



UNIVERSIDAD AUTÓNOMA CHAPINGO

POSGRADO EN INGENIERÍA AGRÍCOLA Y USO INTEGRAL DEL AGUA

PRONÓSTICO DE CAUDALES HORARIOS CON FILTRO DE KALMAN DISCRETO EN LA CUENCA DEL RÍO HUAYNAMOTA

TESIS DE GRADO

Que como requisito parcial
para obtener el grado de:

MAESTRO EN INGENIERÍA AGRÍCOLA Y USO INTEGRAL DEL AGUA

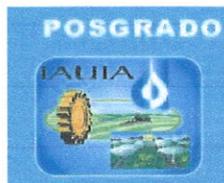
Presenta:

LETICIA ALVARADO HERNANDEZ



DIRECCION GENERAL ACADEMICA
DEPTO. DE SERVICIOS ESCOLARES
SECRETARIA DE EDUCACION PUBLICA

Bajo la supervisión de: LAURA ALICIA IBÁÑEZ CASTILLO, Ph. DOCTORA



Chapingo, Estado de México, diciembre de 2018.

**PRONÓSTICO DE CAUDALES HORARIOS CON FILTRO DE KALMAN
DISCRETO EN LA CUENCA DEL RÍO HUAYNAMOTA**

Tesis realizada por **Leticia Alvarado Hernández** bajo la supervisión del Comité Asesor indicado, aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

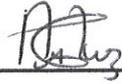
MAESTRO EN INGENIERÍA AGRÍCOLA Y USO INTEGRAL DEL AGUA

DIRECTORA:



Ph.D. LAURA ALICIA IBÁÑEZ CASTILLO

ASESOR:



DR. AGUSTÍN RUIZ GARCÍA

ASESOR:



DR. MARIO ALBERTO VÁZQUEZ PEÑA

CONTENIDO

	Página
CONTENIDO	iii
ÍNDICE DE CUADROS	v
ÍNDICE DE FIGURAS	vi
DEDICATORIA	viii
AGRADECIMIENTOS	ix
DATOS BIOGRÁFICOS	x
RESUMEN GENERAL	xii
GENERAL ABSTRACT	xiii
CAPÍTULO 1. INTRODUCCIÓN GENERAL	1
1.1. Hipótesis	4
1.2. Objetivos	5
1.2.1. General	5
1.2.2. Específicos.....	5
1.3. Organización de la tesis.....	5
CAPÍTULO 2. REVISIÓN DE LITERATURA	6
2.1. Ciclo hidrológico.....	6
2.1.1. Sistema hidrológico.....	7
2.2. Modelación para procesos hidrológicos	11
2.2.1. Clasificación de modelos hidrológicos	11
2.2.2. Modelos lluvia-escorrentía.....	14
2.2.3. Proceso de modelado	20
2.3. Modelo Autorregresivo con entrada exógena (ARX).....	22
2.3.1. Estimación del orden del modelo	23
2.4. Filtro de Kalman discreto (DKF).....	24
2.5. Criterios de eficiencia para modelos hidrológicos	27

2.5.1. Criterio de eficiencia de Nash-Sutcliffe (E).....	27
2.5.2. Raíz del error cuadrático medio (RMSE)	27
2.6. Zona de estudio	28
2.6.2. Cuenca del río Santiago.....	28
2.6.2. Cuenca Chapalagana	29
2.6. Referencias.....	31
CAPÍTULO 3. PRONÓSTICO HORARIO DE CAUDALES MEDIANTE	
FILTRO DE KALMAN DISCRETO EN EL RÍO HUAYNAMOTA	34
3.1. Resumen.....	34
3.2. Introducción	35
3.3. Materiales y métodos.....	37
3.3.1. Cuenca Huaynamota	37
3.3.2. Información climatológica.....	38
3.3.3. Metodología	39
3.3.3.1. Modelo Autorregresivo con entrada exógena (ARX).....	39
3.3.3.2. Filtro de Kalman Discreto (DKF)	40
3.3.4. Modelo ARX+DKF.....	42
3.3.5. Análisis estadístico	44
3.4. Resultados y discusión	44
3.5. Conclusiones	55
3.6. Agradecimientos	56
3.7. Literatura citada	56
CAPÍTULO 4. CONCLUSIONES Y RECOMENDACIONES	59
CAPÍTULO 5. ANEXOS	60
5.1. Cálculo de la precipitación media de la cuenca	60
5.2. Análisis estadístico de los resultados	62
5.3. Forma matricial del filtro de Kalman discreto	68

ÍNDICE DE CUADROS

	Página
CAPÍTULO 2	
Cuadro 1. Ejemplo de análisis hidrológico y síntesis	10
CAPÍTULO 3	
Cuadro 1. Estaciones de registro horario de lluvia y caudal.	39
Cuadro 2. Órdenes del modelo ARX.....	46
Cuadro 3. Estadísticas de pronóstico de caudales del modelo ARX-DXF.	50
CAPÍTULO 5	
Cuadro 1. Tabla de ponderación de precipitación.....	61
Cuadro 2. Estadísticas de la regresión simple para L=1.....	62
Cuadro 3. Tabla ANOVA para L=1.	63
Cuadro 4. Estadísticas de la regresión simple para L=2.....	63
Cuadro 5. Tabla ANOVA para L=2.	63
Cuadro 6. Estadísticas de la regresión simple para L=3.....	64
Cuadro 7. Tabla ANOVA para L=3.	64
Cuadro 8. Estadísticas de la regresión simple para L=4.....	65
Cuadro 9. Tabla ANOVA para L=4.	65
Cuadro 10. Estadísticas de la regresión simple para L=5.....	66
Cuadro 11. Tabla ANOVA para L=5.	66
Cuadro 12. Estadísticas de la regresión simple para L=6.....	67
Cuadro 13. Tabla ANOVA para L=6.	67

ÍNDICE DE FIGURAS

	Página
CAPÍTULO 2	
Figura 1. Esquema de las componentes del ciclo hidrológico (UNESCO, 2006).....	7
Figura 2. Diagrama de bloques del sistema hidrológico global (Chow et al., 1994).....	8
Figura 3. Representación conceptual del proceso del sistema (McCuen, 2005).....	9
Figura 4. Clasificación de modelos hidrológicos (Chong-yu, 2002).....	12
Figura 5. Modelos matemáticos en hidrología (UNESCO, 1985).....	15
Figura 6. Proceso lluvia-escurrimiento (Aparicio, 2013).....	17
Figura 7. Esquema de los pasos en el proceso de modelado (Beven, 2012).....	20
Figura 8. Esquema general del Filtro de Kalman.....	24
Figura 9. Algoritmo del Filtro de Kalman discreto (Welch y Bishop, 2006).....	26
Figura 10. Cuenca del río Santiago.....	29
Figura 11. Cuenca Chapalagana.....	30
CAPÍTULO 3	
Figura 1. Ubicación de la cuenca del río Huaynamota.....	38
Figura 2. Etapas del modelo DKF.....	41
Figura 3. Algoritmo del filtro de Kalman Discreto (Welch y Bishop, 2006).....	42
Figura 4. Proceso de estimación de parámetros y pronóstico de caudales.....	43
Figura 5. Pronóstico de caudales L pasos hacia adelante.....	44

Figura 6. Datos de precipitación y caudales de escurrimiento introducidos al modelo ARX-DKF.....	45
Figura 7. Caudales pronosticados vs observados para órdenes de modelo bajo el criterio AIC.	47
Figura 8. Caudales pronosticados vs observados para órdenes de modelo bajo el criterio MDL.....	48
Figura 9. Caudales pronosticados vs observados con el modelo ARX-DXF para L=1, 4 y 6 horas hacia adelante.....	51
Figura 10. Diagramas de dispersión para caudales pronosticados vs observados.....	52
Figura 11. Datos alejados de la línea a 45° para L=1.	53
Figura 12. Graficas de residuales con franjas para un nivel de confianza de 95%.	55

CAPÍTULO 5

Figura 1. Polígonos de Thiessen en la cuenca Chapalagana.	60
--	----

DEDICATORIA

A mi familia, por ser mi origen y mi destino.

A mis compañeras, Clau y Cyn por franquear juntas este proceso.

A mis amigos, por apoyarme en todo momento.

AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el financiamiento otorgado para realizar mis estudios de maestría.

A la Universidad Autónoma Chapingo, por brindarme la oportunidad de realizar mis estudios de maestría.

Al Posgrado en Ingeniería Agrícola y Uso Integral del Agua por el apoyo recibido durante el transcurso de la maestría.

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

A la Dra. Laura Alicia Ibáñez Castillo por permitirme trabajar con ella y por los conocimientos compartidos a lo largo de este proceso.

Al Dr. Agustín Ruiz García por su apoyo en la elaboración de este trabajo de investigación.

Al Dr. Mario Alberto Vázquez Peña por su apoyo y sugerencias para la realización de este trabajo.

A todos los profesores que contribuyeron a mi formación académica.

DATOS BIOGRÁFICOS

Datos personales

Nombre	Leticia Alvarado Hernández
Fecha de nacimiento	27 de agosto de 1981
Lugar de Nacimiento	Ecatepec de Morelos, Estado de México
CURP	AAHL810827MMCLRT08
Profesión	Ingeniero en Irrigación
Cédula profesional	08737500

Los estudios de bachillerato los cursó en la Preparatoria Oficial No. 14, en Tezoyuca, Estado de México. En el año 2000 ingresó a la universidad Autónoma Chapingo para cursar el propedéutico, posteriormente ingresa al departamento de Irrigación de la misma universidad, concluyendo sus estudios en el año 2005. El título de Ingeniero en Irrigación lo obtuvo en el año 2012 mediante seminario de titulación.

Del mes de septiembre de 2005 a enero de 2007, laboró en la empresa Centro de Estrategias y Redes de Mercado S.C., realizando proyectos productivos, planes de negocio y evaluaciones financieras para la gestión de recursos en diversas dependencias como SAGARPA, FONAES, Secretaría de Economía, FIRA, entre otras.

Del mes de febrero de 2007 a julio de 2016, laboró en la empresa Sanx Ingeniería Integral y Desarrollo S.A. de C.V., como ingeniero de proyectos. Estuvo a cargo de coordinar y participar en los trabajos de campo y gabinete para la elaboración de estudios geohidrológicos para la actualización de mediciones piezométricas,

reactivación de redes de monitoreo y actualización de documentos de disponibilidad media anual de varios acuíferos en el país, así como proyectos de control de erosión de cuencas, diseño de sistemas de riego de baja presión, estudios de evaluación social de obras de protección en cauces, proyectos para la construcción de infraestructura de protección de cauces y estudios de delimitación de zona federal.

En el mes de agosto de 2016 ingresa al programa de maestría en Ingeniería Agrícola y Uso Integral del Agua en la universidad Autónoma Chapingo, finalizando sus estudios en julio de 2018.

RESUMEN GENERAL

En este estudio se utilizó el algoritmo del filtro de Kalman discreto (DKF) para el pronóstico de caudales en conjunto con un modelo autorregresivo con entrada exógena (ARX), el código se programó en Matlab®, con ayuda del Toolbox identificación de sistemas se identificaron las órdenes del modelo ARX (n_a y n_b). Se estableció un periodo P para recalcular los parámetros a y b del mismo modelo, estos parámetros los retoma el algoritmo del filtro de Kalman discreto para el pronóstico de los caudales para cada periodo de tiempo, el pronóstico se realizó seis pasos hacia adelante $L=1, 2, 3, 4, 5$ y 6 horas.

Se emplearon series de datos horaria ($\Delta t=1$ hora) de caudales a la salida de la cuenca y de lluvia de seis estaciones meteorológicas, se analizó información para el periodo de agosto a septiembre del año 2017, la cuenca de estudio se encuentra en el centro del país, siendo ésta una porción de la cuenca del río Huaynamota delimitada a partir de la estación meteorológica Chapalagana ubicada en el estado de Nayarit.

Se obtuvieron resultados satisfactorios, los valores del índice de eficiencia de Nash-Sutcliffe para $L=1$ fue de 0.99 y para $L=6$ fue de 0.85.

Palabras clave: Filtro de Kalman discreto, modelos ARX, pronóstico de caudales.

GENERAL ABSTRACT

In this case of study, we used the Discrete Kalman filter algorithm (DKF) and an autoregressive model with exogenous input (ARX) for the forecasting of streamflow. The model was programmed in Matlab®, with the toolbox identification system the ARX model orders were identified (n_a and n_b) and a P period to recalculate the parameters a and b for the same model was established, these parameters are taking by the discrete Kalman filter algorithm to forecasting the flows for each period of time, the streamflow forecasting was done with six lead times $L = 1, 2, 3, 4, 5$ and 6 hours.

Time data series ($\Delta t = 1$ hour) of flows at the exit of the basin and data of rain from six meteorological stations were used, the information for the period from August to September of 2017 was analyzed, the basin of the study is located in the center of the country, it is a portion of the basin of the Huaynamota river delimited from the Chapalagana meteorological station located in the state of Nayarit.

The obtained results were satisfactory for the forecast with six steps forward in terms of Nash-Sutcliffe and root-mean-square error (RMSE).

Keywords: Discrete Kalman filter, ARX models, forecast of flows.

Thesis: Maestría en Ingeniería Agrícola y Uso Integral del Agua, Universidad Autónoma Chapingo
Author: Leticia Alvarado Hernández
Advisor: Dra. Laura Alicia Ibáñez Castillo

CAPÍTULO 1. INTRODUCCIÓN GENERAL

La variabilidad climática es una medida del rango en que los elementos climáticos como temperatura y precipitación, varían respecto a los promedios normales. Los fenómenos que producen esos contrastes son frentes fríos, huracanes y perturbaciones tropicales. En este sentido los eventos meteorológicos extremos incluyen precipitaciones pluviales excesivas así como sequías prolongadas.

El impacto de los fenómenos hidrometeorológicos, en conjunto con características de topografía, uso de suelo y cubierta vegetal muestran una exposición al riesgo de los sistemas humanos, los efectos se acentúan más en las zonas con rápidos procesos de urbanización. De aquí se deriva la necesidad de comprender los procesos que ocurren en el sistema hidrológico y predecir la función de respuesta de la cuenca.

La comprensión del ciclo hidrológico, específicamente el entendimiento de la respuesta de la cuenca en el tiempo y en el espacio, resulta de gran importancia para la planeación y manejo de los recursos hídricos.

La mayoría de los sistemas hidrológicos son extremadamente complejos y no se puede esperar comprender con detalle sus procesos e interrelaciones tanto superficiales como subsuperficiales, por lo que la abstracción se hace necesaria para la comprensión y control de algunos aspectos de su comportamiento, lo cual es objeto de la modelación de cuencas (Chong-yu, 2002). Los modelos utilizados en hidrología generalmente se basan en modelos físicos y conceptuales, este segundo manifiesta el entendimiento de los procesos y se presentan como modelos matemáticos o físicos, desde el punto de vista matemático, los modelos pueden ser empíricos y teóricos, con solución análoga o numérica (Mendoza *et al.*, 2002), respecto a la distribución espacial de la información pueden ser agrupados o distribuidos, y si alguna de sus variables o parámetros se describe

mediante una función de probabilidad se considera estocástico, en caso de que no suceda es determinístico.

Los primeros modelos hidrológicos se generaron a mediados del siglo XIX, principalmente para resolver problemas de drenaje, se formuló el método racional para determinar la descarga pico de la inundación, la fórmula racional se considera como la primera generación de modelos hidrológicos y se limita a pequeñas cuencas (Chong-yu, 2002). La evolución de los modelos para entender el comportamiento de cuencas grandes y variables en el tiempo y en el espacio, paso por el método racional modificado, modelos agregados conceptuales como el hidrograma unitario, modelos agregados de base física, en los años 70 aparecen modelos de caja negra en los cuales el modelo se adapta a datos de entrada y salida de la cuenca, destacan: Modelo de promedio móvil autorregresivos (ARMA), Filtro de Kalman que es un esquema de asimilación de datos secuencial (Jean-Philippe, 2003), sin embargo, debido a la dinámica cambiante de condiciones de la cuenca como uso de suelo, contaminantes y sedimentos, se generaron modelos distribuidos de base física como los indicados por Gayathri *et al.* (2015), este tipo de modelos son más complejos, aunque no necesariamente los más eficaces.

Beven (1989) y Mendoza *et al.* (2002), indican que la simulación de modelos de base física, ya sea agregados o distribuidos presentan inconvenientes, primeramente por las ecuaciones que representan a los procesos que son una simplificación lo cual introduce errores, la no linealidad de los procesos individuales del ciclo hidrológico y por lo tanto los parámetros de calibración no se reproducen correctamente, la calibración y validación depende de los errores de las observaciones de entradas y salidas, se puede caer en una parametrización excesiva si se consideran varios procesos en el modelado, un factor muy importante es la falta de información: datos de uso de suelo, vegetación, edafología que representen las condiciones reales de la cuenca, datos de precipitación, escurrimiento, velocidad de viento, etc., para periodos largos. Lo anterior da lugar a que este tipo de modelos solo se pueda emplear en cuencas

específicas para los cuales sus parámetros fueron calibrados (Semenova y Beven, 2015).

