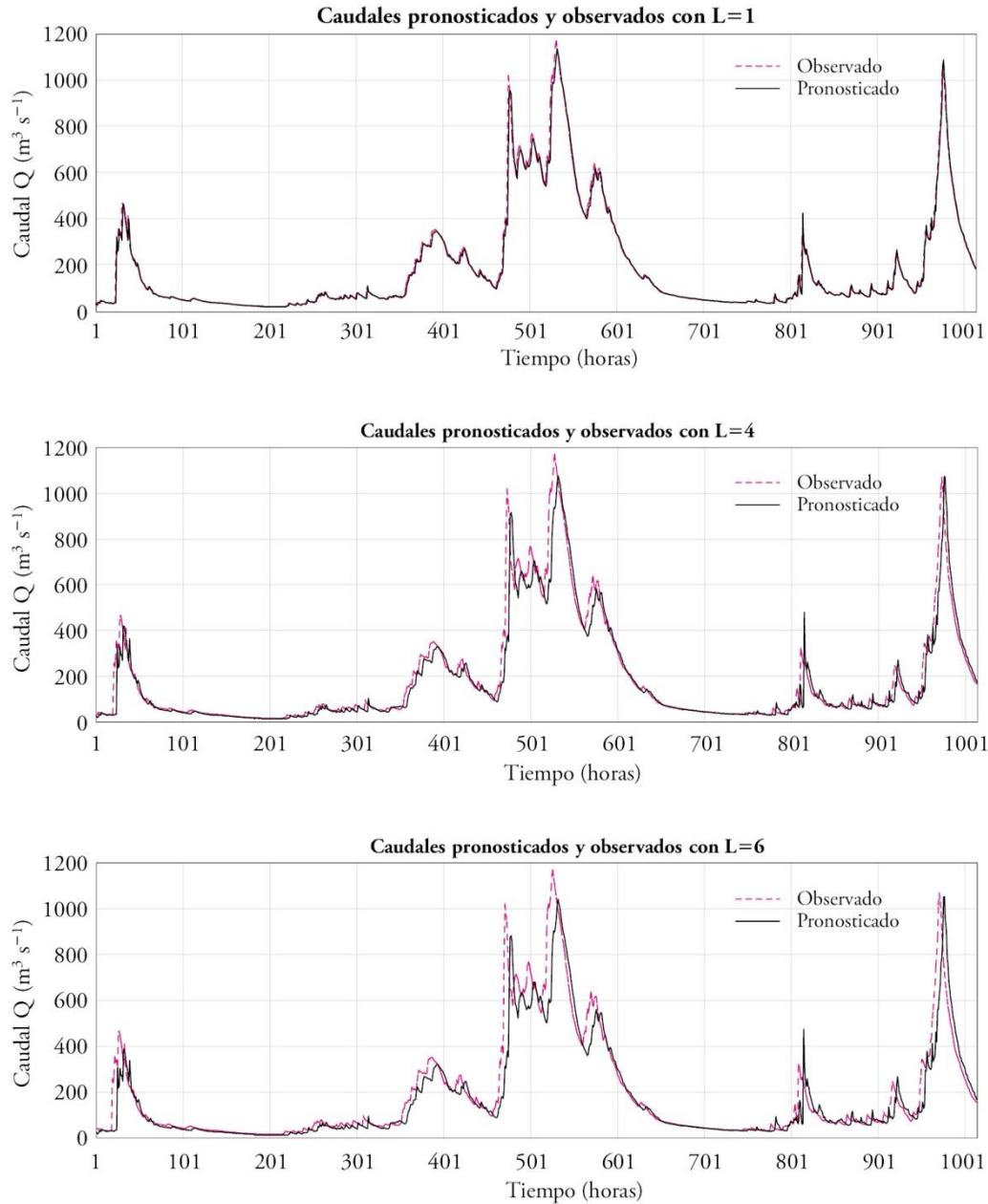


Figura 8. Caudales pronosticados vs. observados con el mejor modelo ARX-DXF para L=1, 4 y 6.
Figure 8. Forecasted vs. observed flows with the best ARX-DXF model for L=1, 4 and 6.

general del hidrograma pronosticado tiene similitud respecto al hidrograma observado, con caudales pronosticados subestimados y desfasados L pasos hacia adelante.

