**UNIVERSIDAD AUTÓNOMA CHAPINGO**

**POSGRADO EN INGENIERÍA AGRÍCOLA Y USO INTEGRAL DEL AGUA**

# PRONÓSTICO DE CAUDALES HORARIOS MEDIANTE EL FILTRO DE KALMAN DISCRETO Y FILTRO DE KALMAN DE CONJUNTOS EN LA CUENCA DEL RÍO HUAYNAMOTA, MEXICO

## T E S I S

COMO REQUISITO PARCIAL PARA OBTENER EL GRADO DE

**DOCTOR EN INGENIERÍA AGRÍCOLA Y USO INTEGRAL DEL AGUA**

PRESENTA:

**ILDEFONSO NARVÁEZ ORTIZ**

BAJO LA SUPERVISIÓN DE:

**DRA. LAURA ALICIA IBÁÑEZ CASTILLO**

APROBADA

Chapingo, Estado de México, 09 de junio de 2021

**PRONÓSTICO DE CAUDALES HORARIOS MEDIANTE EL FILTRO DE KALMAN DISCRETO Y FILTRO DE KALMAN DE CONJUNTOS EN LA CUENCA DEL RÍO HUAYNAMOTA, MEXICO**

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

## DOCTOR EN INGENIERÍA AGRÍCOLA Y USO INTEGRAL DEL AGUA

**DIRECTORA:**

Dra. Laura Alicia Ibáñez Castillo

**ASESOR:**

Dr. Mario Alberto Vázquez Peña

**ASESOR:**

Dr. Ramón Arteaga Ramírez

**LECTOR EXTERNO**

Dra. Rocío Cervantes Osornio
Lectora Externa

## CONTENIDO

**1 INTRODUCCIÓN GENERAL ..... 1**
1.1 Literatura citada ..... 3
**2 REVISION DE LITERATURA ..... 5**
2.1 Variabilidad climática y pronóstico de caudales ..... 5
2.2 Relaciones lluvia - escurrimiento ..... 6
2.3 Modelación y predicción de procesos hidrológicos ..... 7
2.4 Incertidumbre de la predicción hidrológica ..... 7
2.5 Variables y función de distribución ..... 8
*2.5.1 Variables aleatorias ..... 8*
*2.5.2 Función de distribución de probabilidades ..... 9*
2.6 Filtro de Kalman ..... 10
2.7 Asimilación de datos ..... 12
2.8 Recursividad ..... 12
2.9 Filtro de Kalman discreto (Lineal) ..... 13
2.10 Filtro de Kalman no lineal ..... 15
*2.10.1 Filtro de Kalman de conjuntos ..... 15*
*2.10.2 Filtro de partículas ..... 18*
*2.10.3 Filtro de Kalman extendido ..... 18*
*2.10.4 Filtro de Kalman "unscented" ..... 19*
2.11 Asimilación de varios sensores ..... 19
2.12 Modelos autorregresivos ..... 20
2.13 Redes neuronales ..... 20
2.14 Filtro de Kalman e hidrología ..... 22
*2.14.1 Cuenca hidrográfica ..... 22*
*2.14.2 Aplicaciones del Filtro del Kalman ..... 23*
2.15 Rutinas de código ..... 25

2.16 Literatura citada ..... 25
**3 COMPARACIÓN DEL FILTRO DE KALMAN DISCRETO CON EL FILTRO DE KALMAN DE CONJUNTOS PARA PRONÓSTICO DE CAUDALES HORARIOS EN EL RÍO HUAYNAMOTA, NAYARIT, MÉXICO ..... 33**
3.1 Resumen ..... 33
3.2 Abstract ..... 34
3.3 Introducción ..... 34
3.4 Materiales y métodos ..... 35
*3.4.1 Filtro de Kalman Discreto ..... 38*
*3.4.2 Filtro de Kalman de Conjuntos ..... 40*
*3.4.3 Filtro de Kalman discreto y filtro de Kalman de Conjuntos ..... 41*
*3.4.4 Modelo autorregresivo de primer orden (ARX(1,1)) y DKF ..... 42*
3.5 Resultados y Discusión ..... 42
3.6 Conclusiones ..... 53
3.7 Agradecimientos ..... 54
3.8 Referencias ..... 54
**4 ENSEMBLE KALMAN FILTER (ENKF) FOR HOURLY STREAMFLOW FORECASTING IN HUAYNAMOTA RIVER, NAYARIT, MÉXICO ..... 57**
4.1 Abstract ..... 57
4.2 Resumen ..... 58
4.3 Introduction ..... 58
4.4 Materials and Methods ..... 60
4.5 Ensemble Kalman Filter ..... 62
4.6 Discrete Kalman Filter ..... 64
4.7 First-order autoregressive model ..... 65
4.8 Results and discussion ..... 65
4.9 Conclusions ..... 71

4.10 Acknowledgements ..... 72
4.11 References ..... 72
**5 CONCLUSIONES GENERALES ..... 76**
**6 LITERATURA CITADA GENERAL ..... 78**

# LISTA DE CUADROS

## CAPITULO 3

Tabla 3.1. Estadísticos para la aplicación de DKF, EnKF, DKF-EnKF y ARX-DKF. ..... 44
Tabla 3.2. Coeficientes de determinación de las series pronosticadas. ..... 50
Tabla 3.3. Cantidad de valores atípicos por algoritmo y paso. ..... 53

## CAPITULO 4

Table 4.1. Summary of statistics for application of EnKF, DKF, and ARX ..... 66

# LISTA DE FIGURAS

## CAPITULO 2

Figura 2.1. Algoritmo de filtro de Kalman discreto (Welch & Bishop, 2006) ..... 14
Figura 2.2. Algoritmo de EnKF (Gillijns et al., 2006)..... 17

## CAPITULO 3

Figura 3.1. Localización del área de estudio. ..... 36
Figura 3.2. Algoritmo del filtro de Kalman discreto (Welch & Bishop, 2006). ..... 38
Figura 3.3. Algoritmo del filtro de Kalman de conjuntos (Gillijns et al., 2006)..... 40
Figura 3.4. Raíz del cuadrado medio del error con diferentes conjuntos. ..... 43
Figura 3.5. Pronósticos con DKF, ENKF y DKF-EnKF (Avenida del 4/9/2017 16:00 h al 8/9/2017 20:00 h) ..... 46
Figura 3.6. Pronósticos con DKF, EnKF y DKF-EnKF (Avenida del 23/09/17 10:00 h al 28/09/17 00:00 h) ..... 47
Figura 3.7. Caudales observados versus pronosticados. ..... 49
Figura 3.8. Histograma de residuales del DKF. ..... 51
Figura 3.9. Residuales estandarizados. ..... 52

## CAPITULO 4

Figure 4.1. Location of the study area. ..... 60
Figure 4.2. Algorithm of the ensemble Kalman filter. ..... 63
Figure 4.3. Discrete Kalman Filter. ..... 64
Figure 4.4. Root mean square error with different set sizes. ..... 66

Figure 4.5. Observed flow and flows predicted with EnKF, DKF and ARX (Flood of 4/9/2017 16:00 h to 8/9/2017 20:00 h). ..... 68
Figure 4.6. Observed flow and flows predicted with EnKF, DKF and ARX (Flood from 23/09/17 10:00 h al 28/09/17 00:00 h)..... 69
Figure 4.7. Observed versus predicted flows. ..... 70
Figure 4.8. Histographs of residuals..... 71

# ABREVIATURAS

| | |
|---|---|
| AR | Modelo Autorregresivo |
| ARX | Modelo Autorregresivo con variable exógena |
| CFE | Comisión Federal de Electricidad |
| CONAGUA | Comisión Nacional del Agua |
| DKF | Filtro de Kalman Discreto |
| EKF | Filtro de Kalman Extendido |
| EnKF | Filtro de Kalman de Conjuntos |
| HUI | Hidrograma Unitario Instantáneo |
| IPCC | Panel Intergubernamental sobre el Cambio Climático |
| PF | Filtro de Partículas |
| SMN | Servicio Meteorológico Nacional |
| UKF | Filtro de Kalman “Unscented” |

# DEDICATORIA

A mis padres Manuel Narvaez y Raquel Ortiz y mis hermanos Isabel, Herman y María Narvaez Ortiz por su constante apoyo y el aguantarme todos mis caprichos y tareas, quienes hemos comprendido y apoyado que nuestras metas apuntan al bienestar de nuestra familia.