De la información que se requiere para realizar una modelación, las que mayormente se registran en el país por la Comisión Nacional del Agua (CONAGUA), Comisión Federal de Electricidad (CFE) y gobierno de los estados, son precipitación y caudales de escurrimiento registradas en estaciones meteorológicas y estaciones hidrométricas, de esta última es más escasa la información. Mendoza *et al.* (2002) recalcan la importancia de contar con registros de precipitación y caudales de escurrimiento para validar los modelos hidrológicos, por lo que resulta conveniente emplear modelos para predicción de la respuesta de la cuenca que solo considere variables climatológicas y no tanto características físicas de la cuenca.

El filtro de Kalman es una serie 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, Hernández y Medina (2012) evaluaron la bondad de los modelos de asimilación de datos como inserción directa (ID) y filtro de Kalman para la estimación de perfiles de humedad de suelo a partir de la asimilación de valores de humedad superficial basado en el uso de la ecuación de Richards para la modelación del movimiento del flujo de agua. Hobson y Kardynal (2015) emplearon el modelo de filtro de Kalman en el espacio de estado para estimar los parámetros de movimiento de una población del Oeste de Veeries (*Catharus fuscescens*) en la Columbia Británica en Canadá durante su migración hacia la cuenca del Amazonas en Brasil. También se ha empleado en forma conjunta con otros modelos, Padilla *et al.* (2013) emplearon el filtro de Kalman para mejorar la estimación de parámetros de los modelos autorregresivos multivariados para la reconstrucción de registros del electroencefalograma (EEG). En el ámbito de la hidrología se ha empleado como un modelo de regresión con una estimación recursiva empleado para predecir la función de respuesta de la cuenca para el periodo t , es expresado como una

función lineal de caudales observados en intervalos de tiempo anteriores y/o de precipitación (Valdés, et al., 1980).

Recientemente Morales-Velázquez *et al.* (2014) aplicaron el filtro de Kalman discreto para la predicción de caudales en la cuenca de la presa Ángel Albino Corso (Peñitas), el modelo fue evaluado a partir de caudales obtenidos del tránsito inverso modificado en vasos o antitránsito medidos en la estación hidrométrica Sayula, 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 (DKF) junto con un modelo autorregresivo con entrada exógena (ARX) en la cuenca del río Turbio, México, las variables de entrada (precipitación y caudales) y salida (caudales) corresponde a información promedio diaria (24 horas), se 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, análogamente para la serie de datos de 2004 los valores de Nash-Sutcliffe fueron 0.93, 0.82, 0.72 y 0.62.

Con el objetivo de realizar el pronóstico de caudales de escurrimiento más preciso ya que resultan vitales para el manejo y control de los recursos hídricos, se hace necesario evaluar la capacidad de los modelos de caja negra en la predicción de la función de respuesta de la cueca alimentado por datos horarios.

1.1. Hipótesis

A partir de series de tiempo obtenidas de una red de instrumentación en una cuenca, como lluvia y caudales de escurrimiento, y a partir del uso de técnicas de estimación recursiva de estados como el algoritmo de filtro de Kalman discreto, es posible implementar un modelo eficiente y manejable para pronóstico de caudales horarios asociados a la función de respuesta de la cuenca.

1.2. Objetivos

1.2.1. General

Evaluar la eficiencia del pronóstico de caudales horarios con el algoritmo de Filtro de Kalman discreto (DKF) en conjunto con el modelo autorregresivo con entrada exógena (ARX) en la cuenca del río Huaynamota.

1.2.2. Específicos

- Implementar el modelo desarrollado para la cuenca del río Huaynamota.
- Comparar los pronósticos realizados a 1, 2, 4 y 6 horas hacia adelante.
- Evaluar la incertidumbre de los pronósticos considerando medidas de bondad de ajuste como el índice de eficiencia de Nash-Sutcliffe (E) y raíz del error cuadrático medio (RMSE).

1.3. Organización de la tesis

La tesis está organizada en 5 capítulos, el capítulo uno presenta la introducción del estudio realizado, el capítulo dos la revisión de literatura, en el tercero se incluye el artículo científico resultado del presente estudio de investigación conservando la estructura de los escritos para enviar a revistas arbitradas, en el capítulo 4 se encuentran las conclusiones y recomendaciones y el quinto contiene los anexos.

CAPÍTULO 2. REVISIÓN DE LITERATURA

2.1. Ciclo hidrológico

El agua en el mundo se encuentra naturalmente en diferentes formas y ubicaciones, en el aire, en la superficie terrestre, debajo del suelo y en mares y océanos. Existen reservas de agua dulce almacenadas como glaciares (68.7%), agua subterránea (30.1%) y permafrost (0.8%) Unesco (2006), la cual representa el 2.5% del agua en el planeta.

A pesar de que el volumen de agua dulce es importante, se hace necesario analizar los flujos de agua renovables anuales tomando en cuenta dónde y cómo se mueven a través del ciclo hidrológico.

El ciclo hidrológico en un modelo conceptual simplifica como el agua se mueve entre la tierra y la atmosfera en diferentes estados como gas, líquido y sólido, el cual contiene muchas simplificaciones (Davie, 2008).

En el ciclo hidrológico global los principales procesos son; precipitación, evaporación y escurrimiento. En los océanos la evaporación es mucho mayor al agua precipitada sobre ellos, mientras que en los continentes ocurre lo contrario. La diferencia entre precipitación y evaporación en los continentes se denomina escurrimiento, el cual se mueve sobre o debajo de la superficie para retornar a los océanos, con lo cual se completa el ciclo hidrológico.

En una escala más pequeña el ciclo hidrológico incluye los procesos como precipitación, intercepción vegetal, retención por depresiones topográficas, escorrentía, infiltración en el suelo, evaporación y transpiración. Actividades humanas pueden alterar los componentes del ciclo natural a través de

diversiones de uso de la tierra, reutilización y descarga de desechos en las vías de aguas superficiales y subterráneas naturales.

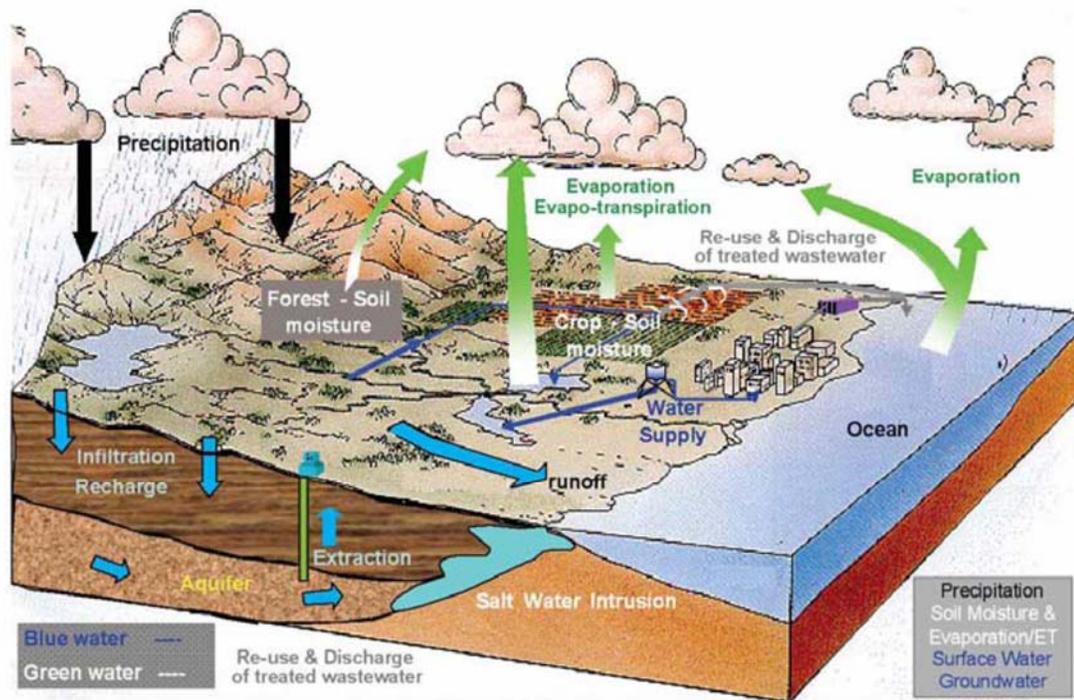


Figura 1. Esquema de las componentes del ciclo hidrológico (UNESCO, 2006).

2.1.1. Sistema hidrológico

Chong-yu (2002) define a un sistema hidrológico como un conjunto de procesos físicos, químicos y/o biológicos que actúan sobre una variable o variables de entrada, para convertirlos en una variable de salida (o variables). En el caso del ciclo hidrológico cada uno de sus componentes se puede agrupar en subsistemas del ciclo total. El ciclo hidrológico global representado como un sistema se muestra en la figura 2, el cual se divide en tres subsistemas; sistema de agua atmosférica, superficial y subsuperficial, cada uno con sus respectivos procesos.

De acuerdo a McCuen (2005) la hidrología requiere del análisis y la síntesis para emplear los conceptos fundamentales en la solución de problemas de ingeniería. Por lo que el funcionamiento de la cuenca puede representarse como un sistema

el cual está integrado por tres partes: la entrada, salida y la función de transferencia.

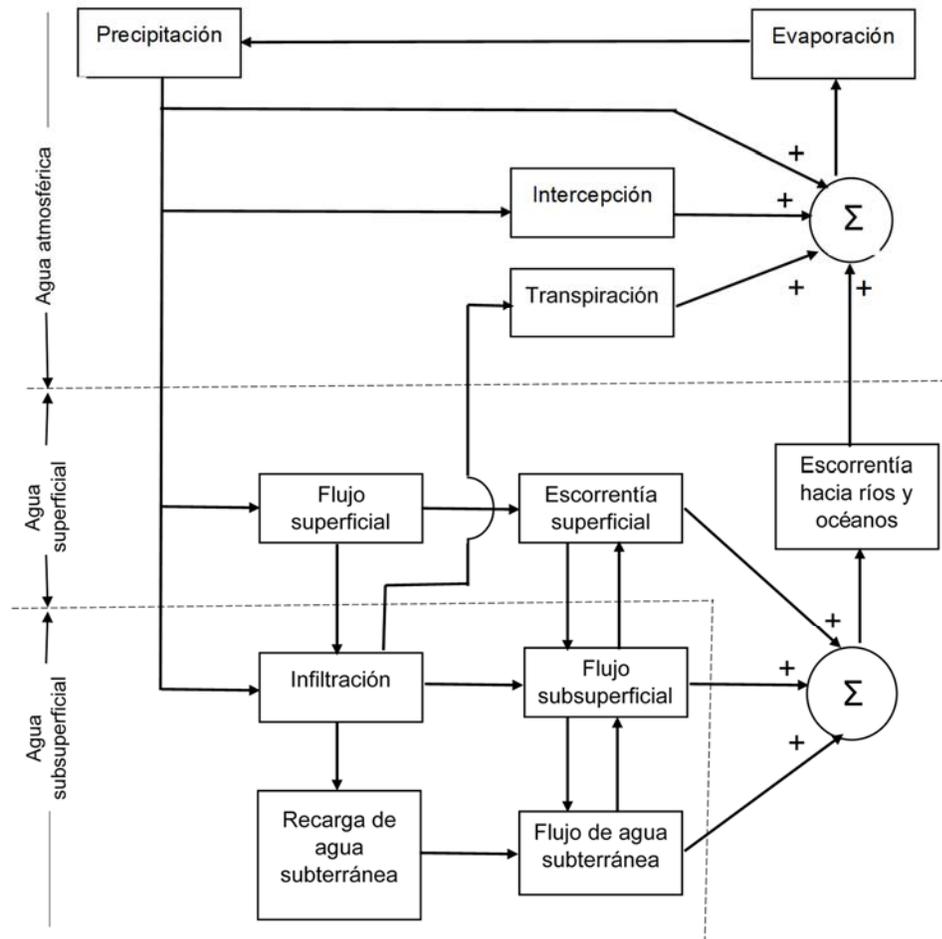


Figura 2. Diagrama de bloques del sistema hidrológico global (Chow *et al.*, 1994).

De acuerdo a McCuen (2005) la hidrología requiere del análisis y la síntesis para emplear los conceptos fundamentales en la solución de problemas de ingeniería. Por lo que el funcionamiento de la cuenca puede representarse como un sistema el cual está integrado por tres partes: la entrada, la salida y la función de transferencia.

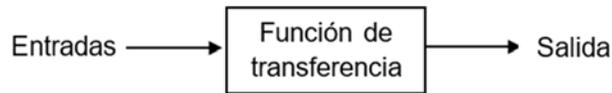


Figura 3. Representación conceptual del proceso del sistema (McCuen, 2005).

En la fase de análisis, la entrada y salida son conocidas, mientras que se debe de encontrar un modelo racional de la función de transferencia. En la fase de síntesis y diseño, la entrada de diseño y el modelo de la función de transferencia son conocidos, la salida del sistema se obtiene como resultado.

En el ciclo hidrológico la precipitación se puede ver como una entrada a la superficie de la tierra, en la superficie existen elementos de almacenamiento como la vegetación y depresiones, mientras que el escurrimiento en la superficie se considera una salida del sistema, lo anterior es la representación de los procesos físicos que controlan la escorrentía superficial. Otros análisis hidrológicos se muestran en la cuadro 1.

Los estudios hidrológicos se elaboran a escalas espaciales y temporales apropiados, de acuerdo al propósito en cuestión en la planificación y gestión de los recursos a mediano y largo plazo para una región, es apropiado una escala mensual o anual a una escala de cuenca, en el caso de diseño de estructuras de una ciudad, una escala temporal horaria o más fina sería la adecuada y una escala espacial menor a la cuenca.

Cuadro 1. Ejemplo de análisis hidrológico y síntesis.

Entradas	Modelo o función de transferencia	Salidas	Análisis		Síntesis	
			Conocida	Desconocido	Conocida	Desconocido
Coeficiente de escurrimiento (C) Intensidad de precipitación (i) Área de drenaje (A)	Fórmula racional $Q_p = CiA$	Descarga máxima (Q_p)	I, A, Q_p	C	C, i, A	Q_p
Hidrograma aguas arriba (I)	Ecuación de Muskingum $O_2 = C_0I_2 + C_1I_1 + C_2O_1$	Hidrograma aguas abajo (O_2)	I, O_2	C_0, C_1, C_2	I, C_0, C_1, C_2	O_2
Índice de erosión (R) Factor de erosionabilidad del suelo (K) Factor topográfico (T) Factor de cobertura (C) Factor de soporte (P)	Ecuación universal de pérdida de suelo $E = RKTCP$	Pérdida de suelo (E)	E, R, K, T	C, P	R, K, T, C, P	E
Longitud (HL) Pendiente (Y) Almacenamiento (S)	SCS ecuación de retraso $L = \frac{HL^{0.8}(S + 1)^{0.7}}{1900Y^{0.5}}$	Tiempo de retraso de la cuenca	L, HL, Y, S		HL, Y, S	L
Almacenamiento (S) Precipitación (P)	$Q = \frac{(P - 0.2S)^2}{P + 0.8S}$	Volumen de escurrimiento (Q)	Q, P	S	P, S	Q

Fuente: McCuen (2005).

2.2. Modelación para procesos hidrológicos

Clarke (1973) define a un modelo como una representación simplificada de un sistema complejo, un modelo implica similitud sin ser idéntico, en este sentido solo simula algunas de las características del sistema. De acuerdo a Chong-yu (2002) los modelos hidrológicos tienen dos objetivos, el primero es obtener una mejor comprensión de los fenómenos hidrológicos que operan en una cuenca, y de cómo los cambios en la cuenca puedan afectar estos fenómenos, el segundo es la generación de secuencias sintéticas de datos hidrológicos para el diseño de las instalaciones o para uso de predicción. Lo anterior permite mejorar la toma de decisiones sobre un problema hidrológico, ya sea en la planificación de recursos hídricos, protección contra inundaciones o mitigación de contaminación, Beven (2012).

2.2.1. Clasificación de modelos hidrológicos

Existen distintas clasificaciones de modelos hidrológicos, una clasificación mencionada por Chong-yu (2002) distingue entre modelos materiales y simbólicos o formales, como se muestra en la figura 4.

Modelo físico

El modelo material o físico corresponde a la representación de un sistema que satisface las leyes de similitud geométrica, cinemática y dinámica, cada ley se aplica de acuerdo a la utilidad del prototipo, ejemplo de ellos son las maquetas de cuencas hidrográficas de laboratorio o el modelo del vertedor de una presa. Respecto al modelo análogo es la representación del sistema físico por un sistema análogo, que simule su comportamiento de forma aproximada, ejemplo de ellos es la simulación del flujo de agua en un cauce o a través de un medio poroso, mediante el flujo de una corriente eléctrica en un circuito compuesto por resistencias y condensadores.