El caudal máximo observado en el periodo analizado fue $1171 \text{ m}^3 \text{ s}^{-1}$ y ocurrió el 7 de septiembre del 2017 a las 13:00; un segundo valor máximo de $1070 \text{ m}^3 \text{ s}^{-1}$ se presentó el 26 de septiembre a las

the forecast is being lost and the covariance of error of the forecasted flows increases at each step of time. However, according to the adjustment criteria, they are acceptable and the Nash-Sutcliffe coefficient for six steps forward $L=6$ is 0.85. What was considered as a possibility at the end of this study was to group data every 12 h and make forecasts at 12, 24 and 36 h, considering the 44-h concentration time. But we did

02:00. La media y desviación estándar de las caudales observados son 190.27 y $231.78 \text{ m}^3 \text{ s}^{-1}$, para los caudales pronosticados un paso hacia adelante $L=1$ son 187.16 y $227.75 \text{ m}^3 \text{ s}^{-1}$, el valor del coeficiente de Nash-Sutcliffe para un paso hacia adelante es 0.99 , RMSE es 28.37 . El pronóstico del primer caudal pico se presentó el 7 de septiembre del 2017 a las 14:00 con un valor de $1135.84 \text{ m}^3 \text{ s}^{-1}$, para el segundo pico se presenta el 26 de septiembre a las 3:00 con un valor de $1087.61 \text{ m}^3 \text{ s}^{-1}$, la diferencia porcentual es menor de 3.1 en ambos casos, y los pronósticos se desfasan un paso de tiempo hacia adelante (Figura 8) respecto a los datos observados. Morales-Velázquez *et al.* (2014) determinaron un coeficiente de Nash-Sutcliffe de 0.98 para el pronóstico de caudales a corto plazo con DFK considerando el hidrograma unitario como función de respuesta de la cuenca, por lo que el modelo implementado en nuestro estudio mejoró sensiblemente este resultado.

Conforme se incrementa el pronóstico de caudales L pasos hacia adelante, el coeficiente de Nash-Sutcliffe decrece, mientras que en RMSE se aumentan (Cuadro 3); esto sugiere que la precisión en el pronóstico se va perdiendo y la covarianza de error de los caudales pronosticados aumenta a cada paso de tiempo. Sin embargo, de acuerdo con los criterios de ajuste, estos son aceptables y el coeficiente de Nash-Sutcliffe para seis pasos hacia adelante $L=6$ es 0.85 . Lo que se consideró como una posibilidad al final de este estudio fue agrupar los datos cada 12 h y realizar pronósticos a 12 , 24 y 36 h , al considerar que el tiempo de concentración es 44 h . Pero no se hizo porque al dejar el pronóstico cada hora y pronosticar 12 h hacia adelante, la confiabilidad del pronóstico va decayendo. O bien, como lo realizaron González-Leiva *et al.* (2015), pronosticar caudal medio cada 24 h , y considerar como variable exógena la lluvia acumulada las 24 h anteriores.

La media y desviación estándar de las caudales observados son 190.93 y $231.51 \text{ m}^3 \text{ s}^{-1}$, para los caudales pronosticados seis pasos hacia adelante $L=6$ son 173.00 y $210.60 \text{ m}^3 \text{ s}^{-1}$, respectivamente (Figura 8). El pronóstico del primer caudal pico se presenta el 07 de septiembre a las 19:00 con un valor de $1042.81 \text{ m}^3 \text{ s}^{-1}$ que subestima el valor máximo observado con una diferencia porcentual de 10.95 , para el segundo pico se presenta el 26 de septiembre a las 8:00 con un valor de $1053.59 \text{ m}^3 \text{ s}^{-1}$ que subestima el caudal con una diferencia porcentual de 1.53 , y el pronóstico se

not because, by leaving the forecast every hour and forecasting 12 h forward, reliability of the forecast declines. Or rather, like González-Leiva *et al.* (2015), half flow can be forecasted every 24 h , considering the rainfall accumulated in the previous 24 h as an exogenous variable.