A mis familiares en Colombia que siempre han creído en mí, y que han sido un soporte para mantener la estabilidad y continuo avance en la pequeña finca cafetera de mi familia.

A los lugares que me han dado mucho, La Unión Nariño que me vio nacer y Yopal Casanare que es mi segundo hogar.

A México por darme la oportunidad de conocer y aprender.

# AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología (CONACyT), por el financiamiento otorgado para formación doctoral, a la Universidad Autónoma Chapingo y al Posgrado IAUIA por darme la oportunidad de culminar una de las metas más anheladas en mi vida. A Unitrópico por fomentar y apoyar la formación a nivel superior.

A Pablo Alan Calderón, Luciana Vega y Max por darme apoyo y la confianza para entrar a su casa durante la fase más difícil de contingencia por el COVID19. Asimismo, a Natali Suárez (Nata), Carolina Romero, Daniela Alvarado y Karen Cáceres por aguantarme “tanto tiempo” en Zacatecas.

A mi familia; Manuel, Raquel, Isabel, Herman y María, quienes a pesar de la distancia que ha estado presente desde ya hace muchos años siempre han sido un apoyo incondicional.

A los doctores Ramon Arteaga Ramírez y Mario Vázquez Peña por su permanente y eficiente asesoría. Al Doctor Carlos Cintora por sus enseñanzas y aclaraciones sobre diversos temas de la estadística.

A la Doctora Laura Alicia Ibáñez Castillo quien desde el inicio fue mi directora de trabajo de investigación y de forma permanente, eficiente y eficaz siguió de cerca no solo el componente investigativo sino también la administración y demás asuntos inherentes a este proceso de formación avanzada.

Al Dr. Mauricio Carrillo García y Mayra, por su colaboración en la coordinación del posgrado. A los profesores que hicieron parte de mi formación académica durante el doctorado. A todos mis compañeros de estudio en especial a Cesar Vázquez, Jorge Jaimes y Natalia Valencia.

# DATOS BIOGRÁFICOS

Ildefonso Narvaez Ortiz, de nacionalidad colombiana, nació en La Unión Nariño, Colombia, el 6 de agosto de 1984. Realizo sus estudios de primaria en la Escuela Nueva La Cañada, y cursó la secundaria en el Instituto Nocturno de Bachillerato Comercial Juanambú en La Unión Nariño. Entre 2004 y 2010 en la Universidad de Nariño realizó estudios de Ingeniería Agroforestal, y en 2012 obtuvo el título como Master en Practica del Desarrollo (Magister en Desarrollo Rural) en el Centro Agronómico Tropical de Investigación y Enseñanza – CATIE en Costa Rica. Se desempeñó como coordinador de la Unidad Agroambiental en la alcaldía del Municipio de la Unión Nariño, Docente invitado en el programa de Maestría en Agroforestería Tropical de la Universidad de Nariño, profesor de cartografía y estadística en el programa de Ingeniería Agroforestal en la Universidad de Nariño, consultor en temas cartográficos y ambientales en la Fundación Cataruben, profesor tiempo completo Fundación Universitaria del Trópico Americano – Unitrópico.

# RESUMEN GENERAL

Pronóstico de caudales horarios mediante el Filtro de Kalman Discreto y Filtro de Kalman de conjuntos en la cuenca del río Huaynamota, México

La variabilidad climática ha provocado el aumento de la frecuencia de ocurrencia de fenómenos climáticos extremos, eventos de lluvias intensas se traducen en cambios bruscos en los caudales y por ende posibles inundaciones o avalanchas. Para prevenir y mitigar los efectos negativos es indispensable contar con información anticipada que permita la toma de decisiones acertadas para la gestión del riesgo de desastres o la administración de infraestructuras. Se evaluó el filtro de Kalman discreto y filtro de Kalman de conjuntos para pronosticar los caudales horarios en la estación Chapalagana con base en los últimos registros de las estaciones Chapalgana y platanitos, en la cuenca del río Huaynamota, en los Estados de Zacatecas, Jalisco y Nayarit al norte de México. La evaluación del ajuste de los pronósticos se hizo mediante la raíz del cuadrado medio del error y el índice de Nash-Sutcliffe, y la determinación del tamaño de conjunto se hizo a través de un análisis de sensibilidad que relaciona el ajuste de la serie pronosticada contra el tamaño del conjunto. El filtro de Kalman discreto obtuvo los mejores ajustes en los seis pasos evaluados, no obstante, tiende a generar desplazamiento de la serie pronosticada. El filtro de Kalman de conjuntos converge con mayor precisión sobre la serie observada una vez que se ejecuta la actualización de estados, identifica el comportamiento no lineal de los caudales y realiza estimación con base a esta tendencia, pero genera sobrestimación de los picos. La estimación de estados en lugar de estimar valores escalares de caudal junto con la incorporación dinámica del tiempo de retraso entre series permite mejorar el ajuste de los pronósticos.

**Palabras clave**: precipitación, asimilación de datos, método recursivo, simulación de Monte Carlo.

---

Tesis de Doctorado en Ingeniería Agrícola y Uso Integral del Agua, Universidad Autónoma Chapingo.
Autor: Ildefonso Narvaez Ortiz
Directora de tesis: Dra. Laura Ibáñez Castillo.

# GENERAL ABSTRACT

Forecast of hourly streamflows using Kalman filters in the Huaynamota river basin, Mexico

Climatic variability has caused increasing frequency of occurrence of extreme weather phenomena, intense rainfall events caused by abrupt changes in flow rates, hence possible floods or avalanches. In order to prevent and mitigate negative effects, it is essential to have advance information that allows the precise decision making for disaster risk management or infrastructure administration. The discrete Kalman filter and Ensemble Kalman filter to forecast hourly streamflow were evaluated at the Chapalagana station based on the latest records of Chapalagana and Platanitos stations in the Huaynamota River basin, States of Zacatecas, Jalisco and Nayarit, Mexico. The evaluation of forecast adjustment was done over root mean square error and the Nash-Sutcliffe index, and the determination of the set size was done through a sensitivity analysis that relates the adjustment of the forecast series against the set (ensemble) size. The discrete Kalman filter obtained the best adjustments in the six steps evaluated, however, it is tending to generate displacement of the predicted series. The Ensemble Kalman filter converges greater precision on the observed series once the state update is executed, identifies the non-linear behavior of the flows and it makes estimates based on this trend, but generates an overestimation of the peaks. State estimation instead of estimate scalar flow values together with the dynamic incorporation of the delay time between series allows to improve forecast adjustment.

**Key words**: precipitation, data assimilation, recursive method, Monte Carlo simulation.

---

Doctoral's Thesis in Agricultural Engineering and Integrated Water Use
University of Chapingo
Author: Ildefonso Narváez Ortiz
Thesis Advisor: Dra. Laura Ibáñez Castillo.

# 1 INTRODUCCIÓN GENERAL

La ocurrencia de eventos extremos relacionados con la precipitación se ha intensificado a causa del cambio climático (IPCC, 2013), situación que se ve reflejada en las variaciones de intensidad y frecuencia con que se presentan fenómenos de lluvia o sequía y que afecta el desarrollo de las actividades agropecuarias e industriales y pone en riesgo la seguridad de las personas y sus posesiones. La gestión del riesgo de desastres se centra en la prevención para proteger a las personas, centros poblados e infraestructura ante fenómenos climáticos adversos, debido a que generan daños físicos, desestabilización social, pérdidas económicas y ambientales, afectación psicológica, entre otros (Hapuarachchi et al., 2011).