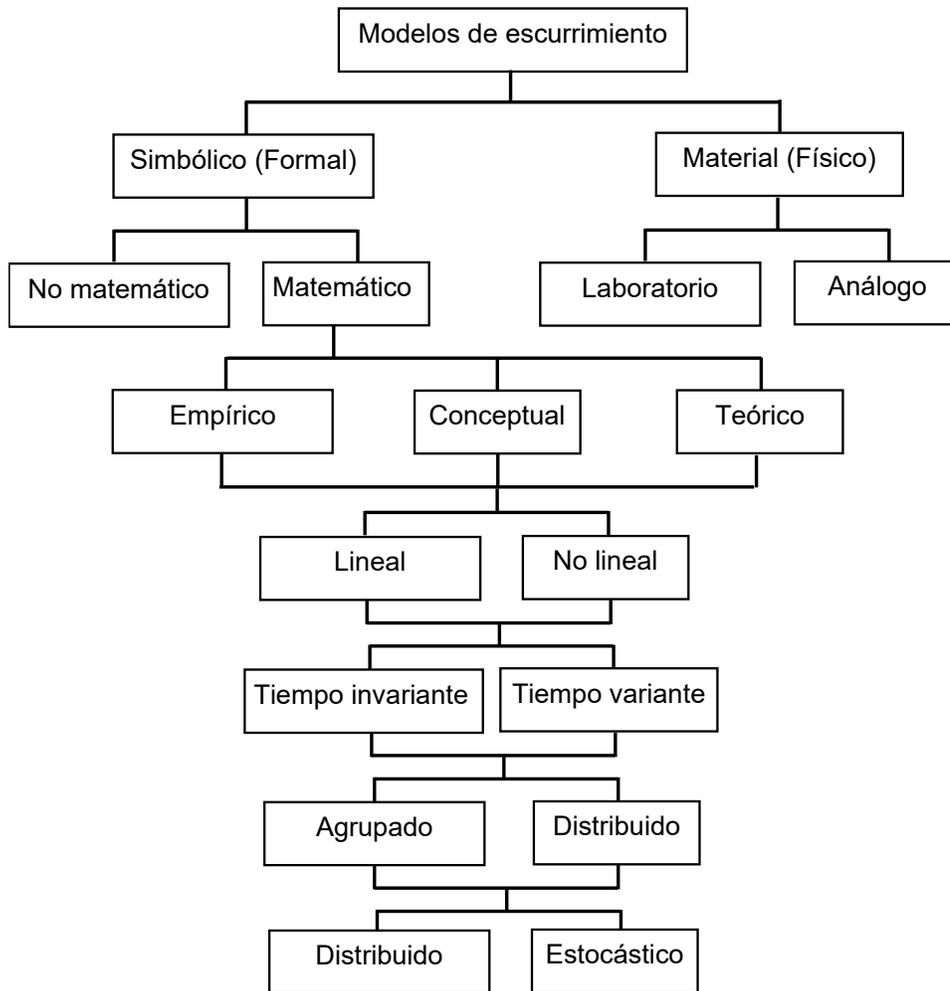


Figura 4. Clasificación de modelos hidrológicos (Chong-yu, 2002).

Modelo simbólico

Un modelo formal es la representación simbólica en términos lógicos de una situación idealizada, relativamente simple, que comparte las propiedades estructurales del sistema original, estas se pueden representar de forma matemática.

Considerando al sistema hidrológico y si se denotan por $x(t)$ e $y(t)$ las variables de entrada y salida del sistema en el tiempo t , un modelo matemático podría ser:

$$f[y(t), x(t), \partial y / \partial t, \partial x / \partial t; \partial y^2 / \partial t^2, \partial x^2 / t^2 \dots \dots, \theta_1, \theta_2 \dots \dots \dots] = 0$$

Donde $f(\bullet)$ es una función de la forma dada, θ_1, θ_2 , son parámetros medidos del sistema o estimados a partir de datos.

Cuando el sistema a modelar es muy complejo, puede ser adecuado para muchos propósitos adoptar una forma relativamente simple $f^*(\bullet)$ para la función $f(\bullet)$, para expresar la falta de ajuste por medio de un residuo o error, se convierte en:

$$f^* = (y_t, x_t; y_{t-1}, x_{t-1}; y_{t-2}, x_{t-2}; \dots \theta_1, \dots \theta_2) + \epsilon_t = 0 \quad (1)$$

Donde ϵ_t es el residual en el tiempo t .

De los modelos matemáticos, el modelo teórico o de base física es la consecuencia de las leyes más importantes que rigen los fenómenos en una cuenca. Los modelos empíricos o de caja negra, no ayudan a la comprensión física directa, solo pueden estimarse mediante el uso de mediciones de entrada y salida, un ejemplo son los modelos estocásticos de series de tiempo. Un modelo intermedio entre los anteriores son los modelos conceptuales, este emplea leyes físicas, pero de forma muy simplificada.

El resto de las clasificaciones se describen a continuación:

Modelo lineal y no lineal: En el sentido de teoría de sistemas es lineal cuando mantiene la validez del principio de superposición, en el sentido estadístico es lineal cuando las relaciones de los parámetros que expresan el comportamiento del sistema son lineales.

Tiempo invariante y variante: El modelo es invariante si su relación entre la entrada-salida no cambia con el tiempo, es decir que la salida depende solo de la forma de la entrada y no del momento en que se esta se aplique, caso contrario son los modelos invariantes. Los sistemas geohidrológicos son variantes en el tiempo, sin embargo, por simplicidad para su análisis se considera invariante.

Agrupado y distribuido: Respecto a la discretización espacial de los parámetros y variables del sistema, los modelos agrupados son aquellos que tratan a la cuenca o subcuencas como un todo homogéneo, o se consideran áreas separadas con características similares, mientras que en los modelos distribuidos el área de la cuenca se divide en una red, en donde en cada nodo de la cuadrícula contiene datos de entrada y se hace el análisis pertinente.

Determinístico y estocástico: Si alguna de las variables $x(t)$ e $y(t)$ son consideradas variables aleatorias que sea definida por una distribución de probabilidad entonces el modelo se considera estocástico, si todas las variables de la Ecuación (1) no tienen una distribución de probabilidad, entonces se considera un modelo determinista.

Otra clasificación que se hace de los modelos matemáticos en la hidrología la presenta UNESCO (1985) y se muestra en la figura 5.

2.2.2. Modelos lluvia-escorrentamiento

Los orígenes de los modelos lluvia escorrentamiento tiene su origen a mitad del siglo XIX, a partir del cual fueron evolucionando para resolver problemas para cuencas de distintos tamaños, así como para predecir la función de respuesta de la cuenca tomando la variabilidad espacial de las condiciones climáticas y de las condiciones físicas de la cuenca, con el desarrollo de la tecnología y de paquetería especializada se han generado modelos cada vez más complejos como los modelos de base física y a su vez modelos más simples como aquellos modelos de caja negra en donde se realiza un análisis estocástico de series de tiempo de variables de entrada y de salida, en los próximos apartados se hace una pequeña descripción de los principales modelos lluvia-escorrentamiento.

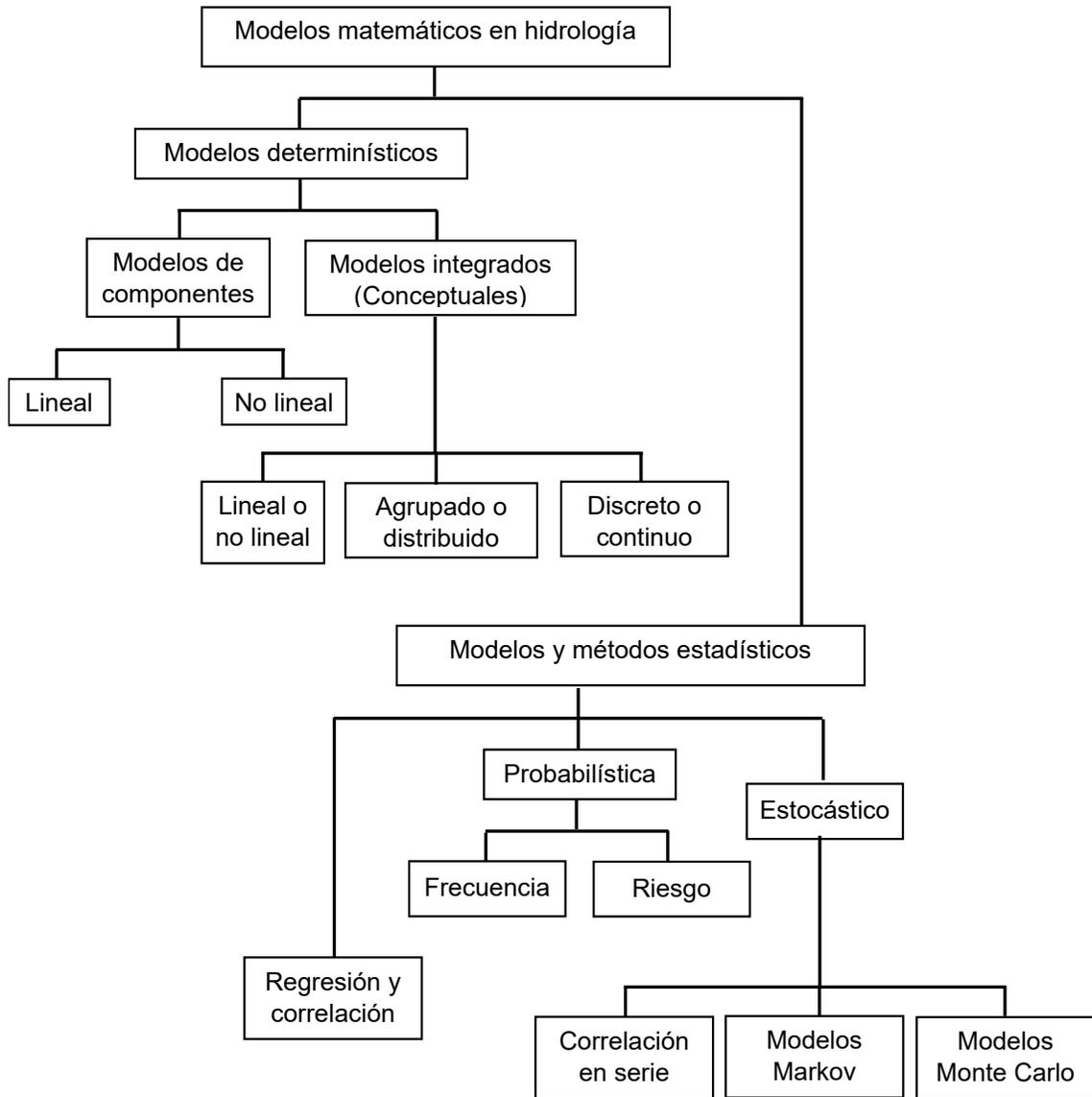


Figura 5. Modelos matemáticos en hidrología (UNESCO, 1985).

Fórmula racional

Esté método fue desarrollado en el año 1889, se emplea para obtener el caudal máximo causado por una precipitación de una intensidad dada en cuencas pequeñas, se introduce en concepto de tiempo de concentración que es el tiempo en que el caudal alcanza su valor máximo, la ecuación de la fórmula racional es:

$$Q_p = 0.278 C i A$$

Para aplicar este método es necesario calcular el tiempo de concentración, en México generalmente se utiliza el método de Kirpich. Se calcula también el coeficiente de escurrimiento que depende del área y uso de suelo, la intensidad se obtiene con apoyo de las curvas intensidad-duración-periodo de retorno (i-d-Tr).

A pesar de que este método en apariencia es sencillo se puede caer en errores de estimación del caudal, sobre todo con el cálculo del parámetro que indica el tiempo de respuesta de la cuenca, Almeida *et al.* (2014) realizaron una revisión y evaluación de distintas expresiones utilizadas para el cálculo del tiempo de concentración.

Hidrograma unitario

Este método fue desarrollado por Sherman (1932) el cual plantea estimar la forma del hidrograma resultante de un evento de precipitación, el hidrograma unitario corresponde al hidrograma generado por una lámina de lluvia en exceso unitaria 1cm, 1mm o 1 pulgada.

De acuerdo a Aparicio (2013) las hipótesis a partir de las cuales se desarrolló el hidrograma unitario son:

- a) Tiempo base constante. Para una cuenca dada, la duración total de escurrimiento directo o tiempo base es la misma para todas las tormentas con la misma duración de lluvia efectiva, independientemente del volumen total escurrido. Todo hidrograma unitario ligado a una duración de lluvia en exceso.
- b) Linealidad o proporcionalidad. Las ordenadas de todos los hidrogramas de escurrimiento directo con el mismo tiempo base, son directamente proporcionales al volumen total de escurrimiento directo, es decir, al volumen total de lluvia efectiva. Como consecuencia, las ordenadas de dichos hidrogramas son proporcionales entre sí.

c) Superposición de causas y efectos. El hidrograma que resulta de un periodo de lluvia dado puede superponerse a hidrogramas resultantes de periodos lluviosos precedentes, como se muestra en la figura 6.

El hidrograma de escurrimiento esta conceptualmente separado en dos partes: escurrimiento directo y caudal base. La escurrimiento directa es la escurrimiento de la tormenta que resulta del exceso de lluvia, la función de transferencia, o hidrograma unitario, es la función que transforma el exceso de lluvia en la escurrimiento directa. El flujo base es la escurrimiento que ha resultado de la acumulación de agua en la cuenca hidrográfica a partir de eventos pasados, los cuales aparecen como flujo de corriente.

El proceso por el cual la tormenta de diseño se combina con la función de transferencia para producir el hidrograma de escurrimiento directa se llama convolución.

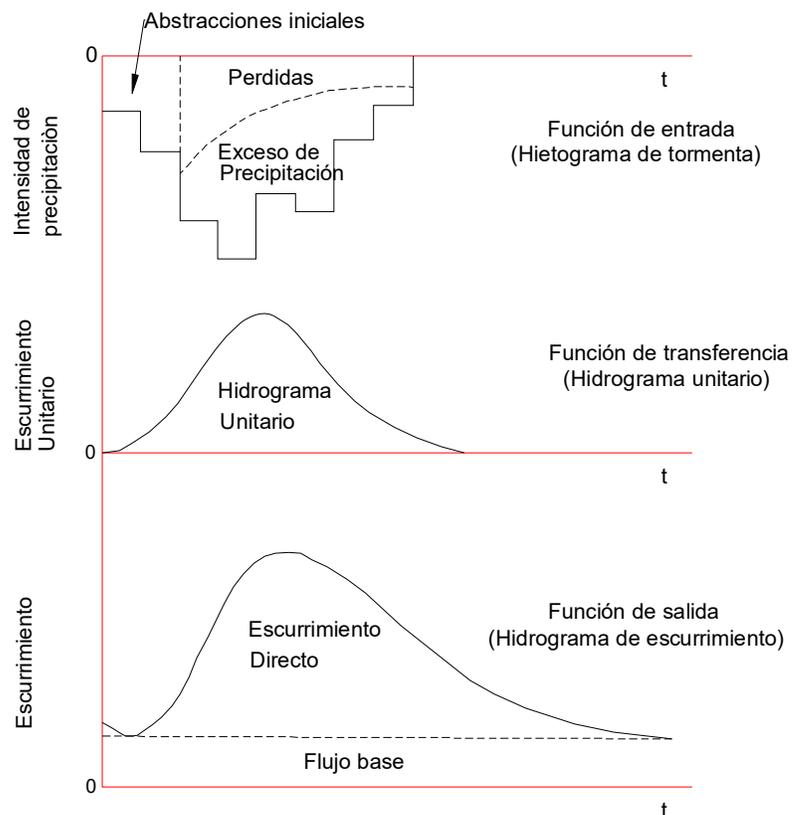


Figura 6. Proceso lluvia-escurrimiento (Aparicio, 2013).

Modelos de base física

Este tipo de modelos son complejos en el sentido que intentan representar matemáticamente los procesos físicos de una cuenca, utiliza variables de estado que son medibles distribuidas espacial y temporalmente, requiere datos de parámetros físicos de la cuenca como contenido de humedad del suelo, topografía, topología, dimensiones de la red hidrográfica, etc. (Gayathrin *et al.*, 2015). Como ejemplo se encuentran modelos como SWAT, MIKE SHE y HEC-HMS.

SWAT

Fue desarrollado para predecir el impacto del manejo del suelo en la circulación de agua, sedimentos y contaminación por sustancias químicas agrícolas en cuencas grandes con distintas condiciones físicas de la cuenca para un periodo prolongado.

El ciclo hidrológico es simulado basado en la ecuación de equilibrio del agua:

$$SW_t = SW_0 + \sum (R_{day} - Q_{surf} - E_a - W_{seep} - Q_{gw})$$

Donde SW_t es el contenido de agua en el suelo, SW_0 es el contenido inicial de agua del suelo, R_{day} es la cantidad de precipitación, Q_{surf} es la cantidad de escorrentía de la superficie, E_a es la cantidad evapotranspirada, W_{seep} es la cantidad de agua percolada en el perfil de suelo y Q_{gw} es la cantidad de flujo de retorno, (Neitsch *et al.* 2011).

Para propósitos de simulación la cuenca hidrográfica se divide en subcuencas y a su vez en unidades de respuesta hidrológica (HRUs) que corresponde a todas las unidades posibles de combinación de cobertura de tierra, suelo y pendiente de la cuenca. La simulación hidrológica se realiza en dos fases: la primera corresponde a la fase terrestre del ciclo hidrológico la cual controla la cantidad de agua, sedimentos, carga de alimentos nutritivos y pesticidas, la segunda fase

corresponde a la fase de enrutamiento del ciclo, el cual define el movimiento del agua, sedimentos a través de la red hidrográfica hacia el vertedero.

MIKE SHE

Este modelo distribuido incluye procesos del flujo superficial, flujo en zona saturada, evapotranspiración basada en la vegetación, flujos subterráneos y flujo en canal totalmente dinámico. Permite que cada proceso se resuelva a su propia escala espacial y temporal relevante, por ejemplo la evapotranspiración la calcula a lo largo del día así como los flujos superficiales ya que tienen una respuesta rápida a la lluvia, mientras que el flujo subterráneo reacciona más lentamente (DHI, 2017).