The mean and standard deviation of the observed flows are 190.93 and $231.51 \text{ m}^3 \text{ s}^{-1}$; for six steps forward $L=6$, are 173.00 and $210.60 \text{ m}^3 \text{ s}^{-1}$, respectively (Figure 8). The forecast of the first peak flow on September 7 at 19:00 h with a value of $1042.81 \text{ m}^3 \text{ s}^{-1}$ underestimates the maximum observed value with a difference of 10.95% . For the second peak on September 26 at 8:00 with a value of $1053.59 \text{ m}^3 \text{ s}^{-1}$, the flow is underestimated by 1.53% , and the forecast is 6 h out of phase ahead of the observed peak flow.

González-Leiva *et al.* (2015), to forecast flows for the ARX+DFK model with average daily (24 h) information for a model order of $na=1$, $nb=2$, and at $L=1, 2, 3$ and 4 steps forward, obtained Nash-Sutcliffe coefficients of $0.92, 0.82, 0.72$ and 0.63 , respectively. In our study, the results improved considerably because when the information is hourly, the response function of the basin is better represented.

The dispersion graph of Figure 9, for one step forward $L=1$, shows that at least 4 values of forecasted flows are underestimated because they move away from the 45° line. This error occurs where the flow increases significantly from one step in time to another, as shown in Figure 10. In the other dispersion graphs the values disperse more because, as we commented above, the forecasted hydrograph becomes out of phase with the observed hydrograph causing large residuals.

Simple linear regression was carried out with a straight-line model of the $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ type,

Cuadro 3. Estadísticas de pronóstico de caudales del modelo ARX-DXF.

Table 3. Statistics of the ARX-DXF model flow forecast.

Pasos hacia adelante L	Nash-Sutcliffe	RMSE
$L=1$	0.99	28.37
$L=2$	0.96	44.72
$L=4$	0.91	69.00
$L=6$	0.85	89.11

desfaza 6 h hacia adelante respecto al caudal pico observado.

González-Leiva *et al.* (2015) para el pronóstico de caudales para un modelo ARX+DFK con información promedio diaria (24 h), obtuvieron para una orden de modelo $na=1$, $nb=2$, y a $L=1, 2, 3$ y 4 pasos hacia adelante, valores del coeficiente de Nash-Sutcliffe de 0.92, 0.82, 0.72 y 0.63, respectivamente. En el presente estudio se mejoraron considerablemente los resultados, porque al ser información horaria la función de respuesta de la cuenca se representa de mejor manera.

La gráfica de dispersión de la Figura 9, para un paso hacia adelante $L=1$, muestra por lo menos 4 valores de caudales pronosticados son subestimados porque se alejan de la línea a 45° ; este error se presenta donde el caudal aumente de manera significativa de un paso de tiempo a otro, como se muestra en la Figura 10. En las otras gráficas de dispersión los valores se van dispersando más, porque como ya se comentó, el hidrograma pronosticado se desfaza del observado provocando residuales grandes.

and the values of parameters β_0 and β_1 are presented in Table 4.

Analysis of the b_1 regression coefficient is important in forecasting flows one step ahead $L=1$. The value $\beta_1=0.9754$ indicates that the forecast slightly underestimates the flows. As the L steps forward increase for flow forecasting, the regression coefficient decreases, and for six steps forward $L=6$, it has the value of $\beta_1=0.8426$, as shown in the dispersion graphs of Figure 9.

The graphs of residuals in Figure 10 show increasing variability in the measure that the flow increases, especially when the forecast L steps ahead is greater. In this case, for the forecast $L=3, 4, 5$ and 6 h ahead, the residues tend to underestimate the flows. This can be explained because in the hydrograph of the forecasted flow is out of phase and, when compared with real values, the variations are greater at the peaks.

Figure 11 shows the high variability in the entire range of flows but, according to the margins with a confidence interval of 95%, it can be better seen as of $760 \text{ m}^3 \text{ s}^{-1}$.