Las variaciones de caudal en los ríos están directamente asociadas con la precipitación y las características físicas de la cuenca. Cuando se dan eventos extremos de lluvia se desencadenan cambios bruscos en los caudales, lo que puede resultar en avalanchas e inundaciones, o daños en la infraestructura construida. Para disminuir el impacto negativo de inundaciones o cambios repentinos del nivel de los caudales se requiere tener la capacidad de pronosticar un evento con suficiente anticipación para alertar a la población. De esta forma se tiene información objetiva para la toma de decisiones sobre evacuación, y prevención de eventos colaterales como deslizamientos o avalanchas (Hapuarachchi et al., 2011).

Dentro del estudio de fenómenos hidrológicos se han propuesto diversos métodos de pronóstico en función de la información disponible. Métodos determinísticos, estadísticos y recursivos son los más frecuentes. Los métodos determinísticos involucran datos precisos de cada componente de la cuenca, mientras que los métodos basados en la estadística clásica como la regresión lineal y los métodos recursivos buscan integrar la información disponible y procesarla mediante algoritmos especializados. En general, todos los modelos requieren la instrumentación de la cuenca y el río, pero los métodos determinísticos son altamente exigentes en registros de datos climáticos y físicos. Por otra parte, los

métodos de asimilación de datos como el filtro de Kalman, permiten la estimación de estados en sistemas dinámicos mediante la actualización continua e incorporación de nuevos registros en función del tiempo (Welch & Bishop, 2006), inicialmente se propuso para sistemas dinámicos lineales (Filtro de Kalman Discreto - DKF) (Kalman, 1960) posteriormente, se adaptó para tratar sistemas dinámicos no lineales mediante el filtro de Kalman Extendido (EKF), filtro de Kalman Unscented (UKF) y Filtro de Kalman de Conjuntos (EnKF) (Wu & Wang, 2012).

En aplicaciones hidrológicas se ha evaluado el filtro de Kalman discreto para el pronóstico de caudales diarios y horarios (Alvarado-Hernández et al., 2020; Morales-Velázquez et al., 2014) obteniendo resultados aceptables. EnKF es frecuentemente utilizado en la estimación de parámetros hidrológicos (S. Wang et al., 2017), la estimación dual de parámetros y estados (Moradkhani et al., 2005), estimación de varios estados simultáneos (D. Wang et al., 2009), estimación de la humedad del suelo (Brandhorst et al., 2017; Medina González et al., 2015; Samuel et al., 2014), detección de patrones temporales en caudales (Pathiraja et al., 2016), integración de varias fuentes de datos para el pronóstico de caudales en cuencas urbanas (Leach et al., 2018), por mencionar algunos.

Dado que los fenómenos hidrológicos responden a un patrón no lineal (Kundzewicz & Napiórkowski, 1986), EnKF puede obtener mejor ajuste en los pronósticos de caudales horarios, utilizando registros en tiempo real de estaciones distribuidas en el área de la cuenca. En este sentido, se comparó el filtro de Kalman discreto (DKF), filtro de Kalman de conjuntos (EnKF), modelo autorregresivo (ARX) y sus combinaciones para hacer pronóstico de caudales horarios en la cuenca del río Huaynamota, México.

Los siguientes cuatro capítulos presentan las bases del filtro de Kalman y su relación con el pronóstico de caudales, resultados de la implementación del filtro de Kalman discreto (DKF) y Filtro de Kalman de Conjuntos (EnKF). El capítulo 2 Revisión de literatura condensa la principal información relacionada con las bases del filtro de Kalman y su aplicación en la estimación de estados en sistemas hidrológicos

dinámicos de cuenca hidrográfica, específicamente los caudales. Los capítulos 2 y 3 presentan los resultados de la implementación de filtros de Kalman con registros horarios de caudal de las estaciones Chapalagana y Platanitos en la cuenca del río Huayamota al norte de México, en el primero se centra en DKF estimando estados comparado con EnKF que hace estimaciones de valores escalares de caudal, además se compara con las combinaciones de ARX-DKF y DKF-EnKF; en el capítulo 3 se tienen los resultados de la implementación de EnKF y DKF estimando estados y con actualización dinámica de los tiempos de retraso entre las series o variables de entrada. Finalmente, el capítulo 4 corresponde a las conclusiones generales derivadas de los dos capítulos previos.

## 1.1 Literatura citada


Alvarado-Hernández, L., Ibanez-Catillo, L., Ruiz-García, A., González-Leiva, F., & Vázquez-Peña, M. (2020). Pronóstico horario de caudales mediante filtro de kalman discreto en el Río Huaynamota, Nayarit, México. Agrociencia, 54, 295–312.

Brandhorst, N., Erdal, D., & Neuweiler, I. (2017). Soil moisture prediction with the ensemble Kalman filter: Handling uncertainty of soil hydraulic parameters. Advances in Water Resources, 110(February), 360–370. https://doi.org/10.1016/j.advwatres.2017.10.022

Hapuarachchi, H. A. P., Wang, Q. J., & Pagano, T. C. (2011). A review of advances in flash flood forecasting. Hydrological Processes, 25(18), 2771–2784. https://doi.org/10.1002/hyp.8040

IPCC. (2013). Cambio climático 2013: Bases físicas. IPCC.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. Journal of Basic Engineering, 82(Series D), 35–45.

Kundzewicz, Z. W., & Napiórkowski, J. J. (1986). Nonlinear models of dynamic hydrology. Hydrological Sciences Journal, 31(2), 163–185. https://doi.org/10.1080/02626668609491038

Leach, J. M., Kornelsen, K. C., & Coulibaly, P. (2018). Assimilation of near-real time data products into models of an urban basin. Journal of Hydrology, 563(April), 51–64. https://doi.org/10.1016/j.jhydrol.2018.05.064

Medina González, H., Hernández Pereira, Y., Santiago Piloto, A. B., & Lau Quan, A. (2015). Modelación de perfil de humedad de suelos empleando un filtro de Kalman de Monte-Carlo. Revista Ciencias Técnicas Agropecuarias, 24(2), 31–37.

Moradkhani, H., Sorooshian, S., Gupta, H. V., & Houser, P. R. (2005). Dual state-parameter estimation of hydrological models using ensemble Kalman filter. Advances in Water Resources, 28(2), 135–147. https://doi.org/10.1016/j.advwatres.2004.09.002

Morales-Velázquez, M. I., Aparicio, J., & Valdes, J. B. (2014). Pronóstico de avenidas utilizando el filtro de kalman discreto. Tecnologia y Ciencias Del Agua, 5(2), 85–110.

Pathiraja, S., Marshall, L., Sharma, A., & Moradkhani, H. (2016). Hydrologic modeling in dynamic catchments: A data assimilation approach. Water Resources Research, 52(5), 3350–3372. https://doi.org/10.1002/2015WR017192

Samuel, J., Coulibaly, P., Dumedah, G., & Moradkhani, H. (2014). Assessing model state and forecasts variation in hydrologic data assimilation. Journal of Hydrology, 513, 127–141. https://doi.org/10.1016/j.jhydrol.2014.03.048

Wang, D., Chen, Y., & Cai, X. (2009). State and parameter estimation of hydrologic models using the constrained ensemble Kalman filter. Water Resources Research, 45(11), 1–13. https://doi.org/10.1029/2008WR007401

Wang, S., Huang, G. H., Baetz, B. W., Cai, X. M., Ancell, B. C., & Fan, Y. R. (2017). Examining dynamic interactions among experimental factors influencing hydrologic data assimilation with the ensemble Kalman filter. Journal of Hydrology, 554, 743–757. https://doi.org/10.1016/j.jhydrol.2017.09.052

Welch, G., & Bishop, G. (2006). An Introduction to the Kalman Filter. In University of North Carolina at Chapel Hil (Issue 1).