Los parámetros básicos de entrada al modelo son la extensión del modelo, topografía, precipitación y dependiendo de los procesos hidrológicos incluidos se requieren datos de evapotranspiración de referencia, temperatura del aire, radiación solar, morfología del río, geología del subsuelo, distribución del suelo y uso de la tierra.

HEC-HMS

El modelo HEC-HMS fue diseñado para simular los procesos lluvia-escorrentía en sistemas de cuencas dendríticas, está diseñado para resolver problemas en áreas geográficas grandes o cuencas pequeñas o urbanas. El modelo calcula el hidrograma producido en una cuenca, para ello se tienen cuatro fases principales del proceso: 1. Separación de la lluvia neta, 2. Cálculo de la escorrentía directa, 3. Suma a la escorrentía directa la escorrentía básica (los cálculos para estas tres fases se realizan a nivel subcuenca), 4. Tránsito de hidrogramas (se calcula cómo evoluciona un hidrograma a medida que el flujo transita a lo largo de un cauce o a través de un depósito o embalse). En los puntos de unión suma los caudales generados por varios elementos, por lo que genera un hidrograma por subcuenca y a la salida de la cuenca de estudio.

En el modelo perceptual el hidrólogo define como la cuenca responde a la lluvia bajo diferentes condiciones, dependiendo del conjunto de sitios y datos analizados, el modelo perceptivo muestra las complejidades de los procesos de flujo. En el modelo conceptual la descripción de procesos se debe de hacer explícitos mediante ecuaciones matemáticas, el modelo puede ser simple o tan complejo para utilizar ecuaciones diferenciales parciales no lineales las cuales pudieran resolverse de forma analítica o con técnicas de análisis numérico en forma de código para realizar el análisis con medios computacionales a lo que se le conoce como modelo de procedimiento.

Para poder utilizar el código para hacer predicciones es necesario antes realizar una calibración de parámetros que definen las características del área de captación o dominio de flujo, se pueden incluir características tales como porosidad o conductividad hidráulica en diferentes horizontes de suelo, generalmente se consideran constantes durante el periodo de simulación, una técnica utilizada en la calibración de parámetros es el ajuste de sus valores hasta alcanzar la mejor coincidencia entre las predicciones del modelo y las observaciones medidas en el punto de salida de la cuenca, los valores de los parámetros se ajustan entre las ejecuciones del modelo, ya sea manualmente por el modelador o por algún algoritmo de optimización computarizado. Cada modelo tiene cierta cantidad de parámetros que se puedan ajustar, y de estos pueden existir una infinidad de combinaciones que puedan dar un buen ajuste en los resultados de salida.

Una vez realizada la calibración de parámetros es posible realizar una nueva simulación para hacer predicciones de la respuesta de la cuenca a la cual se le llama validación.

2.3. Modelo Autorregresivo con entrada exógena (ARX)

Para la predicción del comportamiento de sistemas lineales en tiempo discreto, se puede realizar el análisis series de tiempo, para ello se emplean modelos matemáticos paramétricos, uno de los modelos más simple es el modelo Autorregresivo (AR) en donde solo se analiza información a la salida del sistema a este tipo de modelo se le llama error de ecuación, sin embargo, para mejorar la precisión del modelo, es conveniente introducir variables de entrada en el sistema, en este caso a la entrada se denomina entrada exógena y al modelo se le denomina de entrada/salida, en este sentido el modelo autorregresivo con entrada exógena (ARX) es una extensión del modelo AR. Jairo, *et al.* (2003).

El modelo general ARX (Ljung L. 1988-2017) tiene la siguiente estructura:

$$y_{(t)} + a_1 y_{(t-1)} + \dots + a_{na} y_{(t-na)} = b_1 u_{(t-nk)} + \dots + b_{nb} u_{(t-nk-nb+1)} + e_{(t)}$$

Donde:

$y_{(t)}$, es el vector que representa la variable de salida del sistema en el tiempo t.

$u_{(t)}$, es el vector que representa la variable de entrada en el tiempo t.

na y nb , son los números del paso de tiempo considerados previos a la estimación del estado del sistema.

nk , es el retardo de entrada, es decir es el número de muestras antes de que la entrada afecte a la salida del sistema.

$e_{(t)}$, es el ruido blanco asociado a la variable de salida.

Hsu *et al.* (2009) indican que el modelo ARX que se ha utilizado para predecir el flujo en el cauce utilizando series de tiempo de precipitación y escurrimiento es el siguiente:

$$y_{t+1} = \sum_{i=0}^{na} a_i y_{t-i} + \sum_{j=0}^{nb} b_j r_{t-j} + e_{t+1}$$

Donde:

a_i y b_j , son parámetros.

y_t , es la secuencia de datos de escurrimiento.

r_t , es la secuencia de datos de precipitación.

e_{t+1} , es el error de la estimación del escurrimiento.

2.3.1. Estimación del orden del modelo

El orden del modelo define la complejidad del mismo, en el caso de la estructura del modelo ARX corresponde a los valores de n_a , n_b y n_k . Estos son datos de entrada para el modelo ARX, por lo que se deben de estimar, para ello se utilizará el System Identification Toolbox™ de MATLAB®.

Para determinar el orden del modelo se debe de contar con una serie de datos tanto de la entrada y salida cuando se trata de un modelo ARX con solo una entrada y una salida (SISO), o las entradas y salidas, cuando se trata de un modelo de múltiples entradas y salidas (MIMO). Debe de existir una serie de tiempo para la estimación de parámetros y otra para la validación, si solo se cuenta con una serie de datos, un porcentaje de la misma se puede utilizar en la estimación y el resto en la validación.

Para la elección de los órdenes del modelo se utilizan tres criterios, 0 que elige las órdenes que mejor ajusten a los datos de validación, y los criterios de Rissanen MDL e Información de Akaike (AIC).

2.4. 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 $t - 1$, añadiendo un término de corrección proporcional al error de predicción, este último es minimizado estadísticamente.

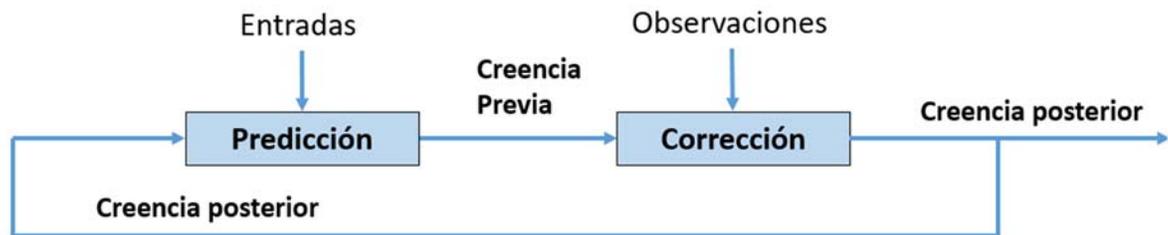


Figura 8. Esquema general del Filtro de Kalman.

La etapa de predicción corresponde a la dinámica del proceso del sistema, mientras que en la etapa de corrección se utiliza la ecuación de medición y se utilizan lecturas de mediciones físicas del estado del sistema que se desee predecir.

De acuerdo a Welch y Bishop (2006) trata de estimar el estado de $X \in \mathbb{R}^n$ de un proceso controlado en tiempo discreto. La ecuación que representa al modelo del proceso del sistema es:

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

Donde:

x_t , es el vector que representa el estado del sistema en el tiempo t .

x_{t-1} , es el vector que representa el estado del sistema en el tiempo $t-1$.

U_t , es el vector de control.

w_t , es el vector que representa el ruido del sistema.

A , es la matriz de transición del estado, de dimensiones apropiadas.

B , es la matriz de transición de las mediciones, de dimensiones apropiadas.

La ecuación que representa al modelo observacional $z \in \mathbb{R}^m$ es:

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

Donde:

z_t , es el vector que contiene las mediciones actuales del estado del sistema.

v_t , es el vector que representa el ruido en las mediciones.

H , es la matriz de dimensiones apropiadas.

Las variables aleatorias w_t y v_t representan el ruido del proceso y de medición, respectivamente, las cuales son independientes entre sí, con ruido blanco, se pueden describir por una función de probabilidad normal.

$$p(w) \sim N(0, Q)$$

$$p(v) \sim N(0, R)$$

En la práctica, la covarianza del ruido de proceso Q y la de medición R puede cambiar a cada paso en el tiempo, sin embargo, generalmente se consideran constantes.

Con las Ecuaciones (2) y (3) el filtro de Kalman trata de obtener un estimador óptimo en el sentido de una menor varianza de predicción, del estado del sistema $\hat{x}_{t/t}$ basado en las observaciones pasadas del estado del sistema:

$$z_t = \{z_1 \ z_2 \ \dots \ \dots \ \dots \ \dots \ z_t\}$$

y del vector de control de entradas:

$$U_{t-1} = \{U_1 U_2 \dots \dots \dots U_{t-1}\}$$

El estimador \hat{x} se define como la media de la distribución condicional de x_t dados z_t y U_{t-1} .

$$\hat{x}_{t/t} = E\{x_t/z_t, U_{t-1}\}$$

Donde el término t/t indica que se ha usado toda la información disponible hasta el tiempo t en la obtención del estimador del estado del sistema. El algoritmo del filtro de Kalman discreto se muestra en la figura 9.

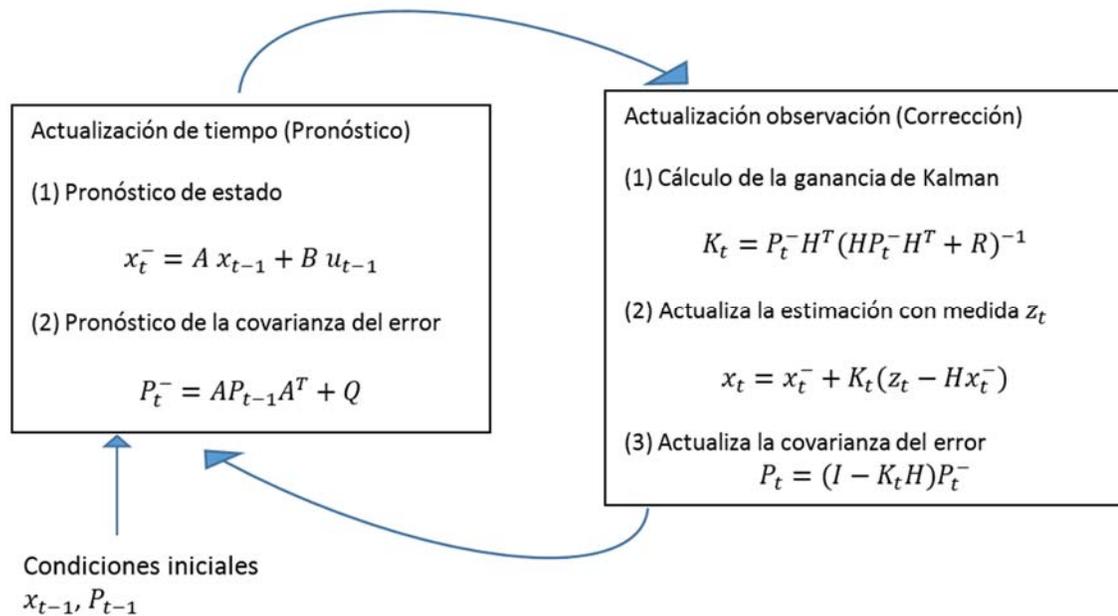


Figura 9. Algoritmo del Filtro de Kalman discreto (Welch y Bishop, 2006).

P_t es la matriz de covarianza del error de la estimación de x_t basado en $z_1, z_2, \dots \dots \dots z_t$. En el estado inicial, Valdés *et al.* (1980) y Morales-Velázquez *et al.* (2014) indican que este 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, I es la matriz identidad.

A la parte $z_t - H\hat{x}_t$ se le llama innovación del filtro y mide la discrepancia entre la estimación del estado en el tiempo x_t , dado la nueva observación z_t .

La covarianza del ruido en la medición R, de acuerdo a Valdés *et al.* (1980) se define como $\alpha \cdot Q_{t-1}$, es decir la varianza corresponde a un valor proporcional del caudal, González-Leiva *et al.* (2015) indican que el valor de α es del orden del 5% que representa el error que probablemente se haya cometido en la medición del caudal en la estación hidrométrica.

2.5. Criterios de eficiencia para modelos hidrológicos

La necesidad de evaluar el desempeño de los modelos radica en estimar la capacidad del modelo para reproducir el comportamiento del sistema, en este caso es la función de la respuesta de la cuenca, para ello existen distintas medidas de bondad de ajuste que permiten evaluar al modelo de modo cuantitativo.

2.5.1. Criterio de eficiencia de Nash-Sutcliffe (E)

El criterio fue propuesto por Nash y Sutcliffe (1970), se utiliza ampliamente en modelos hidrológicos y se define como “uno menos la suma de las diferencias al cuadrado absoluto entre los valores pronosticados y los observados normalizados por la varianza de los valores observados durante el periodo objeto de investigación” (Krause, 2005).

$$E = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2}$$

El rango de E se encuentra entre 0 a 1 siendo este último indicativo de un ajuste perfecto.

2.5.2. Raíz del error cuadrático medio (RMSE)

Esta medida de bondad de ajuste permite conocer la magnitud de la desviación de los valores simulados respecto a los valores observados.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}}$$

El rango de RMSE es de 0 a infinito, el valor mínimo indica al ajuste perfecto, conforme el valor se incrementa el ajuste de los valores pronosticados es menor.

2.6. Zona de estudio

2.6.2. Cuenca del río Santiago

La cuenca del río Santiago pertenece a la Región Hidrológica 12 Lerma-Santiago que se encuentra en los estados de Aguascalientes, Guanajuato, Jalisco, Nayarit y Zacatecas. Se origina por los desbordamientos del Lago de Chapala, controlados por la presa Poncitlán. Entre sus afluentes principales están los ríos Verde, Juchipila, Bolaños, Apozolco y Huaynamota. A partir del lago de Chapala hasta su desembocadura en el mar, el río tiene una longitud de 475 kilómetros, el área de la cuenca es de 78,809 km² y se divide en 6 subcuencas como se muestra en la figura 10.

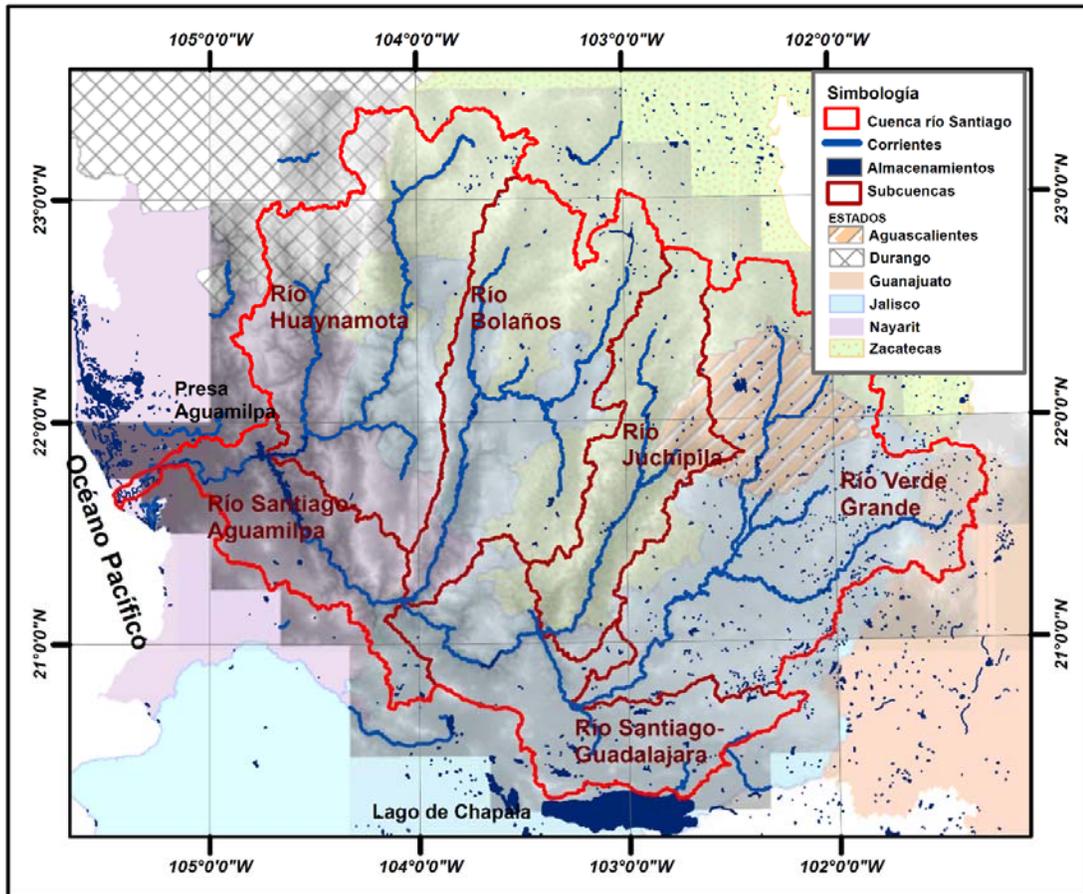


Figura 10. Cuenca del río Santiago.