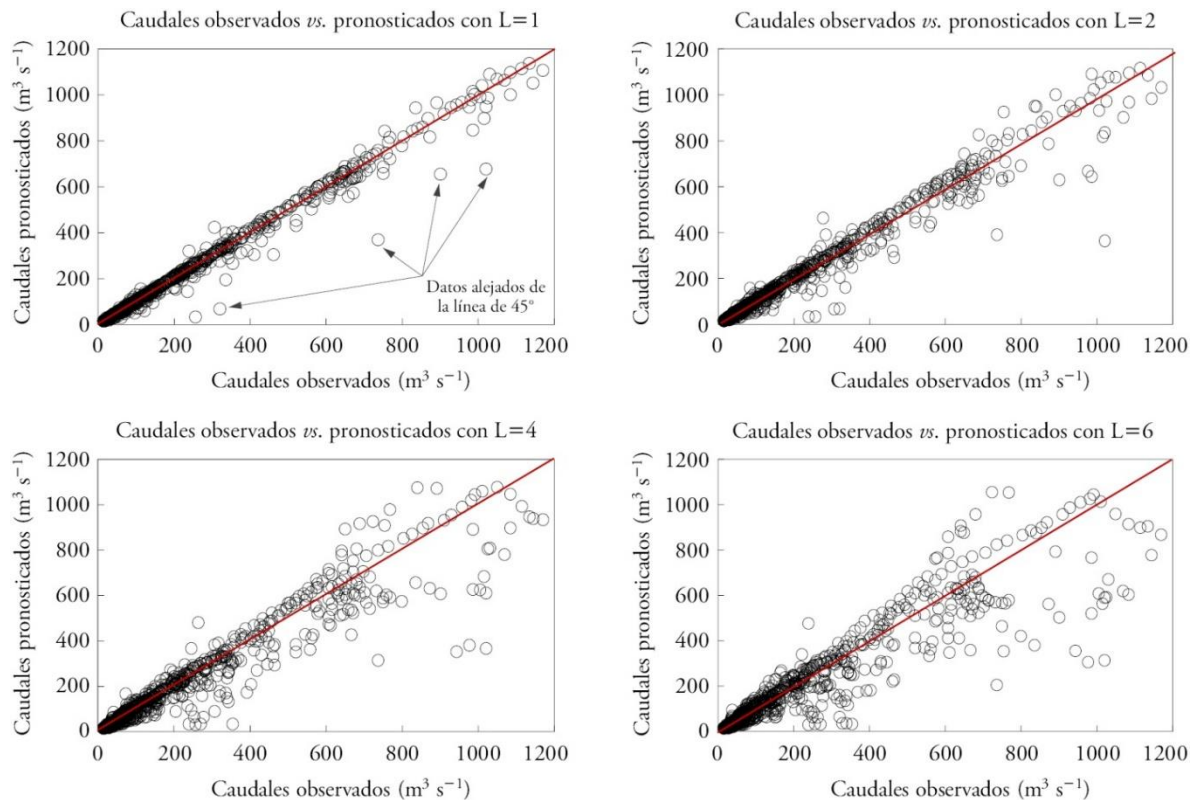


Figura 9. Diagramas de dispersión para caudales pronosticados *vs.* observados.

Figure 9. Dispersion diagrams for forecasted *vs.* observed flows.