Wu, X., & Wang, Y. (2012). Extended and Unscented Kalman filtering based feedforward neural networks for time series prediction. Applied Mathematical Modelling, 36(3), 1123–1131. https://doi.org/10.1016/j.apm.2011.07.052

# 2 REVISION DE LITERATURA

## 2.1 Variabilidad climática y pronóstico de caudales

La variación climática se manifiesta en la ausencia o exceso de precipitación y tiene directa repercusión en el nivel y volumen del caudal de los ríos (IPCC, 2012). Los eventos de lluvia intensos sobrepasan los niveles de absorción de suelo, generan escorrentías y fuertes crecidas espontaneas en los cauces de ríos. Estos eventos ocasionan afectaciones a las personas y sus posesiones, desencadenando crisis social y desestabilidad. Los eventos extremos de caudal se generan por lluvias intensas o situaciones extraordinarias como fallas en represas o estancamiento accidental en los cauces (Hapuarachchi et al., 2011). Con frecuencia las lluvias son las principales causantes de los eventos extremos y de ahí parte el centrarse en los fenómenos climáticos más que en situaciones extraordinarias.

Con la finalidad de obtener pronósticos acertados del clima se han desarrollado diversos métodos, cada uno con características particulares en función a la disponibilidad de datos o la capacidad de cómputo. En general se puede distinguir dos tipos de modelos: de caja blanca y de caja negra (Khan & Khan, 2012).

Los modelos de caja blanca hacen el estudio detallado de la lógica y estructura interna, es decir, las relaciones físicas entre los componentes del sistema, por esto es indispensable tener datos precisos de variables representativas del sistema dinámico cuenca. Frecuentemente, los modelos de caja blanca utilizan datos provenientes de sensores remotos que se comparan con estaciones en tierra (Hapuarachchi et al., 2011), estos métodos requieren algoritmos complejos y calibrados para obtener pronósticos adecuados. De igual manera, se encuentran limitaciones de resolución espacial e incertidumbre en los pronósticos (Zou et al., 2017a). Es adecuada su utilización cuando se tiene detallada información climática y física, resultante de la adecuada instrumentación de la cuenca.

Por su parte, los modelos de caja negra no requieren del conocimiento sobre el comportamiento y estructura interna del sistema dinámico, solo se analiza los elementos esenciales (Khan & Khan, 2012). Son adecuados en situaciones de falta de registros y pobre instrumentación, no obstante, deben ser acompañados con técnicas probabilísticas para estimar el error y confiabilidad de los resultados. En situaciones donde se tiene grandes volúmenes de información climática ruidosa y escasa instrumentación de cuencas los modelos de caja negra son ideales. Debido a que no requieren parámetros detallados que medien el proceso, y en su lugar se modela internamente para producir una salida sin entrar a detallar el proceso.

Procesos hidrológicos como los caudales son complejos e intervienen muchas variables físicas que implican alto costo y dificultad para la precisa medición. Por esta razón, los métodos probabilísticos son frecuentemente utilizados para obtener una respuesta que se incluye dentro de intervalos de confianza (Aparicio, 2017). Frecuentemente, las distribuciones asimétricas como Gumbel, Logística o Pearson se utilizan para analizar variables hidrológicas, y, además, la distribución normal se utiliza para modelar los errores o residuales. Se requiere ajuste del modelo probabilístico mediante métodos como mínimos cuadrados o máxima verosimilitud.

## 2.2 Relaciones lluvia - escurrimiento

La cuenca hidrográfica es un sistema con diversos componentes y entradas y salidas que lleva inmerso relaciones que son representativas de una superficie en particular, donde intervienen las características biofísicas predominantes. Una de las relaciones fundamentales es la lluvia – escurrimiento que adquiere aún más relevancia cuando se trata de cuencas hidrográficas con limitada instrumentación tanto en tiempo como en espacio (Aparicio, 2017). Los métodos determinísticos o probabilísticos recurren a parámetros de la cuenca que intervienen en la circulación del agua desde el momento de la precipitación hasta la formación de los escurrimientos o caudales. El área de la cuenca, forma de la cuenca, distribución de la precipitación, geomorfología, entre otros, son

fundamentales para implementar métodos de pronóstico de caudales (Te Chow et al., 1996).

## 2.3 Modelación y predicción de procesos hidrológicos

Comprender como funcionan los fenómenos naturales es la base fundamental para mitigar los impactos del cambio climático y también para el diseño de estructuras hidráulicas. Sin embargo, la interacción de los componentes físicos, biológicos y las sociedades generan alta complejidad de los sistemas de cuenca hidrográfica, por esta razón, la abstracción del sistema a través de modelos que se basan en variables de entrada representativas del sistema (Te Chow et al., 1996) es indispensable para analizar y comprender su funcionamiento. La esencia de los modelos son ecuaciones matemáticas que representan las interacciones entre las variables representativas de la cuenca, el ajuste de estas ecuaciones requiere de conocimientos matemáticos e implica conocer el sistema cuenca (WMO, 2011). La complejidad del modelo es proporcional a la magnitud de la cuenca hidrográfica, y así mismo, el ajuste y actualización dinámica del modelo tiende a ser complejo.

Una vez ajustado un modelo se puede hacer pronósticos con base a los registros históricos de las variables de entrada. No obstante, cuando se trata de pronósticos en tiempo real se requiere la constante actualización del modelo, exigiendo la intervención en el proceso de ajuste y también el uso de sistemas de cómputo capaces de administrar grandes volúmenes de datos. Al tratarse de modelos probabilísticos en los pronósticos se tiene incertidumbre inmersa que es importante estimarse para definir la confiabilidad del modelo.

## 2.4 Incertidumbre de la predicción hidrológica

El objetivo fundamental al hacer predicciones de caudal es ofrecer información con un alto grado de confiabilidad que permita tomar acciones que impidan o limiten la afectación por parte de fenómenos climáticos extremos. La compleja configuración de los sistemas climáticos e hidrológicos representados por

modelos y variables aleatorias que pueden tomar infinitos valores hacen que las predicciones contengan incertidumbre, es decir, que los valores obtenidos del modelo sean exactos. Esta situación ha llevado a desarrollar procedimientos que permitan estimar el grado de confianza que se puede tener en un modelo dado, en otras palabras, determinar el potencial error que está inmerso en los resultados obtenidos (Zou et al., 2017a).

El enfoque probabilístico es frecuentemente adoptado en hidrología, ya que permite hacer predicciones y estimar el error asociado (Gouweleeuw et al., 2005). Las diferentes versiones del filtro de Kalman agrupan las propiedades probabilísticas, la recursividad y la asimilación de datos en tiempo real (Evensen, 2009; Han & Coulibaly, 2017) lo que confiere capacidad para ser aplicado en el pronóstico de caudales en tiempo real. La estimación del error se hace con base en las matrices de covarianza tanto del proceso como de la medición (Kalman, 1960) o con base en simulaciones de Monte Carlo que perturban las observaciones (Evensen, 1994, 2009).

## 2.5 Variables y función de distribución

### 2.5.1 Variables aleatorias

Un insumo indispensable para los procedimientos estadísticos es el conjunto de datos que representan un fenómeno bajo estudio (Infante & Zaraté, 2012). Dicho conjunto puede ser obtenido mediante mediciones bajo condiciones controladas o bien secuencias de registros de fenómenos naturales. Estos datos generalmente se encierran en un rango de posibles valores que puede tomar la variable aleatoria bajo estudio, y se denomina Espacio Muestral. Una variable aleatoria es una función que a cada observación de un espacio muestral le asocia un número real (Naghettini & De Andrade, 2007).