2.6.2. Cuenca Chapalagana

La zona de estudio para el modelo lluvia-escorrimento es la subcuenca del río Huaynamota al oeste de la cuenca del río Santiago, la cual se delimitó considerando la disponibilidad de información, el punto para evaluar la predicción de caudales es la estación Chapalagana, por lo que solo se consideró el afluente de la margen izquierda como se muestra en la figura 11, la cuenca de estudio tiene un área aproximada de 12,079.54 km².

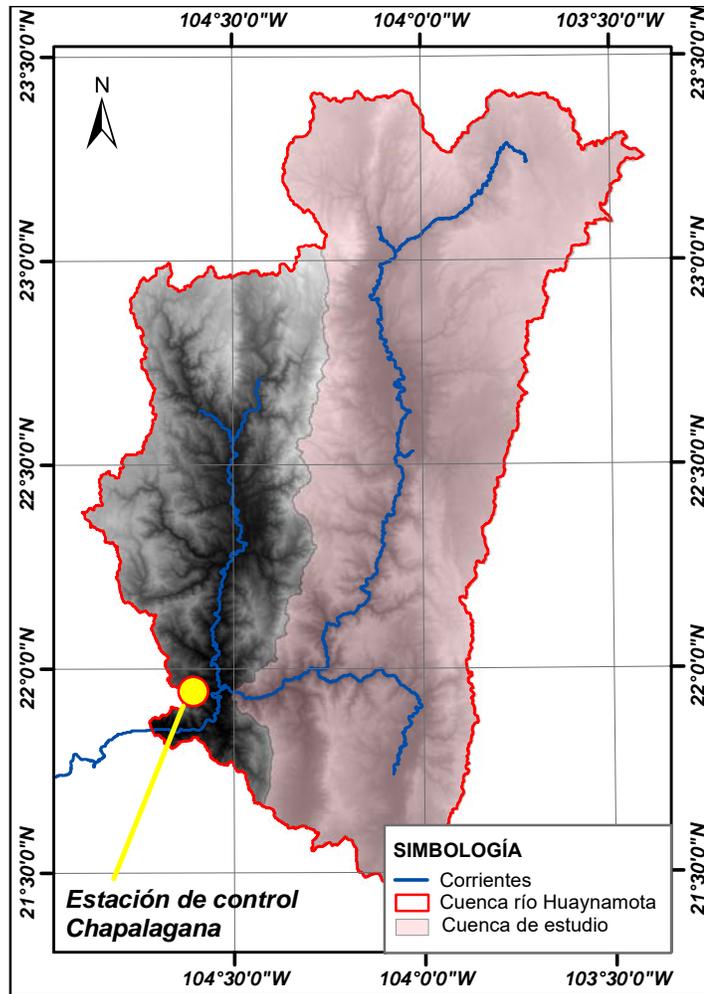


Figura 11. Cuenca Chapalagana.

2.6. Referencias

- Almeida K., I., Almeida K., A., Anache A, J. A., Steffen, J. L., Alves Sobrinho, T. (2014). Estimation of time of concentration in overland flow in watersheds: A review. *Geosciences*, 33 (4), 661–671.
- Aparicio M. F. J. (2013). *Fundamentos de Hidrología de Superficie*. Limusa, México, 303p.
- Beven, K. (1989). Changing ideas in hydrology-the case of physically-based models. *Journal of Hydrology*, 105, 157-172.
- Castañeda C., J. A., Nieto A. M. A. y Ortiz B. V. A. (2013). Analysis and application of the Kalman filter to a signal with random noise. *Scientia en Technica*. *Scientia en Technica*, 18(1), 267-274.
- Clarke, R. T., (1973). A review of some mathematical models used in hydrology, with observations on their calibration and use. *Journal of Hydrology*, 19, 1-20.
- Chong-yu Xu. (2002). *Hydrologic models*. Uppsala University. Suecia, 168p.
- Chow, V. T., Maidment, D. R. y Mays, L.W. (1994). *Hidrología Aplicada*. McGraw-Hill. Bogotá, Colombia. 584p.
- Comisión Federal de Electricidad. (2018). *Sistema de Monitoreo de Cuencas de la CFE*.
<https://h06814.ife.org.mx/cuencas/logon.aspx?ReturnUrl=%2fcuencas%2fdefault.aspx>. (Fecha de consulta: Septiembre 28 del 2017).
- Davie T. (2008). *Fundamentals of Hydrology*. Routledge Taylor & Francis Group, New York, USA. 200p.
- DHI (2017). *MIKE SHE, User Manual Volume1: User Guide v. 2017*, Danish Hydraulic Institute, Denmark, 420 pp
- Jean-Philippe, D., J. Kalman filtering in hydrological modeling. DAIHM Technical Report 2003-1, DHI Water & Environment, 2003.
- Gayathri, K. D., Ganasri, B. P. and Dwarakish, G. S. (2015). A Review on Hydrological Models. *Aquatic Procedia*, 4, 1001-1007
- González-Leiva, F., Ibáñez-Castillo, L. A., Valdés, J. B., Vázquez-Peña, M. A., & Ruiz-García, A. (2015). Pronóstico de caudales con Filtro de Kalman Discreto en el río Turbio. *Tecnología y Ciencias del Agua*, 6(4), 5-24.

- Hsu, K.-L., Moradkhani, H., and Sorooshian, S. (2009). A sequential Bayesian approach for hydrologic model selection and prediction. *Water Resources Research*, 45, W00B12. doi:10.1029/2008WR006824.
- Huang, W. C. (1999). Kalman Filter Effective to Hydrologic Routing?. *Journal of Marine Science and Technology*, 7(1), 65-71.
- Jairo S., J, Escobar D., A, Peña O., R. (2003). Identificación con modelos discretos para sistemas lineales. *Modelo matemático y aplicaciones. Ciencia, Investigación, Academia y Desarrollo*, 8(2), 47-55. doi: 10.14483/23448393.2691.
- Krause, P., D.P. Boyle y F. Base. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences* 5: 89-97.
- Leqiang S., Ousmane S., Ioan N. & Kailei L. (2015). Review of the Kalman type hydrological data assimilation. *Hydrological Sciences Journal*. doi: 10.1080/02626667.2015.1127376.
- Ljung L. (1988-2017). *System Identification Toolbox for Use with MATLAB*. The Mathworks INC. Massachusetts, USA. 226 p.
- McCuen, R.H. (2005). *Hydrologic Analysis and Desing*. Prentice-Hall, New Jersey, USA. 814p.
- Mendoza, M., Bocco, G., Bravo, M., Siebe, C. y Ortiz M. A. (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. *Investigaciones Geográficas*, 47, 36-58.
- Morales-Velázquez, M. I., Aparicio, J., & Valdés, J. B. 2014. Pronóstico de avenidas utilizando el filtro de Kalman discreto. *Tecnología y Ciencias del Agua*, 5(2), 85-110.
- Nash, J. E. and Sutcliffe, J. V. (1970). River flow forecasting through conceptual models, Part I - A discussion of principles, *Journal of Hydrology*, 10, 282–290.
- Neitsch, S. L., Arnold, J.R., Kiniry, J. R., Williams, J, R. (2011). *Soil and Water Assessment Tool Theoretical Documentation*. Version 2009. Blackland Research Center, Texas Agricultural Experimental Station, Temple, TX. 618p.
- Semenova, O y Beven, K. (2015). Barriers to progress in distributed hydrological modelling. *Hydrological Processes*, 29, 2074-2078. doi: 10.1002/hyp.10434

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.

UNESCO. (1985). Teaching aids in hydrology, Universitaires de France, Vendome. Recuperado de :
<http://unesdoc.unesco.org/images/0006/000653/065326eo.pdf>

Valdés, J. B., Velázquez, M. J. and Rodríguez-Iturbe, I. (1980). Filtros de Kalman en la Hidrología: Predicción de descargas fluviales para la operación óptima de embalses. Informe Técnico No 80-2. Universidad Simón Bolívar, Decanato de estudios de Posgrado, Posgrado en Planificación e Ingeniería de Recursos Hídricos. Venezuela. 69p.

Welch, G., & Bishop, G. (2006). An Introduction to the Kalman Filter. Chapel Hill: University of North Carolina at Chapel Hill, TR 95-041.

CAPÍTULO 3. PRONÓSTICO HORARIO DE CAUDALES MEDIANTE FILTRO DE KALMAN DISCRETO EN EL RÍO HUAYNAMOTA

STREAMFLOW FORECASTING HOURS FOR THE HUAYNAMOTA RIVER USING THE DISCRETE KALMAN FILTER¹

3.1. Resumen

Derivado de los eventos de precipitación extrema provocada por el cambio climático así como la alteración acelerada de las condiciones naturales de las cuencas por el crecimiento poblacional, es importante pronosticar los caudales que generan las cuencas derivado de los eventos de precipitación. En este estudio se utilizó un algoritmo de Filtro de Kalman Discreto (DKF) para el pronóstico de caudales, en conjunto con un modelo autorregresivo con entrada exógena (ARX). Los parámetros del modelo son definidos, y, posteriormente son recalculados cada cierto periodo de tiempo. El pronóstico de caudales se realizó seis pasos hacia adelante ($L = 1, 2, 3, 4, 5$ y 6 horas). La cuenca de estudio forma parte de la cuenca del río Huaynamota, delimitada hasta la estación meteorológica Chapalagana, ubicada aguas arriba de la presa Aguamilpa, en el estado de Nayarit. La cuenca del río Huaynamota es un tributario del río Santiago. Se emplearon series de datos horarias, tanto para precipitación como para caudal, para el periodo de agosto a septiembre de 2017. El modelo de pronóstico DKF-ARX reportó índices de eficiencia de Nash-Sutcliffe entre 0.99 y 0.85 para $L=1$ y $L=6$, respectivamente.

Palabras clave: Filtro de Kalman discreto, modelo ARX, predicción caudal.

¹ Leticia Alvarado Hernández, Laura A. Ibáñez Castillo, Agustín Ruíz García, Mario A. Vázquez Peña. Se enviara para su revisión a una revista para su publicación.

3.2. Introducción

La comprensión de la función de respuesta de la cuenca resulta de gran importancia para la planeación y manejo de los recursos hídricos, sin embargo, debido a la complejidad de todos los procesos que intervienen en el ciclo hidrológico, así como sus interrelaciones superficiales y subsuperficiales, se hace necesaria la abstracción que permita la comprensión y control de algunos aspectos de su comportamiento, lo cual es objeto de la modelación de cuencas (Chong-yu Xu, 2002).

Los modelos del proceso lluvia-escorrentía se han desarrollado a partir del 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 Gayathri *et al.* (2015), este tipo de modelos son más complejos, aunque no necesariamente los más eficaces, Beven (1989) y Mendoza *et al.* (2002) indican que esto es debido 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 Wei *et al.* (2018).

Uno de los mayores inconvenientes para la generación de modelos de lluvia-escorrentía de cualquier tipo, es la falta de información para alimentarlos, así como para validarlos. Las bases de datos climatológicos que se generan en el país son principalmente precipitación y caudales de escorrentía, 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; como es el caso de la cuenca del río Santiago, del cual forma parte la zona de estudio en este trabajo. Sobre el río Santiago, en Nayarit, se encuentran tres presas para generación de energía eléctrica: Aguamilpa, El Cajón y La Yesca. Considerando este tipo de información para

predecir la respuesta de la cuenca, se han desarrollado 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. (Hernández y Medina, 2012; Hobson y Kardynal, 2015; Padilla et al., 2013; y Valdés et al. 1980).

Recientemente Morales-Velázquez et al. (2014) aplicaron el filtro de Kalman discreto para la predicción de caudales horarios en la cuenca de la presa Ángel Albino Corso (Peñitas), el modelo fue evaluado a partir de caudales obtenidos del tránsito inverso modificado en vasos o antitránsito medidos en la estación hidrométrica Sayula, 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; el autor señala que no 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, análogamente para la serie de datos de 2004 los valores de Nash-Sutcliffe fueron 0.93, 0.82, 0.72 y 0.62.

Considerando 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 a ser utilizada en los sistemas de alerta

temprana o en los planes de operación del sistema de presas hidroeléctricas, por lo que el objetivo de esta investigación es 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 horas de anticipación, en la cuenca del río Huaynamota en los estados de Durango, Nayarit, Jalisco y Zacatecas.

3.3. Materiales y métodos

Se utilizó el algoritmo del filtro de Kalman discreto (DKF) en conjunto con un modelo autorregresivo con entrada exógena (ARX) para predecir los caudales, considerando series de tiempo pasadas de precipitación y caudales. La corriente considerada, río Chapalagana o río Atengo, es uno de los dos principales tributarios de la cuenca del río Huaynamota, delimitada hasta la estación meteorológica Chapalagana ubicada en el estado de Nayarit; 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 (12080 km²), se une al río Jesús María (5185 km²) y juntos forman el río Huaynamota. Una vez unidos, 25 kilómetros aguas abajo, se encuentra la presa hidroeléctrica Aguamilpa (INEGI, 2018).

3.3.1. Cuenca Huaynamota

La cuenca del río Huaynamota se localiza en el centro de México con coordenadas extremas 21°24'36.82" y 23°25'3.26" latitud norte y 104°30'34.03" y 103°24'26.73" longitud oeste, abarca parte de los estados de Durango, Jalisco, zacatecas y Nayarit (figura 1). El área de la cuenca es de 12,079.5 km², las elevaciones se encuentran en un rango de 216 a 3148 msnm, la elevación mínima corresponde a la salida de la cuenca en la estación Chapalagana. La pendiente media de la cuenca es del 31%. La longitud y pendiente de su cauce

principal son de 321 km y 0.68%. El tiempo de concentración de la cuenca es de 44 horas.

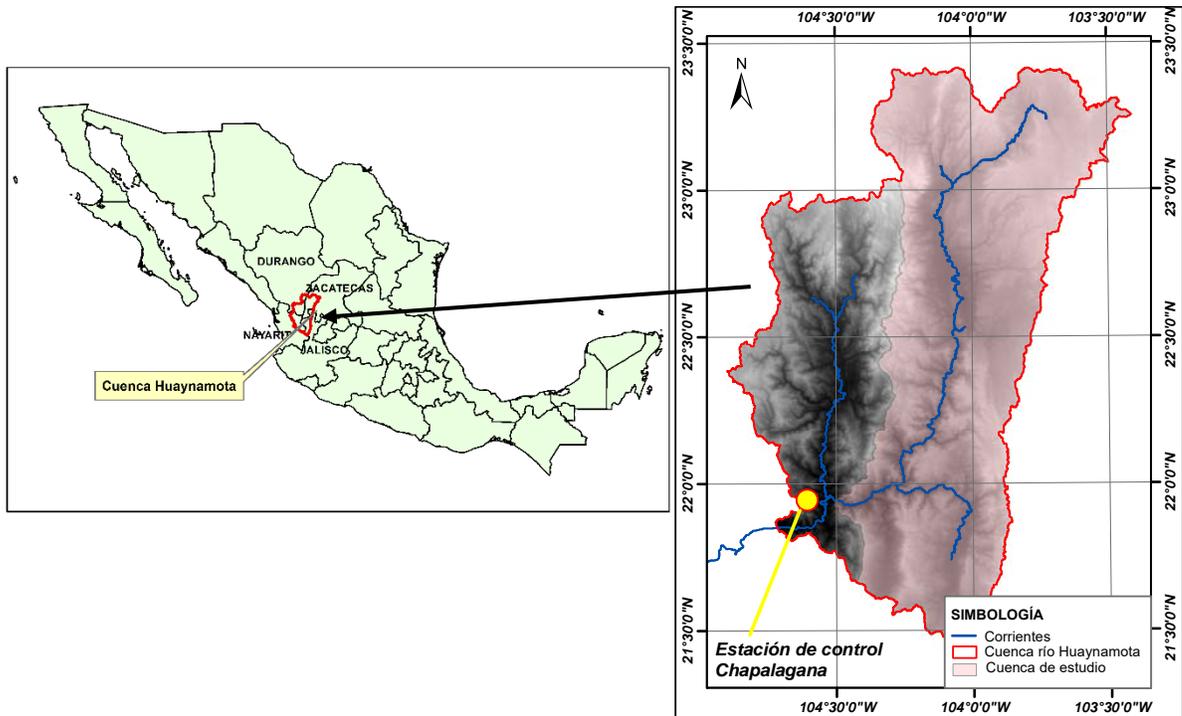


Figura 1. Ubicación de la cuenca del río Huaynamota.

3.3.2. Información climatológica

La información climatológica del periodo de trabajo del año 2017, se obtuvo de la red de estaciones climatológicas de la Comisión Federal de Electricidad (CFE) ubicadas en la cuenca del río Santiago, las estaciones correspondientes son: Jesús María, Chapalagana, Bolaños, Platanito y Florida. Adicionalmente se utilizó la información de lluvia registrada en la estación meteorológica automática la Michila, operada por la Comisión Nacional del Agua (CONAGUA). 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, siendo esta la salida de la cuenca.

Cuadro 1. Estaciones de registro horario de lluvia y caudal.

Estación	Latitud	Longitud	Operador
Chapalagana	21.9450000	-104.5080556	CFE
Bolaños	21.8250000	-103.7833333	CFE
La Florida	22.6861111	-103.6025000	CFE
Jesús María	22.2552778	-104.5161111	CFE
El Platanito	22.6113889	-104.0508333	CFE
La Michila	23.3875000	-104.2469444	CONAGUA

Nota: todas las estaciones de CFE registran lluvia y caudal cada hora; la estación CONAGUA registra lluvia cada 10 minutos.