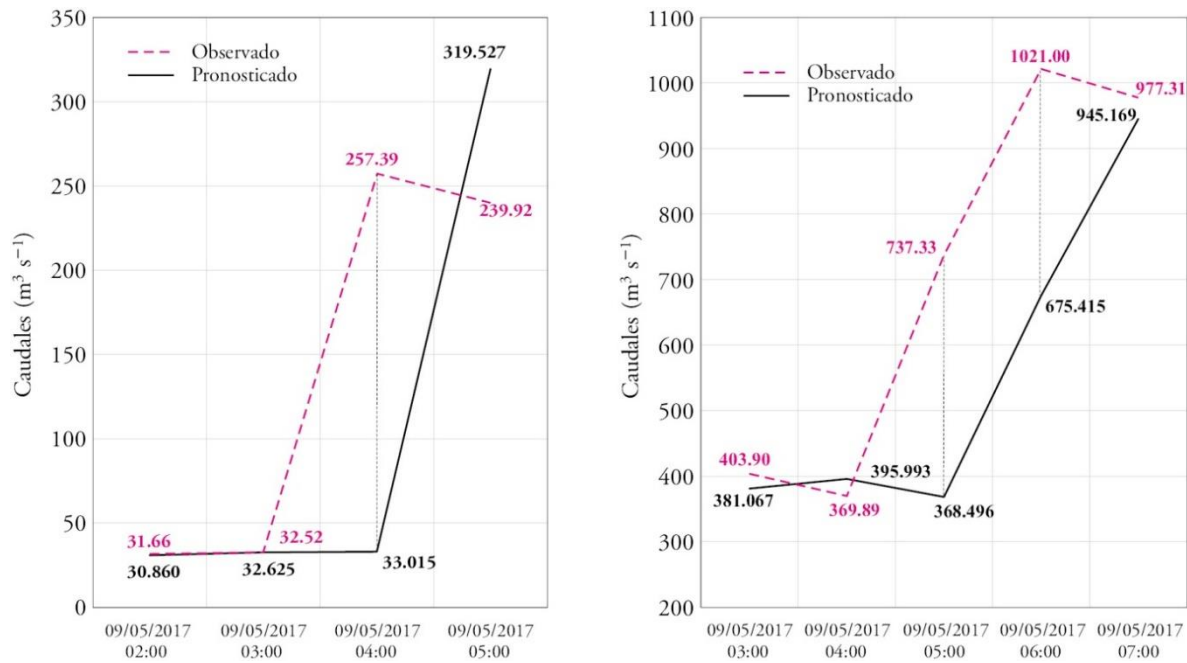


Figura 10. Datos alejados de la línea a 45° para L=1.
Figure 10. Data removed from the 45° line for L=1.

La regresión lineal simple se realizó con un modelo de línea recta del tipo $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$, y los valores de los parámetros β_0 y β_1 se presentan en el Cuadro 4.

El análisis del coeficiente de regresión β_1 es importante para el pronóstico de caudales un paso hacia adelante L=1 el valor $\beta_1=0.9754$, lo cual indica que el pronóstico subestima ligeramente los caudales, y conforme aumentan los L pasos hacia adelante para el pronóstico de caudales el coeficiente de regresión disminuye y para seis pasos hacia adelante es L=6 el valor de $\beta_1=0.8426$, como se muestra en las gráficas de dispersión de la Figura 9.

Las gráficas de residuales en la Figura 10 muestran una variabilidad creciente a medida que los caudales aumentan, sobre todo cuando el pronóstico L pasos hacia adelante es mayor. En este caso para el pronóstico L=3, 4, 5, y 6 h hacia adelante los residuos muestran una tendencia a subestimar los caudales. Esto se explica debido al desfase del hidrograma de caudal pronosticado, que al compararse con los valores reales las variaciones son mayores en los picos.

La Figura 11 muestra la alta variabilidad en todo el rango de caudales, pero de acuerdo con las franjas con un intervalo de confianza de 95%, se aprecia más desde 760 m³ s⁻¹.

CONCLUSIONS

The correct identification of model orders and the period for recalculating the values of the a and b parameters of the ARX model is relevant for obtaining good quality flow forecasts.

The Discrete Kalman Filter (DKF) model, together with an autoregressive model with exogenous input (ARX) performed well in forecasting hourly flows, as indicated by their statistical efficiency criteria (Nash-Sutcliffe and RSME) when comparing observed and forecasted flows. The model, even in the most critical condition for forecasting flows six steps ahead (L=6), produces a value of $E=0.85$; that is, the flow forecast considering hourly data gives satisfactory

Cuadro 4. Parámetros de la línea de regresión.
Table 4. Parameters of the regression line.

Pasos hacia adelante L	β_0	β_1
L=1	1.5837	0.9754
L=2	3.3356	0.9506
L=4	7.4034	0.8977
L=6	12.1199	0.8426