El caudal de un río se comporta como una variable aleatoria que puede asumir cualquier valor, por tanto, no es posible predecir un valor exacto dentro de un periodo de tiempo predefinido. Su análisis se basa en probabilidades y

distribuciones de probabilidades ajustadas a los datos medidos de la variable aleatoria. Se considera los registros históricos de caudales como una muestra con parámetros desconocidos, y con base en la estadística probabilística se ajustan distribuciones que permitan describir el comportamiento de la variable (Naghettini & De Andrade, 2007).

### 2.5.2 Función de distribución de probabilidades

El concepto de probabilidad está directamente ligado a la frecuencia relativa de las observaciones o eventos (Infante & Zaraté, 2012). Una función de distribución se encarga de asignar valores de probabilidad para cada evento, puede ser representada por graficas o tablas (Kaps & Lamberson, 2004). Cuando cada evento es independiente de los demás, se hace la asignación de probabilidad de manera directa de acuerdo con la frecuencia relativa. En cambio, cuando un evento depende de otro la probabilidad se actualiza cada vez que se obtiene un nuevo dato y puede irse complementando hasta tener un valor de mayor precisión (Naghettini & De Andrade, 2007), en el ámbito de la estadística se conoce como probabilidad condicional y se denota como $P(A|B)$ probabilidad de A dado B.

Una aplicación del concepto de probabilidad condicional es el Teorema de Bayes, que consiste en el cálculo de la probabilidad a priori con base a probabilidades conocidas de eventos separados, y que se van complementando a medida que se tienen nuevos eventos (Infante & Zaraté, 2012; Naghettini & De Andrade, 2007), esta dinámica ha dado origen a los métodos estadísticos bayesianos. El filtro de Kalman es un método bayesiano, dado que, actualiza de manera constante la distribución de probabilidades y además, es un método recursivo (Evensen, 2009; Maybeck, 1979), en otras palabras, posee una memoria interna que le permite ejecutarse sin la necesidad de mantener la serie de datos completa.

Para el estudio de los fenómenos naturales se han propuesto diversas distribuciones según el comportamiento predominante. La distribución normal es la más conocida y que representa un amplio número de variables o funciones,

para estudios específicos en el área hidrológica se encuentran las distribuciones asimétricas como Gumbel, Pearson, Weibull, entre otras.

## 2.6 Filtro de Kalman

El filtro de Kalman es un método recursivo de asimilación de datos (data assimilation) con fundamento bayesiano (Dumedah & Coulibaly, 2013; Han & Coulibaly, 2017), se concibe como un método de caja negra para procesos estocásticos (Archer et al., 2015). Estos métodos tratan de obtener la máxima precisión en la estimación de estados mediante la combinación estadística de las observaciones, una distribución de probabilidades y actualización continua de los estados (Quadros et al., 2016). El filtro de Kalman discreto (DKF) (Kalman, 1960) y las variaciones como el filtro de Kalman de Conjuntos (EnKF) (Evensen, 1994) y el Filtro de Kalman Extendido (EKF) corresponden a técnicas de asimilación de datos que se usan con frecuencia en estudios hidrológicos (Abaza et al., 2015; Liu et al., 2012; Samuel et al., 2014). (Abaza et al., 2014).

Los métodos recursivos son frecuentemente utilizados en áreas hidrológicas (Dumedah & Coulibaly, 2013; Pathiraja et al., 2016; Samuel et al., 2014) para hacer pronósticos o estimaciones distribuidas. Dentro de estos métodos el filtro de Kalman y sus variantes sobresalen (Pathiraja et al., 2016; Quadros et al., 2016) gracias a su capacidad de asimilar varias fuentes de datos, combinar observaciones y simulaciones, y actualizar su estado continuamente (DeChant & Moradkhani, 2012; Han & Coulibaly, 2017), con eficiencia computacional superior a modelos como el de Promedios Bayesianos (Vrugt & Robinson, 2007).

En general el algoritmo del filtro de Kalman realiza la estimación en tiempo real de las variables del estado del sistema, dentro de un ciclo infinito donde cada repetición representa un nuevo registro. El ciclo se divide en dos fases; predicción y corrección. Con base en la información del instante anterior y el modelo dinámico o ecuación de estado en la fase de predicción se obtiene una estimación “a priori” del estado en el momento $k$, sin tener disponible los datos más recientes de las variables, posteriormente, en la fase de corrección el algoritmo obtiene la

estimación “a posteriori” del estado en el momento $k$, en este caso ya se incorpora los datos más recientes de las variables de entrada. Generalmente se inicia haciendo el pronóstico del estado y después se corrige incorporando los datos de la medición más reciente, las fases del algoritmo son complementarias y se repiten indefinidamente. La importancia de la corrección recae en que reduce el error sistemático que puede darse al utilizar únicamente la ecuación de dinámica o pronóstico de estado, en cambio, si únicamente se usa los datos de los sensores, se puede obtener grandes oscilaciones en la predicción debido a las variaciones dentro de cada serie de tiempo. La ganancia de Kalman ($K_k$) equivale a factores de corrección para cada una de las variables de entrada respecto a la variable de salida.

Para la aplicación del Filtro de Kalman en el pronóstico de caudales, la cuenca hidrográfica se asume como un sistema dinámico conformado por elementos que interactúan para procesar entradas y generar salidas, la relación entre las entradas y las salidas se conoce como función de respuesta (Hidrograma Unitario Instantáneo) (Valdés et al., 1980). En el caso que nos ocupa acerca de los caudales, como variable representativa de entrada se tiene la precipitación y como salida esta propiamente los caudales. La precipitación que cae sobre la cuenca es atenuada por la vegetación, tipo del suelo, humedad antecedente, usos que tenga el agua, temperatura, evaporación, entre otras, que hacen que se suavice la serie y el agua se distribuya temporal y espacialmente hasta llegar a los escurrimientos superficiales y al caudal de salida. El hidrograma unitario de manera abstracta hace la estimación de la atenuación y relaciona la precipitación con el caudal de salida.

El Hidrograma Unitario Instantáneo consiste en la estimación de la función de respuesta de la cuenca en un momento dado, lo que le confiere la habilidad de identificar la variación temporal de la precipitación. Por esta razón, se utiliza en los modelos de pronóstico en donde se actualizan de forma cíclica.

El filtro de Kalman hace la estimación de estados en sistemas dinámicos, dichos estados equivalen a los coeficientes de la función de respuesta que representan la variable de salida, es decir los caudales. Cuando se incluye en el algoritmo términos autorregresivos, se genera un vector de estados que es similar al Hidrograma Unitario Instantáneo. De esta forma el Filtro de Kalman hace la estimación de estados de la cuenca en función del tiempo y el espacio para un instante dado, y se actualiza por cada ciclo o paso de tiempo que se ejecuta.

## 2.7 Asimilación de datos

La asimilación de datos consiste en combinar información complementaria de varias fuentes o modelos que representan el fenómeno bajo estudio, para obtener una estimación optima con base en la incertidumbre de las mediciones y de los pronósticos del modelo, es decir, mejorar su precisión asumiendo una distribución estadística (Abaza et al., 2015; Reichle, 2008). Asimismo, mediante la asimilación de datos se incorpora nuevos datos en cada ciclo de ejecución. Para el pronóstico de caudales se busca asimilar registros de precipitación o caudal en tiempo real del mayor número de estaciones ubicadas en sitios representativos de la cuenca, los datos en tiempo real permiten mantener actualizado el algoritmo a fin de obtener pronósticos de mayor precisión.