3.3.3. Metodología

3.3.3.1. Modelo Autorregresivo con entrada exógena (ARX)

El modelo ARX que relaciona las entradas con las salidas del sistema mediante una ecuación lineal en diferencias con coeficientes constantes, es decir programado en OFF LINE (Hsu *et al.* 2009) se presenta en la ecuación 1.

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

Donde y_t es el vector que representa la variable de salida, en este caso el caudal de escurrimiento en la estación Chapalagana, r_t es el vector que representa la variable de entrada exógena, siendo esta 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 conocidos. Los índices na y nb indican el orden del modelo.

El modelo ARX se programó en Matlab® con apoyo del 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. Ljung (2017) indica que debido a que se utilizó el mismo conjunto de datos para la estimación y validación se deben de utilizar los criterios de Rissanen MDL e información de Akaike (AIC) para obtener las órdenes del modelo.

3.3.3.2. 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 $t-1$ añadiendo un término de corrección proporcional al error de predicción, este último es minimizado estadísticamente.

Welch y Bishop (2006) indican que el Filtro de Kalman discreto estima el estado $X \in \mathbb{R}_n$ 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 $z \in \mathbb{R}_m$, que es ecuación 3.

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

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, 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.

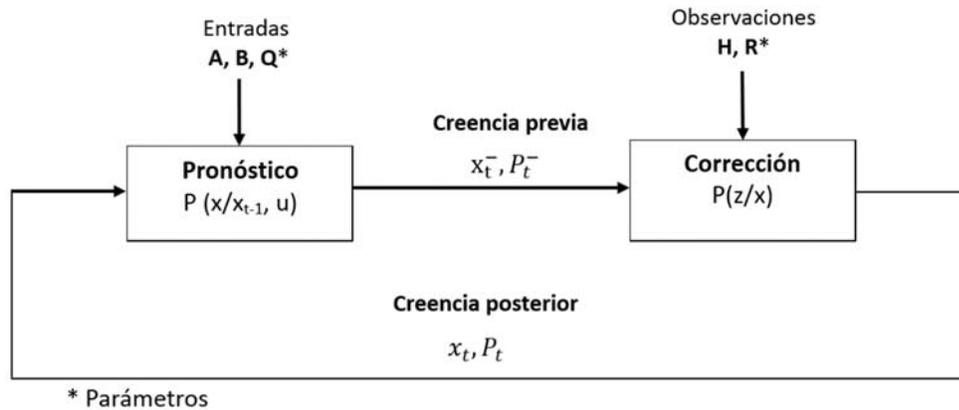


Figura 2. Etapas del modelo DKF.

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

Como se indicó anteriormente el filtro de Kalman opera en dos fases, las ecuaciones que se emplean en ambos tiempos se presentan en el algoritmo de la figura 3.

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, I es la matriz identidad.

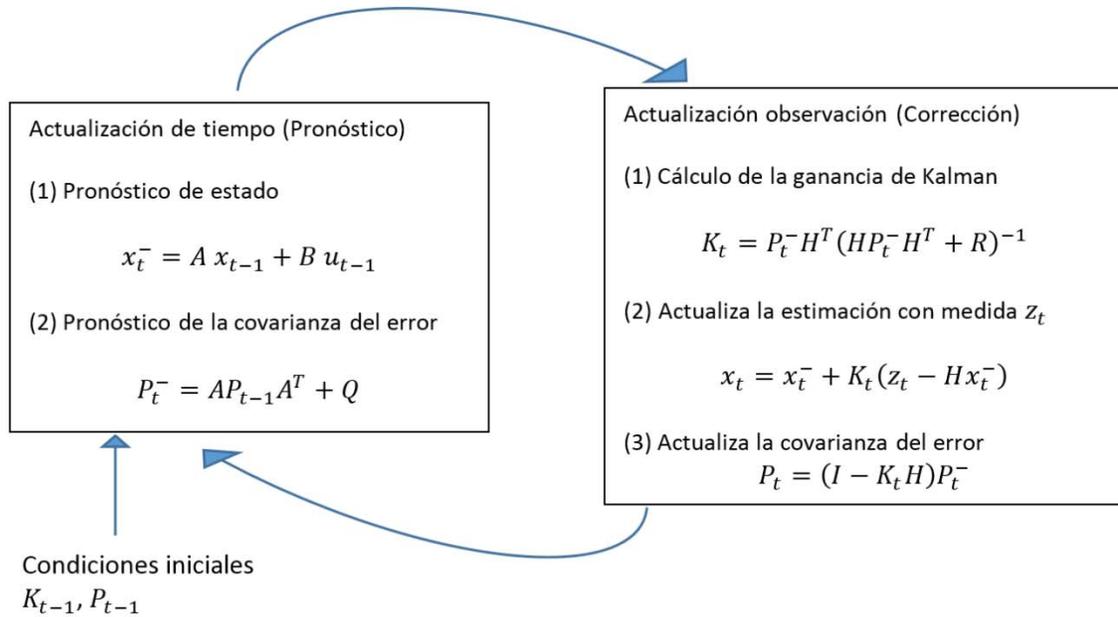


Figura 3. Algoritmo del filtro de Kalman Discreto (Welch y Bishop, 2006).

3.3.4. 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 siendo estos: 10%:90%, 25%:75% y 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 considerando los criterios de AIC y MDL en todos los casos.

El modelo en conjunto ARX+DKF utiliza inicialmente un periodo para la estimación de los parámetros a y b del modelo ARX, posteriormente 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 pudiera cambiar la función de respuesta de la misma, se planteó recalculer los parámetros a y b del modelo ARX cada cierto periodo, como se indica en la figura 4.

Para determinar las órdenes del modelo ARX más adecuado (n_a y n_b) se evaluaron en el modelo en conjunto ARX+DKF los periodos para estimación de parámetros; 10% (136 horas), 25%(340 horas) y 50% (680 horas) 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 horas, posteriormente se realizó la etapa de actualización para el caudal pronosticado en el tiempo $L=1$ y a partir de éste se pronosticó nuevamente seis pasos hacia adelante, como se muestra en la figura 5.

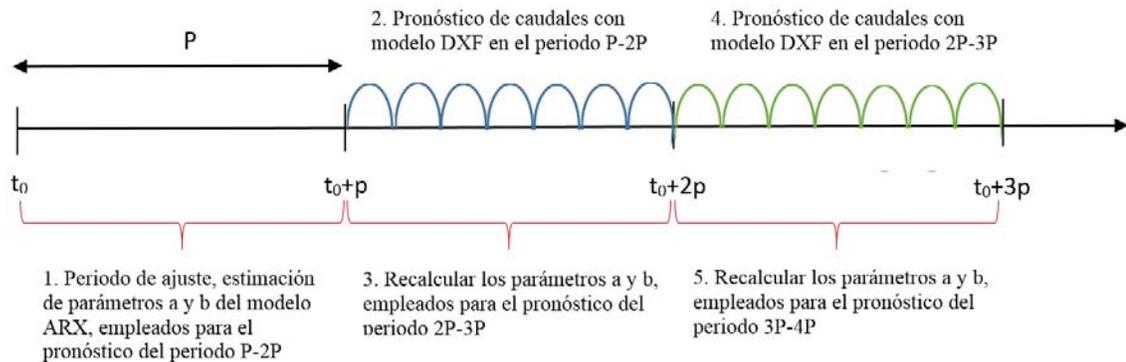
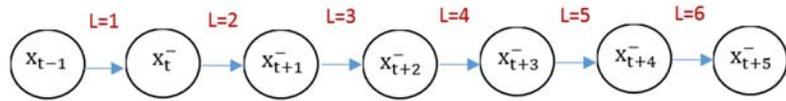
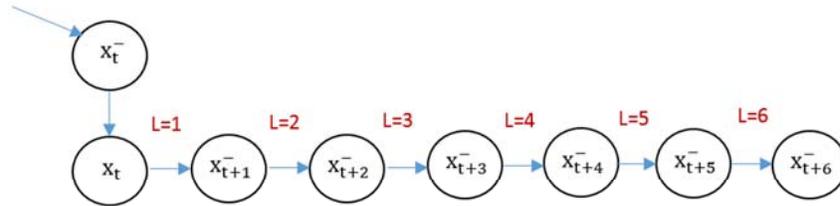


Figura 4. Proceso de estimación de parámetros y pronóstico de caudales.



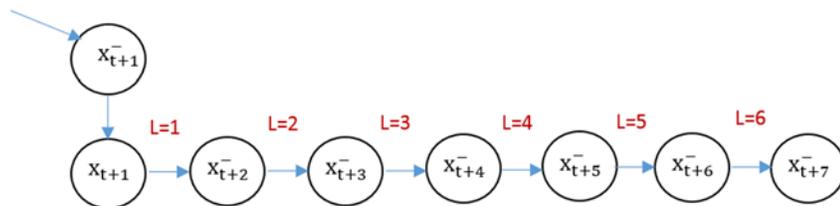
1. Primera etapa de pronóstico a partir del periodo de ajustes de parámetros a y b del modelo ARX

Actualización Zt



2. Se actualiza el pronóstico con L=1 y se vuelve a pronosticar con los 6 pasos hacia adelante

Actualización Zt



3. Se actualiza el pronóstico con L=1 y se vuelve a pronosticar con los 6 pasos hacia adelante

Figura 5. Pronóstico de caudales L pasos hacia adelante.

3.3.5. Análisis estadístico

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

3.4. Resultados y discusión

Se introdujeron 1360 datos, cada hora, para alimentar al modelo, los cuales reflejan las precipitaciones y las funciones de respuesta de varios eventos en la cuenca del río Huaynamota.



Figura 6. Datos de precipitación y caudales de escurrimiento introducidos al modelo ARX-DKF.

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 $L=1$.

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

Cuadro 2. Órdenes del modelo ARX.

Subconjunto	Horas para estimación de parámetros	<i>na</i>	<i>nb</i>	Nash- Sutcliffe	RMSE
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
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

na: número de retrasos de la variable caudal de escurrimiento.

nb: número de retrasos de la variable precipitación media de la cuenca.

AIC: criterio de información de Akaike.

MDL: criterio de Rissanen.

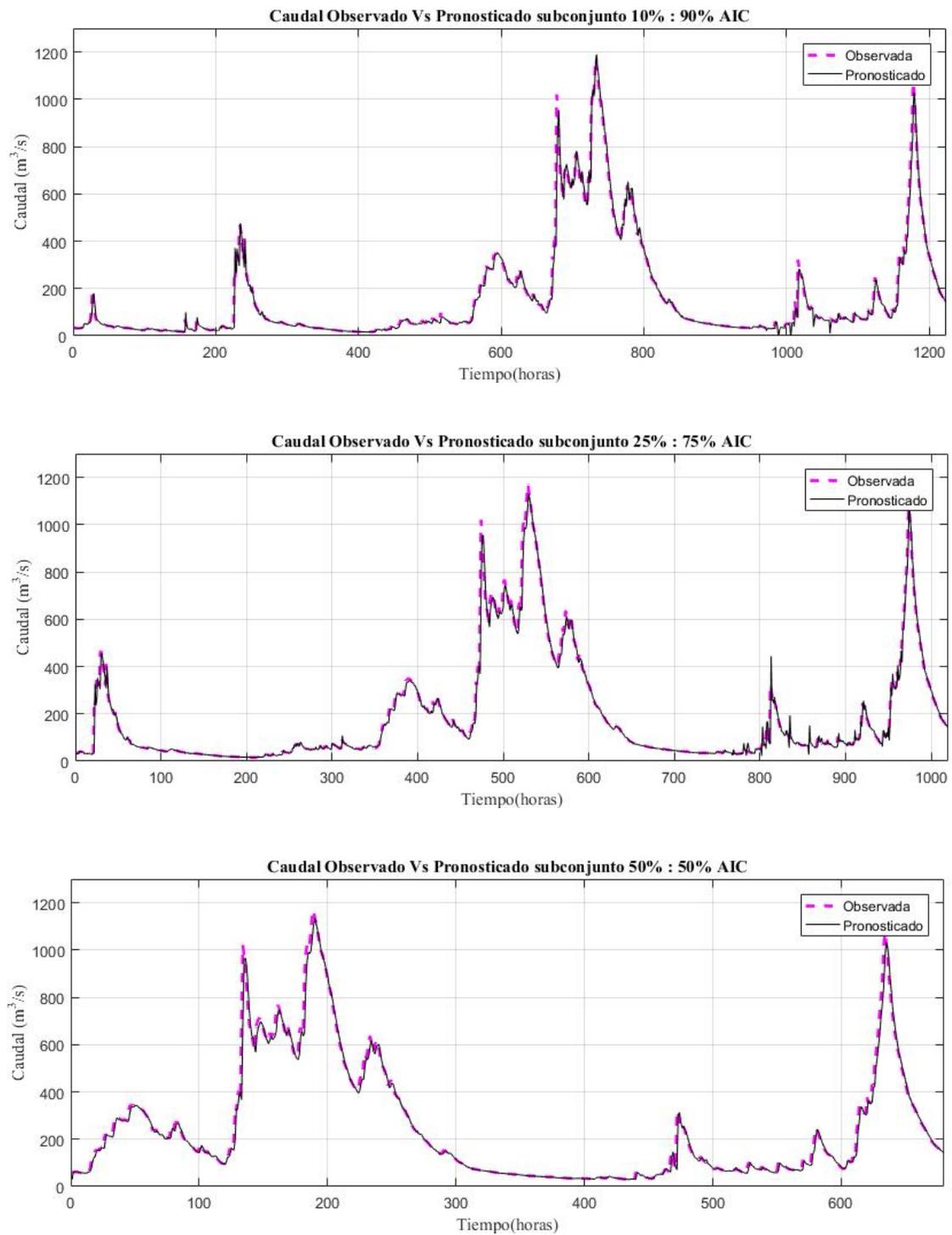


Figura 7. Caudales pronosticados vs observados para órdenes de modelo bajo el criterio AIC.

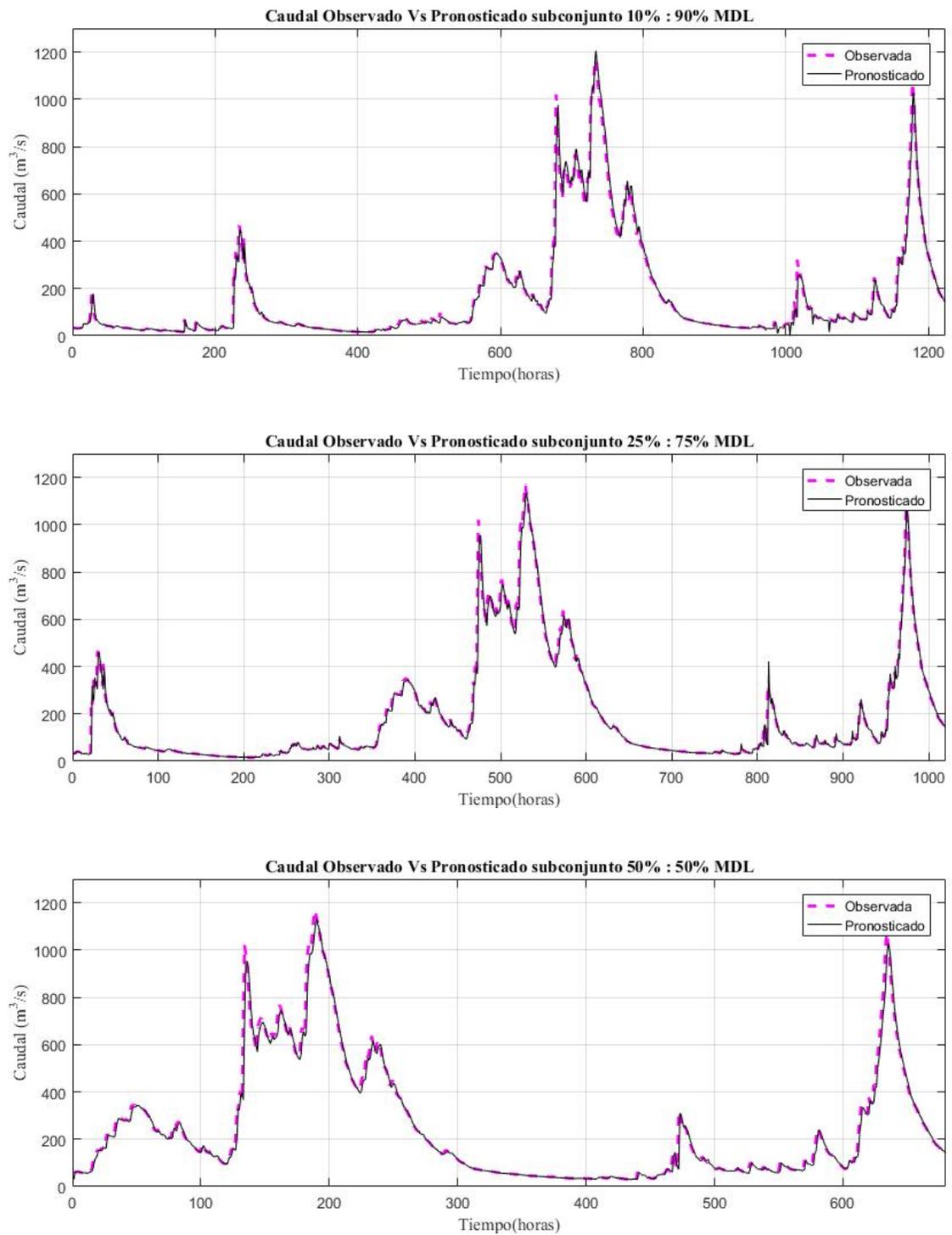


Figura 8. Caudales pronosticados vs observados para órdenes de modelo bajo el criterio MDL.