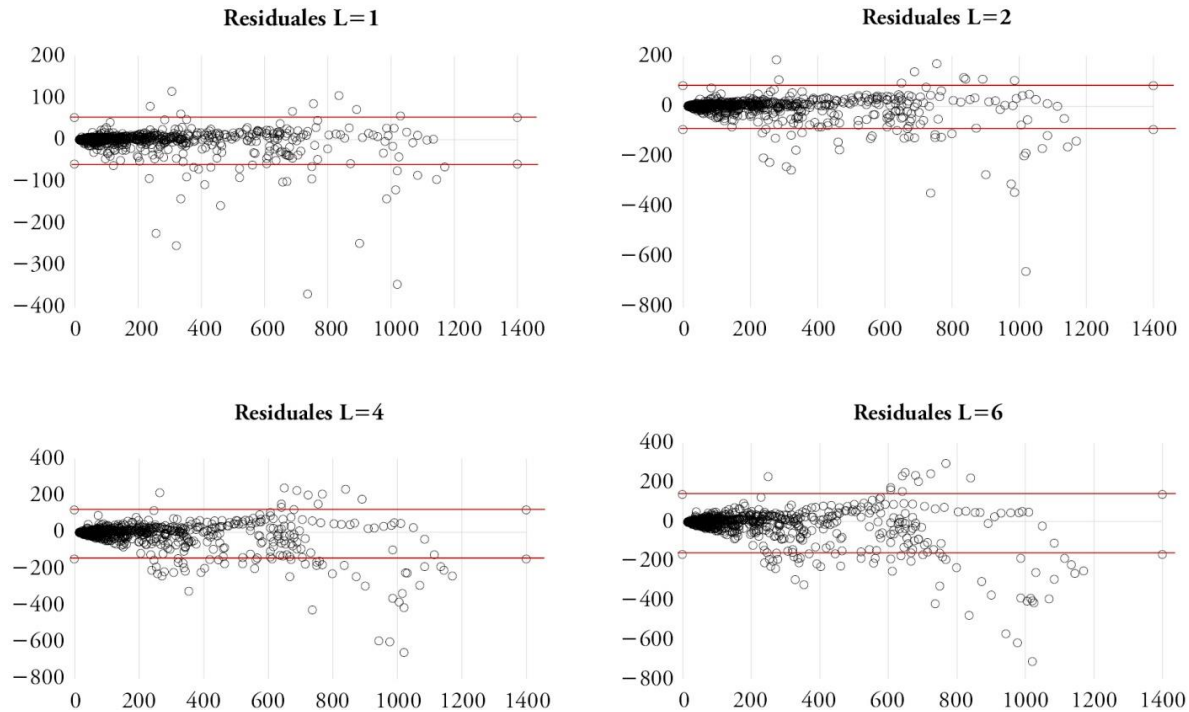


Figura 11. Graficas de residuales con franjas para un nivel de confianza de 95%.
Figure 11. Graphs of residuals with margins for a confidence level of 95%.

CONCLUSIONES

La identificación correcta de las órdenes del modelo y el periodo para recalcular los valores de los parámetros a y b del modelo ARX, son relevantes para obtener una buena calidad de caudales pronosticados.

El modelo de filtro de Kalman Discreto (DKF) junto con un modelo autorregresivo con entrada exógena (ARX) fueron buenos en el pronóstico de caudales horarios, porque sus criterios de eficiencia estadística (Nash-Sutcliffe y RSME) lo indican al momento de comparar los caudales observados contra los pronosticados. El modelo, aún en la condición más crítica para el pronóstico de caudales seis pasos hacia adelante ($L=6$) da un valor de $E=0.85$, es decir el pronóstico de caudales considerando datos horarios da resultados satisfactorios que pudieran emplearse para sistema de alerta temprana. A pesar del desfase de los caudales picos pronosticados para los diferentes pasos hacia adelante, no se pierde la utilidad práctica de estos valores, al considerar que el tiempo de concentración de la cuenca Huaynamota es 44 h.

results that can be used for an early warning system. Despite the lack of coincidence of the forecasted peak flows for different steps ahead, there is no loss of practical usefulness of these values, considering that concentration time in the Huaynamota basin is 44 h.

Because of the lack of hydroclimatological information in the country to implement this type of models for forecasting flows, planning for instrumentation of the basins is necessary, especially to improve the forecasting model that is fed hourly information.

—End of the English version—

Debido a la falta de información hidroclimática en el país, para implementar este tipo de modelos para pronóstico de caudales es necesaria una planeación para la instrumentación de las cuencas, en especial para mejorar el pronóstico el modelo que se alimenta con información horaria.

AGRADECIMIENTOS

A la Comisión Federal de Electricidad por permitir acceso a su base de datos hidrometeorológica, a través del sitio web administrado por el Instituto Nacional de Electricidad y Energías Limpias.