## 2.8 Recursividad

El termino recursivo hace referencia a que su funcionamiento no necesita que todos los datos previos estén almacenados y que sean reprocesados cada vez que se obtiene una nuevo registro (Maybeck, 1979; Sun, Seidou, Nistor, & Liu, 2016), internamente el algoritmo opera un sistema de memoria que se actualiza con cada nuevo registro. Esta propiedad le confiere al Filtro de Kalman versatilidad a la hora de implementación, dado que, no exige alta capacidad de almacenamiento, en otras palabras, de las series que son variables de entrada al filtro solo se debe almacenar los últimos registros, según los términos autorregresivos, y los datos más allá se pueden eliminar sin afectar el desempeño del algoritmo. De igual forma, el hecho de tener una memoria interna facilita el

proceso de cálculo ya que no debe repetir procesos matemáticos sino actualizar con base en los nuevos datos que se alleguen, lo que le confiere al filtro flexibilidad para implementarse en sistemas de cómputo simples.

## 2.9 Filtro de Kalman discreto (Lineal)

El algoritmo de filtro de Kalman discreto (DKF) fue publicado en 1960 como una opción recursiva para estimar el comportamiento de sistemas dinámicos lineales (Cloke & Pappenberger, 2009; Gillijns et al., 2006; Kalman, 1960; Welch & Bishop, 2006). Se ha aplicado con éxito en la estimación de órbitas y velocidad con base en datos provenientes de sensores ruidosos (Brammer, 1974) y ha tenido amplia difusión y aplicaciones en áreas de electrónica y robótica (Bader et al., 2017; Odry et al., 2018) frecuentemente para estimación de posición y velocidad. Fue la primera versión del filtro y es la base para desarrollar modificaciones o adaptaciones para dar mayor versatilidad en las aplicaciones.

Con base en el Filtro de Kalman discreto se han desarrollado algoritmos para tratar sistemas dinámicos no lineales. Entre las variantes utilizadas con más frecuencia están; filtro de Kalman extendido (Extended Kalman Filter – EKF), filtro de Kalman “Unscented” (Unscented Kalman Filter – UKF), filtro de Kalman de conjuntos (Emsemble Kalman Filter – EnKF) y filtro de Kalman con modelos autorregresivos.

El algoritmo del DKF basa su formulación en el modelo espacio estado, el cual supone la existencia de una o más variables con las cuales se puede describir el estado en que se encuentra el sistema (Shumway & Stoffer, 2017), se compone de dos grupos de ecuaciones, el primero realiza la predicción de estados y covarianza del error y el segundo se encarga de actualizar la estimación y la covarianza del error con base en la ganancia de Kalman (Figura 1). Dentro del algoritmo la ganancia de Kalman ($K_k$) es el componente esencial que hace que este filtro tenga las propiedades de estimador óptimo (Welch & Bishop, 2006).

Ecuación de estado

$$x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1}$$

Ecuación de medición

$$z_k = Hx_k + v_k$$

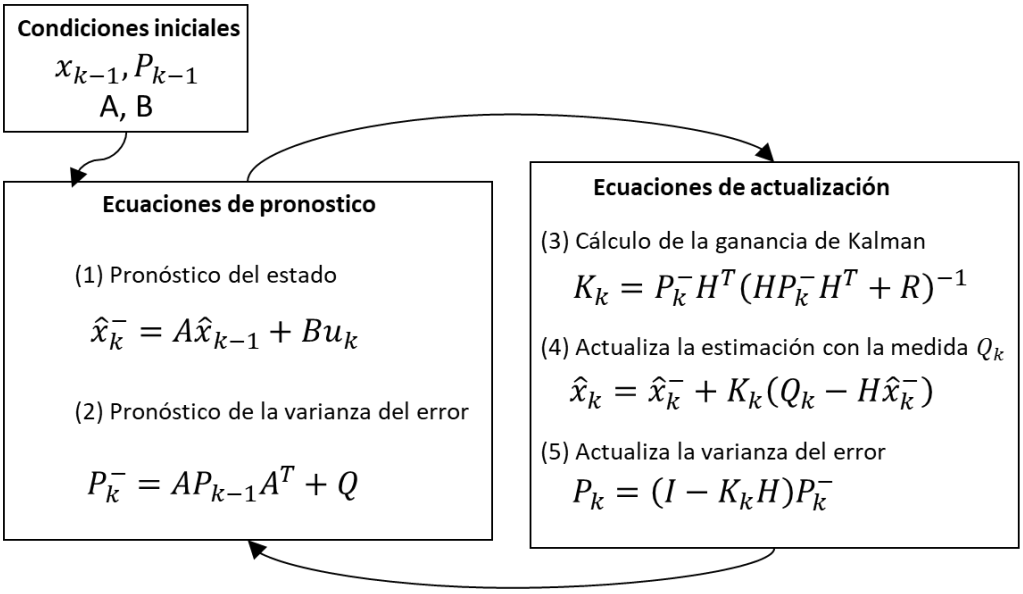



Figura 2.1. Algoritmo de filtro de Kalman discreto (Welch & Bishop, 2006)

La matriz $A$ de la ecuación de estado es de tamaño $n \times n$ y relaciona el estado en el tiempo $k-1$ con el estado en el tiempo actual $k$. La matriz $B$ de tamaño $n \times l$ relaciona la entrada opcional de control $u$ con el estado $x$ y tiene dimensión de tamaño $l$. Por último, la matriz H en la ecuación de medición relaciona el estado con la medición $z_k$. Cada matriz puede cambiar a lo largo de la ejecución del algoritmo, pero por lo general se asumen constantes (Welch & Bishop, 2006). Las variables aleatorias $w_k$ y $v_k$ corresponden al ruido del proceso y la medición respectivamente (Welch & Bishop, 2006), es decir las varianzas que se tiene en cada serie. Estos componentes de ruido se asumen como valores independientes de ruido blanco con distribución de probabilidades normal, algebraicamente de la siguiente forma:

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

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

Donde, Q y R representan las matrices de covarianza del ruido del proceso y covarianza del ruido de la medición, al igual que las anteriores matrices pueden ser cambiantes a lo largo del proceso pero normalmente se mantienen como constantes (Welch & Bishop, 2006).

El Filtro de Kalman discreto se concibe como un estimador óptimo, y se puede definir de varias formas en función de los criterios que se elijan para evaluar el ajuste. Una manera de ver la estimación óptima es que el Filtro de Kalman gestiona toda la información que se le pueda proveer independientemente de su precisión a través del conocimiento del sistema y mediciones continuas, asumiendo distribuciones estadísticas para el ruido y cualquier información que pueda configurarse en las condiciones iniciales (Maybeck, 1979).

## 2.10 Filtro de Kalman no lineal

Las variaciones del filtro de Kalman tratan de mejorar su respuesta ante situaciones de no linealidad. Se destaca el Filtro de Kalman Extendido (EKF) y Filtro de Kalman de conjuntos (EnKF), EKF basa su funcionamiento en la expansión de las series de Taylor de primer orden, lo que conlleva un alto costo computacional y puede bajar su eficiencia en tiempo real. Por su parte EnKF maneja adecuadamente la no linealidad y reduce el número de grados de libertad en comparación con EKF (Reichle, 2008) y su algoritmo es de baja complejidad lo que facilita su implementación (Rafieeinasab et al., 2014). Por estas razones, el EnKF es aplicado con frecuencia en áreas hidrológicas (Fan et al., 2017; Maxwell et al., 2018; Medina González et al., 2015; S. Wang et al., 2017), aunque se debe tener en cuenta que el EnKF no es adecuado donde hay fuerte no linealidad (Rafieeinasab et al., 2014).