El índice de eficiencia Nash-Sutcliffe no es concluyente para evaluar los subconjuntos respecto a los dos criterios, sin embargo, considerando el análisis de las figuras 7 y 8, se aprecian mejores resultados con las órdenes del modelo del criterio MDL, además que estas órdenes resultan ser menores que las del criterio AIC. Las órdenes del modelo que arrojan mejores resultados son los del subconjunto 25%:75%, siendo que el valor de RMSE es menor. En este sentido y considerando los seis pasos hacia adelante en el pronóstico de caudales, el periodo en el que se recalcularon los parámetros fue de 338 horas y órdenes del modelo $n_a=2$ y $n_b=1$.

La forma general del hidrograma pronosticado tiene similitud respecto al hidrograma observado, con caudales pronosticados subestimados y desfasados L pasos hacia adelante, como se observa en la figura 9.

El caudal máximo observado en el periodo analizado fue de $1171 \text{ m}^3 \text{ s}^{-1}$ y ocurrió el 7 de septiembre de 2017 a las 13:00 pm, hubo un segundo valor máximo de $1070 \text{ m}^3 \text{ s}^{-1}$ el 26 de septiembre a las 2:00 am. La media y desviación estándar de los caudales observados son 190.27 y $231.78 \text{ m}^3 \text{ s}^{-1}$, respectivamente, para el caso de 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 de 0.99 , mientras que RMSE es de 28.37 . Respecto al pronóstico del primer caudal pico se presenta el 7 de septiembre del 2017 a las 14:00 pm 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 am, con un valor de $1087.61 \text{ m}^3 \text{ s}^{-1}$, la diferencia porcentual es menor de 3.1 en ambos casos, se aprecia que los pronósticos se desfasan un paso de tiempo hacia adelante respecto a los datos observados. Morales-Velázquez et al. (2014) determinaron un coeficiente de Nash-Sutcliffe para el pronóstico de caudales a corto plazo con DFK considerando el hidrograma unitario como función de respuesta de la cuenca de 0.98 , por lo que el modelo implementado en este trabajo mejoró sensiblemente este resultado.

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

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

Conforme se incrementan el pronóstico de caudales L pasos hacia adelante, el coeficiente de Nash-Sutcliffe decrece, mientras que RMSE se incrementa, lo que sugiere que la precisión en el pronóstico se va perdiendo y la covarianza de error de los caudales pronosticados se incrementa a cada paso de tiempo. Sin embargo, de acuerdo a los criterios de ajuste, estos resultan aceptables, siendo el coeficiente de Nash-Sutcliffe para seis pasos hacia adelante L=6 de 0.85. La media y desviación estándar de las caudales observados son 190.93 y 231.51 $\text{m}^3 \text{s}^{-1}$, para el caso de los caudales pronosticados seis pasos hacia adelante L=6 son 173.00 y 210.60 $\text{m}^3 \text{s}^{-1}$, respectivamente. El pronóstico del primer caudal pico se presenta el 07 de septiembre a las 19:00 pm, con un valor de 1042.81 $\text{m}^3 \text{s}^{-1}$, subestimando 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 am, con un valor de 1053.59 $\text{m}^3 \text{s}^{-1}$ subestimando el caudal con una diferencia porcentual del 1.53, el pronóstico se desfasa 6 horas hacia adelante respecto al caudal pico observado.

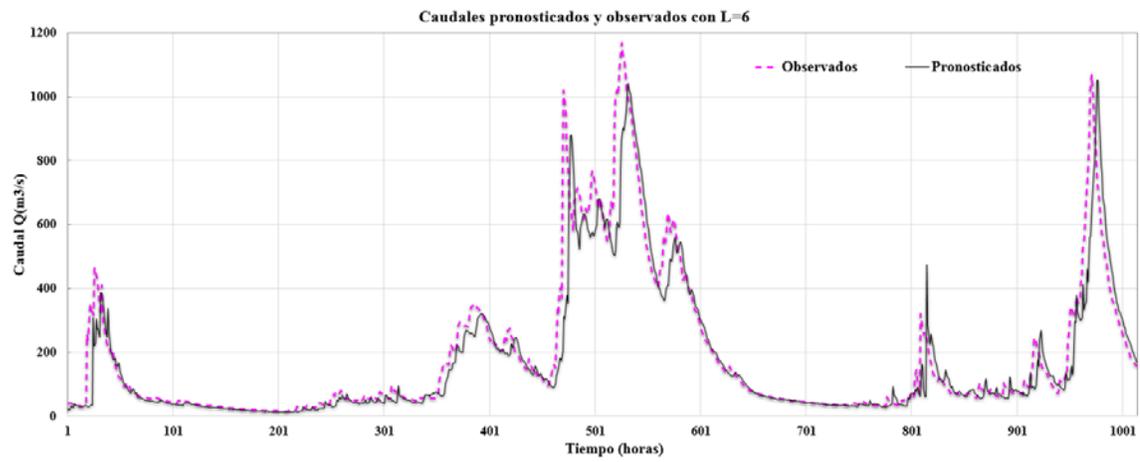
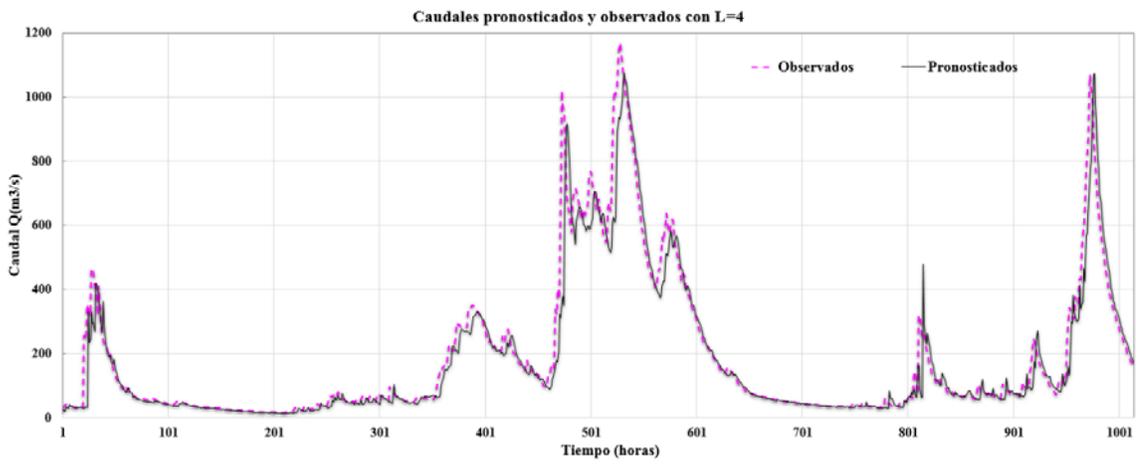
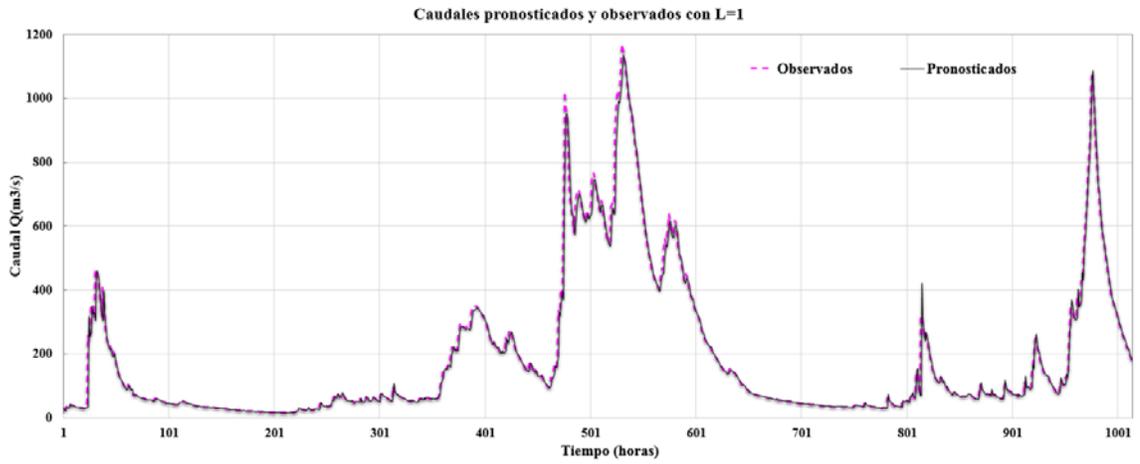


Figura 9. Caudales pronosticados vs observados con el modelo ARX-DXF para L=1, 4 y 6 horas hacia adelante.

González-Leiva et al. (2015) para el pronóstico de caudales para un modelo ARX+DFK considerando información promedio diaria (24 h), obtuvo para una orden de modelo $n_a=1$, $n_b=2$, 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. Respecto a lo obtenido en este estudio se mejoraron considerablemente los resultados, esto es debido a que al ser información horaria la función de respuesta de la cuenca se representa de mejor manera.

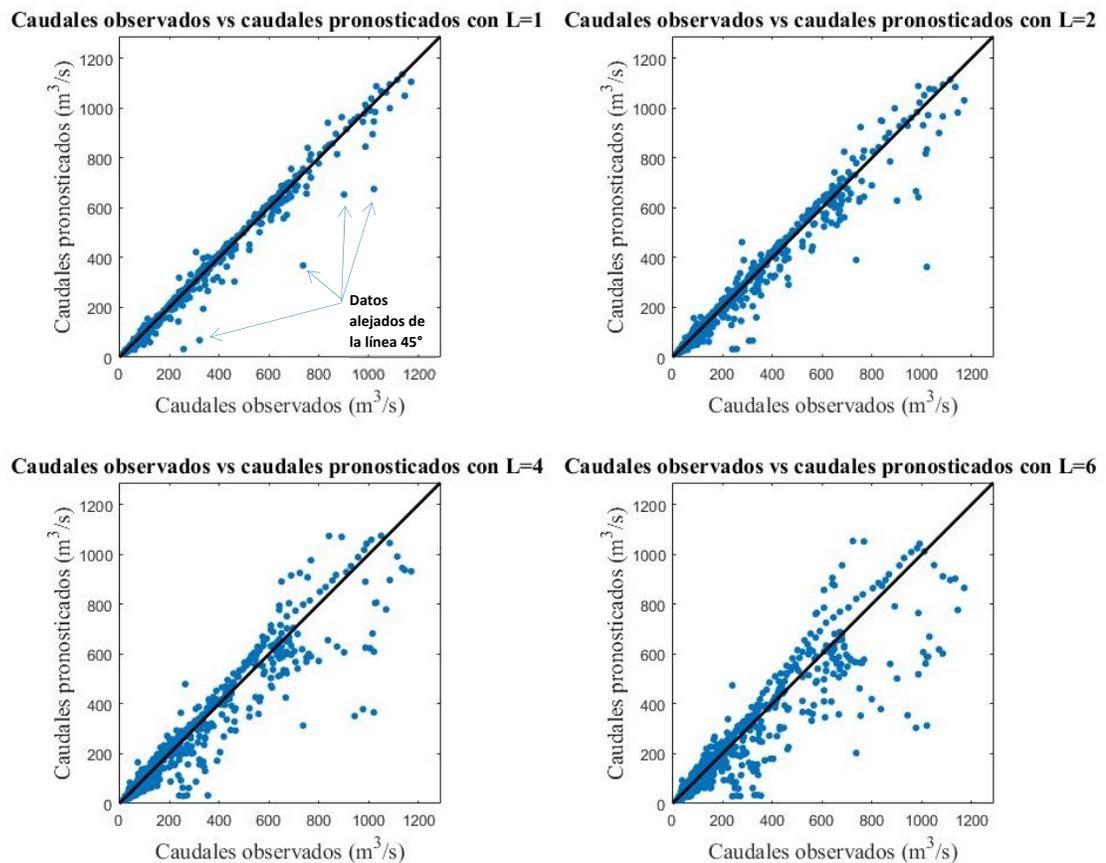


Figura 10. Diagramas de dispersión para caudales pronosticados vs observados.

La gráfica de dispersión de la figura 10, para un paso hacia adelante $L=1$ muestra por lo menos 4 valores de caudales pronosticados subestimados ya que se alejan de la línea a 45° , este error se presenta en donde el caudal se incremente de manera significativa de un paso de tiempo a otro como se muestra en la figura 11. En el resto de las gráficas de dispersión los valores se van dispersando más, ya que como se comentó anteriormente el hidrograma pronosticado se desfasa del observado provocando residuales grandes.

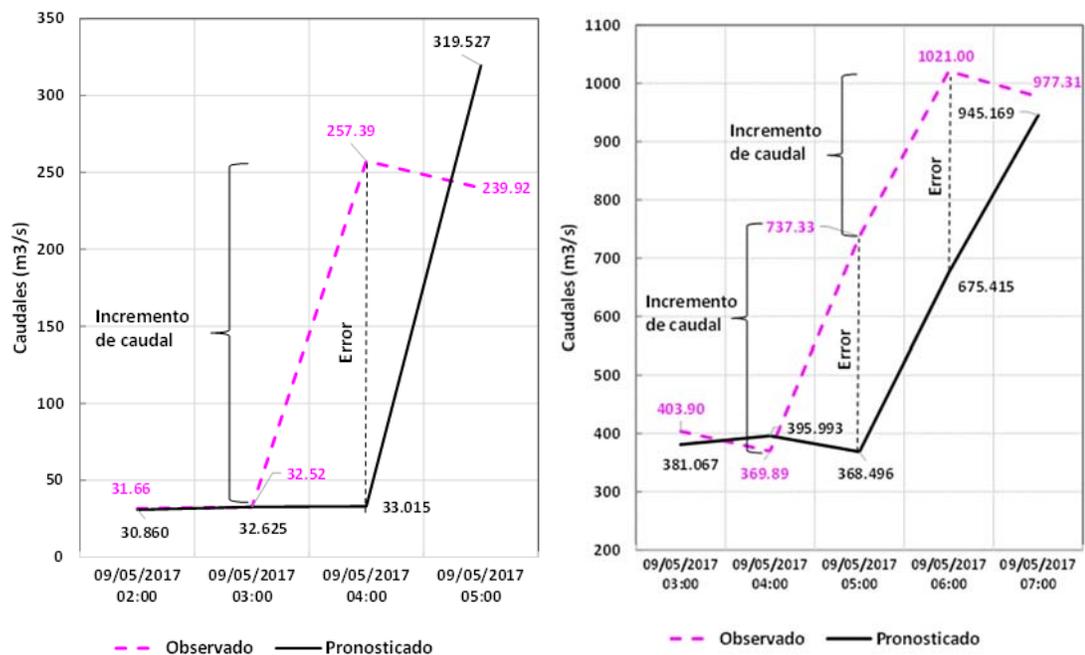


Figura 11. Datos alejados de la línea a 45° para $L=1$.

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

Cuadro 4. Parámetros de la línea de regresión.

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

Es de importancia el análisis del coeficiente de regresión β_1 , para el pronóstico de caudales un paso hacia adelante L=1 el valor $\beta_1 = 0.9754$, lo que indica que el pronóstico subestima ligeramente los caudales, conforme se incrementan los L pasos hacia adelante para el pronóstico de caudales el coeficiente de regresión disminuye, siendo para seis pasos hacia adelante L=6 el valor de $\beta_1 = 0.8426$, como los reflejan las gráficas de dispersión presentadas en la figura 10.

Las gráficas de residuales presentadas en la figura 12, muestran una variabilidad creciente a medida que se incrementan los caudales, sobre todo cuando el pronóstico L pasos hacia adelante es mayor, en este caso para el pronóstico L=3, 4, 5, 6 horas 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 resultan ser mayores en los picos.

La alta variabilidad se presenta en todo el rango de caudales, sin embargo, de acuerdo a las franjas con un intervalo de confianza del 95%, se aprecia mayormente a partir de los $760 \text{ m}^3 \text{ s}^{-1}$.