LITERATURA CITADA

- Beven, K. 1989. Changing ideas in hydrology-the case of physically-based models. *J. Hydrol.* 105: 157-172.
- Castañeda C., J. A., M. A. Nieto A., and V. A. Ortiz B. 2013. Analysis and application of the Kalman filter to a signal with random noise. *Scientia et Technica* 18: 267-274.
- Chong-yu Xu. 2002. *Hydrologic Models*. Uppsala University. Suecia. 168 p.
- Comisión Federal de Electricidad. 2017. Sistema de Monitoreo de Cuencas de la CFE. <https://h06814.ije.org.mx/cuencas/logon.aspx?ReturnUrl=%2fcuencas%2fdefault.aspx>. (Consulta: septiembre 2017).
- Devi K., G., B. P. Ganasri, and G. S. Dwarakish. 2015. A review on hydrological models. *Aquatic Procedia* 4: 1001-1007.
- González-Leiva, F., L. A. Ibáñez-Castillo, J. B. Valdés, M. A. Vázquez-Peña, y A. Ruiz-García. 2015. Pronóstico de caudales con Filtro de Kalman Discreto en el río Turbio. *Tecnol. Ciencias del Agua* 6: 5-24.
- Hernández P., Y. y H. Medina G. 2012. Estimación de la humedad del suelo mediante técnicas de asimilación de datos. *Rev. Ciencias Téc. Agropec.* 21: 30-35.
- Hsu, K.-L., H. Moradkhani, and S. Sorooshian. 2009. A sequential Bayesian approach for hydrologic model selection and prediction, *Water Resour. Res.* 45: W00B12.
- Hobson, K. A., and K. J. Kardynal. 2015. Western Veeries use an eastern shortest-distance pathway: New insights to migration routes and phenology using light-level geolocators. *The Auk* 132: 540-550.
- INEGI. 2018. Red hidrográfica 1:50,000, edición 2.0. http://www.inegi.org.mx/geo/contenidos/topografia/regiones_hidrograficas.aspx. (Consulta: enero 2018).
- Krause, P., D. P. Boyle, and F. Base. 2005. Comparison of different efficiency criteria for hydrological model assessment. *Adv. Geosci.* 5: 89-97.
- Ljung, L. 2017. *System Identification Toolbox. Getting Started Guide*. The MathWorks INC. 226 p.
- Mendoza, M., G. Bocco, M. Bravo, C. Siebe, y M. A. Ortiz. 2002. Modelamiento hidrológico espacialmente distribuido: una revisión de sus componentes, niveles de integración e implicaciones en la estimación de procesos hidrológicos en cuencas no instrumentadas. *Invest. Geog.* 47: 36-58.
- Morales-Velázquez. M. I., J. Aparicio, y J. B. Valdés. 2014. Pronóstico de avenidas utilizando el Filtro de Kalman Discreto. *Tecnol. Cienc. Agua* 5: 85-110.
- Padilla B., J. I., L. D. Avendaño V., y G. Castellanos D. 2013. Estimación mejorada de modelos AR multivariados en el análisis de señales EEG. *Rev. Ing.* 38: 20-26.
- Servicio Meteorológico Nacional. 2017. Datos Históricos de Estaciones Meteorológicas Automáticas operadas por el SMN. (Consulta: septiembre 2017).
- Solera R., A. 2003. El Filtro de Kalman. Documento de trabajo del Banco central de Costa Rica, elaborado en la División Económica, Departamento de Investigaciones Económicas. 33 p.
- Vargas-Castañeda G., L. A. Ibáñez-Castillo, and R. Arteaga-Ramírez. 2015. Development, classification and trends in rainfall-runoff modeling. *INAGBI* 7: 5-21.
- Vargas-Castañeda. G., L. A. Ibáñez-Castillo, R. Arteaga-Ramírez, and G. Arévalo-Galarza. 2018. Kinematic wave hydrologic model of the Turbio River basin, Guanajuato, Mexico. *INAGBI* 10: 33-47.
- Valdés, J. B., J. Mejía V., e I. Rodríguez I. 1980. Filtros de Kalman en la hidrología: predicción de descargas fluviales para la operación óptima de embalses. Informe Técnico No. 80-2. Caracas: Universidad Simón Bolívar, Decanato de Estudios de Posgrado, Posgrado en Planificación e Ingeniería de Recursos Hídricos. 37 p.
- Zhen-lei, W., X. Yue-Ping, S. Hong-yue, W. Xie, and W. Gang. 2018. Predicting the occurrence of channelized debris flow by an integrated cascading model: A case study of a small debris flow-prone catchment in Zhejiang Province, China, *Geomorphology* doi:10.1016/j.geomorph.2018.01.027
- Welch, G., and G. Bishop. 2006. *An Introduction to the Kalman Filter*. Chapel Hill: University of North Carolina at Chapel Hill. 16 p.