### 2.10.1 Filtro de Kalman de conjuntos

Dada la restricción del Filtro de Kalman discreto a sistemas dinámicos lineales, se ha propuesto modificaciones que permitan extender su funcionalidad a sistemas dinámicos no lineales (Sun, Seidou, Nistor, & Liu, 2016). Las

modificaciones más importantes son el Filtro de Kalman de Conjuntos (EnKF) y Filtro de Kalman Extendido (EKF) (Sun, Seidou, Nistor, & Liu, 2016), EnKF se divide en dos líneas; EnKF root square (Bishop et al., 2001; Whitaker & Hamill, 2002) y EnKF de observaciones perturbadas (Evensen, 1994), este último tiene mejor capacidad para gestionar no linealidades (Lawson & Hansen, 2004). Los algoritmos de EKF y EnKF son estimadores subóptimos de estados en sistemas no lineales (Reichle et al., 2002), el primero basa su funcionamiento en la linealización mediante la expansión de primer orden de la serie de Taylor (Wu & Wang, 2012) y la resolución de la matrices dinámicas jacobianas, este último proceso genera alto costo computacional (Gillijns et al., 2006; Krener & Duarte, 1996). El segundo, evita el cálculo de matrices jacobianas y hace la estimación del error estadístico mediante simulaciones de Monte Carlo (Gillijns et al., 2006; Reichle, 2008) que se generan en grupos de números centrados en cero denominados conjuntos (Ensembles) (Evensen, 1994, 2003, 2009), los cuales deben tener un tamaño lo suficientemente grande para cumplir los requisitos de precisión (Han & Coulibaly, 2017).

El Filtro de Kalman de Conjuntos se basa en los mismos principios que DKF pero enfocado al tratamiento de sistemas dinámicos no lineales. Su algoritmo guarda alta similitud al tener dos grupos de ecuaciones que hacen la predicción y actualización de estados o estimaciones (Figura 2). En las ecuaciones de predicción, la primera hace la estimación de estados y para esto se requiere definir una función que represente el fenómeno bajo estudio ya sea lineal o no lineal. La operatividad de esta modificación del Filtro de Kalman se basa en simulaciones de Monte Carlo que son incorporadas como ruido blanco en la ecuación de estimación de estados y como perturbaciones agregadas a cada nuevo dato que se incorpora al algoritmo en la actualización (ecuación 7 Figura 2), es decir, se representan por los términos $v_k^i$ y $w_k^i$ que corresponden a ruido blanco generado aleatoriamente con media cero y bajo distribución normal, asociados a las matrices de covarianza $R_k$ y $Q_k$ (Gillijns et al., 2006). La matriz $h$ relaciona los estados con las mediciones perturbadas.

Al grupo de valores obtenidos con la simulación de Monte Carlo se les denomina conjunto (Ensemble) y la cantidad de miembros o elementos adecuada que debe contener se debe definir con base en un análisis de sensibilidad que relacione un estadístico de ajuste contra diversos tamaños $q$ (Quiroz et al., 2019). A mayor cantidad de miembros se espera mejor ajuste de las series pronosticadas, para estudios hidrológicos se sugiere entre 50 y 300 miembros por conjunto (Gillijns et al., 2006; Quiroz et al., 2019). EnKF pertenece a la categoría de filtros de partículas (Kotecha & Djuric, 2003) derivado como un estimador subóptimo para un problema bayesiano general (Evensen & van Leeuwen, 2000).

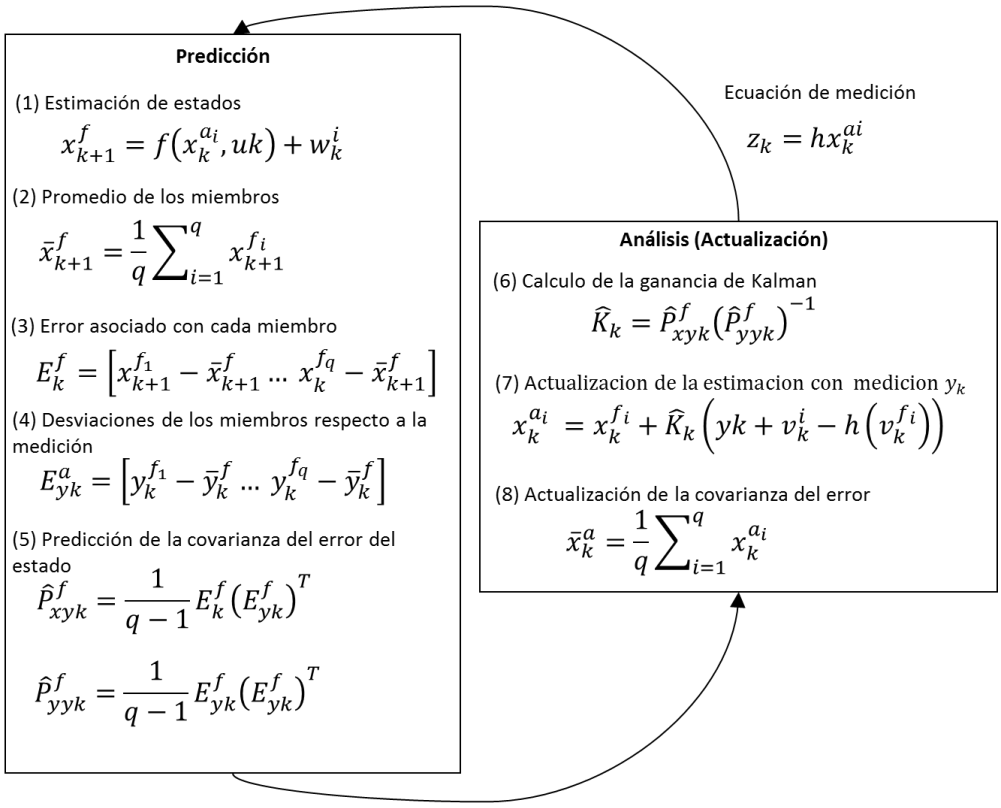



Figura 2.2. Algoritmo de EnKF (Gillijns et al., 2006).

El tamaño de conjunto es uno de los parámetros más importantes del EnKF y depende del propósito del pronóstico (Han & Coulibaly, 2017; Herr & Krzysztofowicz, 2010). Cuando se trata de pronósticos del estado de un rio en un espacio dado o alerta temprana, el tamaño debe ser de varios cientos. Mientras que para pronósticos del proceso de flujo del río (operación de vías fluviales) el tamaño de conjunto debe ser de varios miles (Herr & Krzysztofowicz, 2010). Sin

embargo, varios centros de pronóstico alrededor del mundo utilizan conjuntos que oscilan entre 15 y 51 miembros (Cloke & Pappenberger, 2009; Cuo et al., 2011). Para tener información respecto a la magnitud del error y el tamaño conjunto se deben hacer análisis de sensibilidad (S. Wang et al., 2017).

Específicamente para el pronóstico de caudales a corto plazo (3 y 240 horas) Abaza, Anctil, Fortin, y Turcotte (2014) evaluaron el desempeño del EnKF concluyendo que 20 miembros producen resultados confiables. Además, recomiendan el EnKF para pronósticos de corto plazo en cuencas dominadas por procesos de lluvia escurrimiento (Abaza et al., 2014). Asimismo, se ha aplicado con éxito en la evaluación de la humedad del suelo (Brandhorst et al., 2017; Medina González et al., 2015) y evaluación de la evapotranspiración a nivel de cuenca utilizando sensores remotos (Zou et al., 2017a).

A diferencia del Filtro de Kalman discreto, EnKF se asume como un estimador subóptimo en razón a que no opera directamente sobre los datos, sino que son perturbados con números aleatorios bajo distribución normal (Evensen, 2003; Evensen & van Leeuwen, 2000; Gillijns et al., 2006).

### 2.10.2 Filtro de partículas

El algoritmo de filtro de partículas utiliza muestras aleatorias para aproximar el modelo de estado, y son actualizadas por simulación de Monte Carlo. A diferencia del EnKF no considera la aproximación gaussiana, pero, requiere de muestras más grandes (Fan et al., 2017).