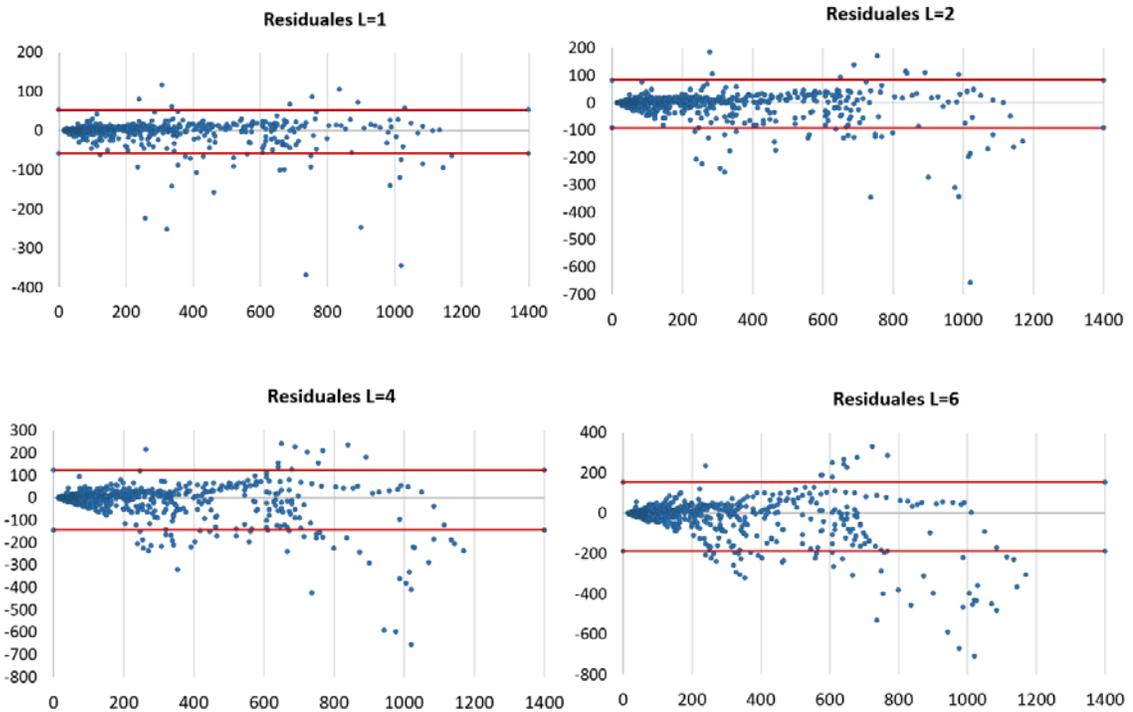


Figura 102. Graficas de residuales con franjas para un nivel de confianza de 95%.

3.5. 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.

Los coeficientes de eficiencia Nash-Sutcliffe para los resultados del modelo conjunto ARX-DXF muestra resultados excelentes para todos los pasos de tiempo de anticipación $L=1, 2, 4,$ y 6 , la condición más crítica para el pronóstico de caudales seis pasos hacia adelante $L=6$ arroja un valor de $E=0.85$, lo que indica que el pronóstico de caudales considerando datos horarios arroja resultados satisfactorios que pudieran emplearse para sistema de alerta temprana.

A pesar del desfase de los caudales picos pronosticados, no se pierde la utilidad práctica de estos valores, sobre todo considerando que el tiempo de concentración de la cuenca Huaynamota es de 44 horas.

Debido a la falta de información hidroclimatológica existente en el país, si se quisiera implementar este tipo de modelos para pronóstico de caudales, es necesario realizar una planeación para la instrumentación de las cuencas correspondientes, especialmente cuando queriendo mejorar el pronóstico el modelo se alimenta con información horaria.

3.6. 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.

3.7. Literatura citada

Beven, K. 1989. Changing ideas in hydrology-the case of physically-based models. *Journal of Hydrology*, 105: 157-172.

Castañeda C., J. A., M.A. Nieto A. y V. A. Ortiz B. 2013. Analysis and application of the Kalman filter to a signal with random noise. *Scientia en Technica*, 18(1): 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.iiie.org.mx/cuencas/logon.aspx?ReturnUrl=%2fcuencas%2fdefault.aspx>. (Fecha de consulta: Septiembre 28 del 2017).

Drécourt, J. Kalman filtering in hydrological modeling. DAIHM Technical Report 2003-1, DHI Water & Environment, 2003.

Gayathri K., D., B.P. Ganasri y G.S. Dwarakish. 2015. A Review on Hydrological Models. *Aquatic Procedia* 4: 1001-1007.

- Grewal M., S. and Andrews A. P. 2008. Kalman filtering theory and practice using MATLAB, 2nd edn. Wiley, New York. 401 p.
- González-Leiva, F., Ibáñez-Castillo, L. A., Valdés, J. B., Vázquez-Peña, M. A., & Ruiz-García, A. 2015. Pronóstico de caudales con Filtro de Kalman Discreto en el río Turbio. *Tecnología y Ciencias del Agua* 6(4): 5-24.
- Hernández P., Y. y H. Medina G. 2012. Estimación de la humedad del suelo mediante técnicas de asimilación de datos. *Revista Ciencias Técnicas Agropecuarias* 21(4): 30-35.
- Hsu, K.-L., H. Moradkhani and S. Sorooshian. 2009. A sequential Bayesian approach for hydrologic model selection and prediction, *Water Resources Research*, 45, W00B12, DOI:10.1029/2008WR006824.
- Hobson, K. A. y K. J. Kardynal. 2015. Western Veeries use an eastern shortest-distance pathway: New insights to migration routes and phenology using light-level geolocators. *The American Ornithologists' Union* 132(3): 540-550.
- Krause, P., D.P. Boyle y F. Base. 2005. Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences* 5: 89-97.
- Ljung, L. 1988-2017. 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. *Investigaciones Geográficas* 47: 36-58.
- Morales-Velázquez. M. I., J. Aparicio, y J. B. Valdés. 2014. Pronóstico de avenidas utilizando el Filtro de Kalman Discreto. *Tecnología y Ciencias del Agua* 5(2): 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. *Revista de Ingeniería* 38: 20-26.
- Semenova, O and K. Beven 2015. Barriers to progress in distributed hydrological modelling. *Hydrological Processes* 29: 2074-2078. DOI: 10.1002/hyp.10434.
- Servicio Meteorológico Nacional. 2017. Datos Históricos de Estaciones Meteorológicas Automáticas operadas por el SMN. (Fecha de consulta: Septiembre 10 del 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., R. Arteaga-Rámirez. 2015. Development, classification and trends in rainfall-runoff modeling. *Ingeniería Agrícola y Biosistemas* 7(1): 5-21. doi: 10.5154/r.inagbi.2015.03.002.
- Vargas-Castañeda. G., L. A. Ibáñez-Castillo., R. Arteaga-Rámirez y G. Arévalo-Galarza. 2018. Kinematic wave hydrologic model of the Turbio River basin, Guanajuato, Mexico. *Ingeniería Agrícola y Biosistemas* 10(1): 33–47. doi: 10.5154/r.inagbi.2017.07.012.
- Valdés, J. B., J. Mejía V., y 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. 37p.
- Zhen-lei, W. X. Yue-Ping, S. Hong-yue, W. Xie, 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.

CAPÍTULO 4. CONCLUSIONES Y RECOMENDACIONES

El modelo implementado en este estudio a partir de la técnica recursiva del filtro de Kalman discreto en conjunto con el modelo paramétrico ARX mostró tener buenos resultados en el pronóstico de caudales horarios en la cuenca del río Huaynamota, siendo que la estructura del modelo permite considerar la variabilidad climática a lo largo del tiempo así como los posibles cambios en las condiciones físicas de la cuenca que pudieran modificar su función de respuesta a eventos de precipitación. Los resultados fueron satisfactorios al utilizar información horaria tanto de precipitación y caudales de escurrimiento de la cuenca, esta información permite describir con detalle la variable de entrada a la cuenca que genere los escurrimientos superficiales reflejados en el eje de las ordenadas del hidrograma a la salida de la cuenca, comparado con los resultados de estudios previos, los resultados de este estudio sugieren que a un Δt más fino de datos de entrada y salida del modelo los pronósticos realizados tienen un mejor ajuste con los datos observados, sin embargo, por practicidad en el manejo de la información y considerando los resultados de las medidas de bondad de ajuste con un $\Delta t=1$ hr este es un buen periodo de tiempo de análisis.

Este tipo de modelos podría implementarse en cuencas con características y tamaños diversos, sin embargo, su mayor inconveniente es la obtención de los datos de entrada al modelo, en el país son pocas las instituciones que generan información con la calidad y distribución espacial requerida como es el caso de Comisión Federal de Electricidad que solo monitorea cuencas con infraestructura de centrales hidroeléctricas. Por lo que de querer implementar este modelo en cuencas en donde haya la necesidad de pronosticar caudales de escurrimiento por cuestiones de operación de infraestructura hidráulica o por seguridad de centros de población, sería necesario diseñar y establecer una red de monitoreo administrada por la instancia de interés.

CAPÍTULO 5. ANEXOS

5.1. Cálculo de la precipitación media de la cuenca

Para determinar la precipitación promedio en la zona de estudio se utilizó el método de polígonos de Thiessen, los factores de ponderación así como las áreas correspondientes a cada estación meteorológica se muestran en la figura 1 y cuadro 1.

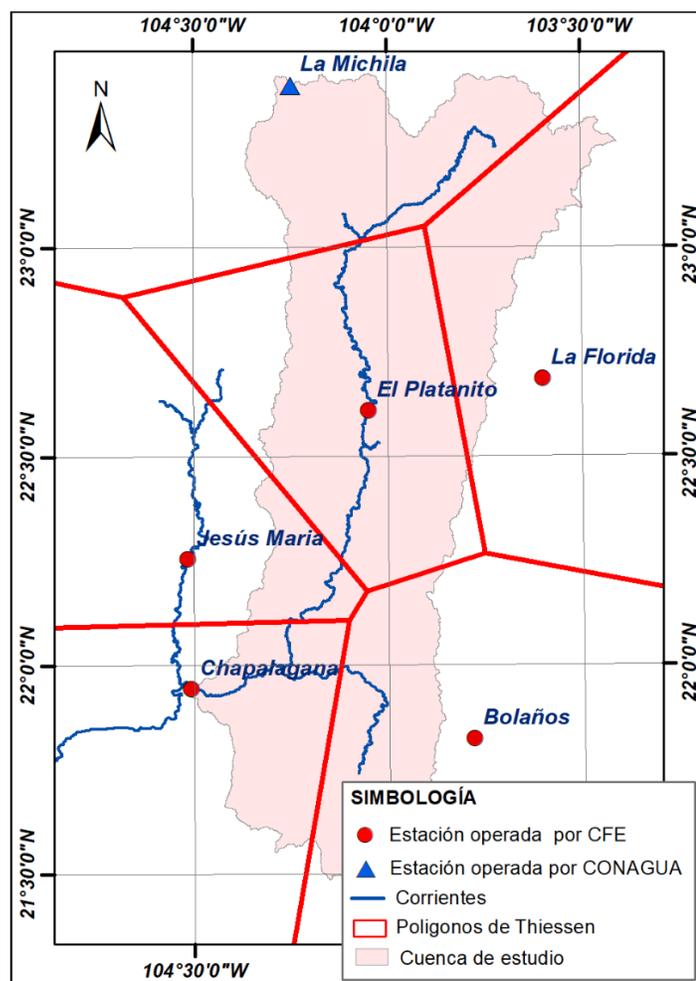


Figura 1. Polígonos de Thiessen en la cuenca Chapalagana.

Cuadro 1. Tabla de ponderación de precipitación.

Estación	Área (km²)	Factor de ponderación
Chapalagana	1,571.885	0.130
Bolaños	2,074.683	0.172
Florida	1,604.676	0.133
Jesús María	655.752	0.054
Platanito	3,778.588	0.313
La Michila	2,393.948	0.198
Total	12,079.532	1.000

La precipitación promedio introducida al modelo de simulación se muestra en la figura 2, los datos incluyen varios eventos de precipitación que puedan definir la respuesta de la cuenca y que se reflejen en los caudales medidos en la estación Chapalagana.

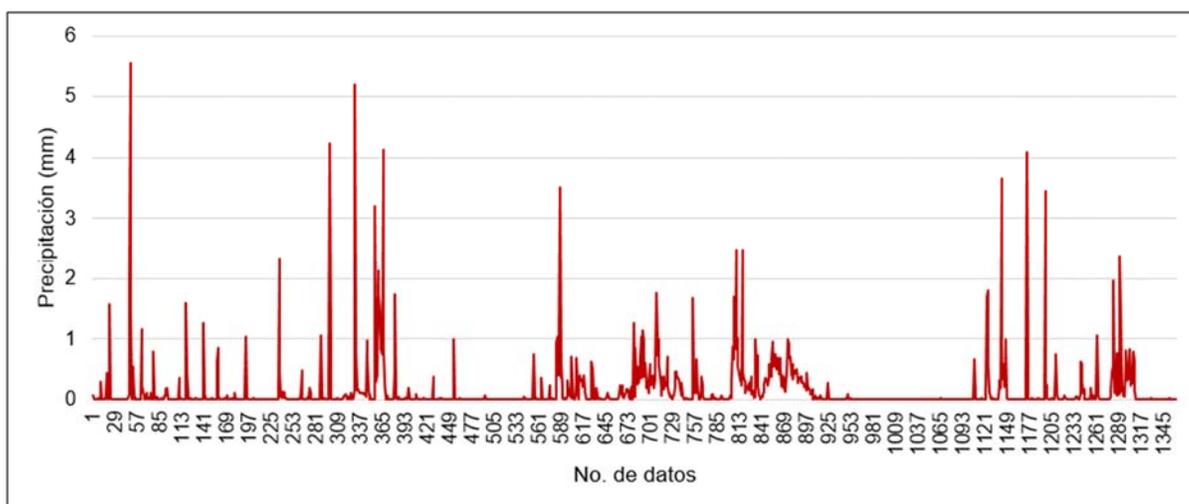


Figura 2. Precipitación promedio en la cuenca de estudio.

5.2. Análisis estadístico de los resultados

Cuadro 2. Estadísticas de la regresión simple para L=1.

Estadística	Valor
Coefficiente de correlación múltiple	0.992612847
Coefficiente de determinación R ²	0.985280264
R ² ajustado	0.985265718
Error típico	27.64506257
Observaciones	1014

Cuadro 3. Tabla ANOVA para L=1.

	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	Valor crítico de F
Regresión	1	51769672.3	51769672.3	67739.23092	0
Residuos	1012	773420.4782	764.2494844		
Total	1013	52543092.78			

Cuadro 4. Estadísticas de la regresión simple para L=2.

Estadística	Valor
Coeficiente de correlación múltiple	0.981623055
Coeficiente de determinación R ²	0.963583822
R ² ajustado	0.963547837
Error típico	42.84051557
Observaciones	1014

Cuadro 5. Tabla ANOVA para L=2.

	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	Valor crítico de F
Regresión	1	49145643.08	49145643.08	26777.84632	0
Residuos	1012	1857333.492	1835.309774		
Total	1013	51002976.58			

Cuadro 6. Estadísticas de la regresión simple para L=3.

Estadística	Valor
Coeficiente de correlación múltiple	0.96938295
Coeficiente de determinación R ²	0.939703304
R ² ajustado	0.939643722
Error típico	54.27417638
Observaciones	1014

Cuadro 7. Tabla ANOVA para L=3.

	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	Valor crítico de F
Regresión	1	46458398.55	46458398.55	15771.67256	0
Residuos	1012	2981034.456	2945.686221		
Total	1013	49439433.01			

Cuadro 8. Estadísticas de la regresión simple para L=4.

Estadística	Valor
Coeficiente de correlación múltiple	0.956141258
Coeficiente de determinación R ²	0.914206105
R ² ajustado	0.914121328
Error típico	63.7227192
Observaciones	1014

Cuadro 9. Tabla ANOVA para L=4.

	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	Valor crítico de F
Regresión	1	43788174.69	43788174.69	10783.71105	0
Residuos	1012	4109311.961	4060.584942		
Total	1013	47897486.65			

Cuadro 10. Estadísticas de la regresión simple para L=5.

Estadística	Valor
Coeficiente de correlación múltiple	0.941964528
Coeficiente de determinación R ²	0.887297173
R ² ajustado	0.887185806
Error típico	71.87592522
Observaciones	1014

Cuadro 11. Tabla ANOVA para L=5.

	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	Valor crítico de F
Regresión	1	41160599.9	41160599.9	7967.366578	0
Residuos	1012	5228142.41	5166.148626		
Total	1013	46388742.31			

Cuadro 12. Estadísticas de la regresión simple para L=6.

Estadística	Valor
Coeficiente de correlación múltiple	0.926262932
Coeficiente de determinación R ²	0.857963019
R ² ajustado	0.857822666
Error típico	79.40829876
Observaciones	1014

Cuadro 13. Tabla ANOVA para L=6.

	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	Valor crítico de F
Regresión	1	38546010.28	38546010.28	6112.905037	0
Residuos	1012	6381346.047	6305.677912		
Total	1013	44927356.32			

5.3. Forma matricial del filtro de Kalman discreto

Modelo lineal en espacio de estados que representa al proceso en el sistema:

$$x_t = Ax_{t-1} + BU_t + w_t$$

$$\begin{bmatrix} x(t) \\ x_{(t-1)} \\ x_{(t-2)} \\ \vdots \\ x_{(t-na+2)} \\ x_{(t-na+1)} \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 & \cdots & a_{na-2} & a_{na-1} & a_{na} \\ 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_{(t-1)} \\ x_{(t-2)} \\ x_{(t-3)} \\ \vdots \\ x_{(t-na+1)} \\ x_{(t-na)} \end{bmatrix}$$

$$+ \begin{bmatrix} b_1 & b_2 & b_3 & \cdots & b_{nb-2} & b_{nb-1} & b_{nb} \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} u_{(t-1)} \\ u_{(t-2)} \\ u_{(t-3)} \\ \vdots \\ u_{(t-nb+2)} \\ u_{(t-nb+1)} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} w$$

$$z_t = Hx_t + v_t$$

$$z_{(t)} = [1 \quad 0 \quad 0 \quad \cdots \quad 0 \quad 0] \begin{bmatrix} x_{(t)} \\ x_{(t-1)} \\ x_{(t-2)} \\ \vdots \\ x_{(t-na+2)} \\ x_{(t-na+1)} \end{bmatrix} + v_{(t)}$$