### 2.10.3 Filtro de Kalman extendido

El filtro de Kalman extendido (EKF) surge como respuesta a la necesidad de tratar datos con relación no lineal (Jonsson & Palsson, 1994; Welch & Bishop, 2006). La actualización de las estimaciones se basa en el mínimo error cuadrático medio mediante las series de Taylor de primer orden para expansión no lineal (Wu & Wang, 2012). La estimación de la distribución del estado se hace por medio de variables aleatorias gaussianas (van der Merwe & Wan, 2001). Esta

aproximación se acerca al tratamiento de variables no lineales pero el costo computacional es alto (Reichle, 2008). Se denota como debilidad la baja capacidad para tratar variables con fuerte no linealidad (Medina González et al., 2015; S. Wang et al., 2017)

### 2.10.4 Filtro de Kalman “unscented”

El algoritmo de filtro de Kalman “unscented” (UKF) al igual que el EKF estima la distribución del estado mediante variables aleatorias gaussianas (Wu & Wang, 2012), pero con base en un grupo mínimo de puntos seleccionados cuidadosamente (Julier & Uhlmann, 1997; van der Merwe & Wan, 2001), es decir, una muestra determinística (Gillijns et al., 2006). Esta muestra, representa con mayor precisión el promedio y la covarianza, y al tratarse de un sistema no lineal se aplica la expansión de las series de Taylor de segundo orden (van der Merwe & Wan, 2001).

## 2.11 Asimilación de varios sensores

Una de las propiedades del filtro de Kalman es la capacidad para asimilar datos de varias fuentes simultáneamente, y con base en este grupo de datos se hace estimaciones y actualización de los estados del sistema. El algoritmo básico presentado en las Figuras 1 y 2 solo admite una entrada y se debe adaptar a la cantidad de variables o puntos de entrada. Para asimilar un número mayor de sensores o variables de entrada se han planteado varias configuraciones: la primera es la *actualización paralela* que consiste en hacer al mismo tiempo la incorporación de todas las variables de entrada al algoritmo repitiendo rutinas de código, la segunda es la *actualización secuencial* incorpora cada variable una tras otra y puede hacerse ejecutando de forma repetitiva una misma rutina de código dentro de un bucle, y por último la *compresión de datos* que implica el tratamiento algebraico de las ecuaciones de actualización de tal manera que reciban las variables de entrada y retornen un solo valor ya actualizado (Willner et al., 1976). De estas tres opciones, la que brinda igual resultado e implica menores cambios o adaptaciones es la actualización secuencial, dado que se

usa una misma línea de código dentro de un ciclo que se repite tantas veces como sensores o variables se tenga.

## 2.12 Modelos autorregresivos

Los modelos autorregresivos (AR) son una representación de variables aleatorias suponiendo que el valor futuro es influenciado por observaciones pasadas y que su relación tiene un comportamiento lineal. No obstante, en hidrología se encuentran con frecuencia no linealidades que hacen que los modelos autorregresivos tengan limitada aplicación, por esto, para modelar caudales diarios y horarios inicialmente se evaluó la normalización y transformación de los registros (Rajagopalan et al., 2010).

Los modelos AR son una de las primeras aproximaciones para el estudio de series de tiempo, y se basan en la autocorrelación que se presenta dentro de la misma serie de datos. El modelo puede ser solamente autorregresivo (AR) o bien puede tener una o varias variables exógenas (ARX) (Box et al., 2016; Ljung, 1999). Algebraicamente se representa de la siguiente forma:

$$y_{t+1} = -a_1 * y_t - a_2 * y_{t-1} - \cdots - a_{na} * y_{t-na} + b_1 * u_t + b_2 * u_{t-1} + \cdots + b_{nb} * u_{t-nb} \quad (2.1)$$

Donde: y corresponde a la variable de salida con los retrasos según el grado de polinomio, y u corresponde a la variable de entrada exógena con los retrasos según el grado de polinomio.

## 2.13 Redes neuronales

Las técnicas de inteligencia artificial han tomado gran relevancia en la actualidad, logrando un considerable avance en cuanto a desarrollo de algoritmos, asimismo, en las aplicaciones en diversas áreas (Martin & Sanz, 2017). Para el pronóstico de caudales se han realizado evaluaciones basadas en técnicas de inteligencia artificial como redes neuronales y varios algoritmos de entrenamiento. Los métodos comunes para el entrenamiento son: algoritmo de retro propagación

(backpropagation), aprendizaje recurrente en tiempo real y el filtro de Kalman de conjuntos (Muluye, 2011). El entrenamiento y algoritmo que se utilice son decisivos para lograr pronósticos precisos.

Se encuentra extensa literatura sobre redes neuronales entrenadas mediante el algoritmo retro propagación (Saba et al., 2010), pero es limitada en relación al entrenamiento con las variantes del filtro de Kalman, más aún el EnKF. A la fecha se han realizado importantes avances con relación a las redes neuronales artificiales. Lo que ha generado interés en las distintas disciplinas y a su vez ha permitido que se avance en la solución de problemas cada vez más complejos (Lopez & Caicedo, 2009). Se han utilizado en mercadeo, climatología, tránsito vehicular, agronomía, entre otros (Bravo, 2016; Garces & Gomez, 2004; Mejía & Eliana, 2004; Torres et al., 2011).

Las redes neuronales pueden ser entrenadas para realizar pronósticos de caudal en tiempo real (Elsafi, 2014; Pujol Reig, 2009; Zealand et al., 1999). Y a diferencia del filtro de Kalman discreto, tienen la habilidad de tratar con la no linealidad (Wu & Wang, 2012), lo que puede otorgarle mayor precisión en los pronósticos de caudal.

Wu y Wang (2012) compararon el desempeño de una red neuronal “feedforward” entrenada con el filtro de Kalman extendido (EKF) y filtro de Kalman “unscented” (UKF) en el pronóstico de series de tiempo. Concluyeron que las dos combinaciones tuvieron resultados con adecuada precisión, sin embargo, EKF genera menor carga computacional. Por otra parte, Saba, Cedeño, y Ortega (2010) evaluaron una red neuronal entrenada mediante EKF y PF para pronosticar series de precipitaciones. Donde concluyeron que el EKF permite obtener resultados más precisos que el algoritmo de entrenamiento por retro propagación.

La principal ventaja de las técnicas de inteligencia artificial es la facilidad para tratar con relaciones no lineales entre variables (Han & Coulibaly, 2017; Pujol Reig, 2009; Zealand et al., 1999). Además, el entrenamiento con el filtro de

Kalman le transfiere dinamismo y precisión, sin embargo, son temas poco explorados (Pujol Reig, 2009). Además, el Filtro de Kalman representa ventajas al momento de actualizarse de manera continua sin necesidad de un proceso de entrenamiento a través de un algoritmo adicional, esto facilita su implementación y ofrece la opción de mantenerse actualizado en función de los cambios que se dan en la serie bajo estudio, asimismo, las variaciones del filtro de Kalman como EnKF y EKF son capaces de estimar estados en sistemas dinámicos no lineales (Evensen, 2009; Welch & Bishop, 2006) y de esta forma se pueden aplicar en hidrología.

## 2.14 Filtro de Kalman e hidrología

### 2.14.1 Cuenca hidrográfica

Una cuenca hidrográfica se configura como un sistema dinámico no lineal, donde tiene entradas, interacciones y salidas (Kundzewicz & Napiórkowski, 1986; Sun, Seidou, Nistor, & Liu, 2016). En el caso de pronóstico de caudales, se tiene como principales variables de entrada la precipitación y los caudales mismos, las cuales se repiten de acuerdo con la cantidad de estaciones meteorológicas distribuidas en el área de la cuenca, es decir, se tienen tantas series como estaciones meteorológicas en la cuenca. La variable de salida corresponde a los caudales reportados en la desembocadura del cauce principal de la cuenca. Como el algoritmo del filtro de Kalman se ejecuta en línea, es decir, en tiempo real, es necesario que el registro de nuevas mediciones se pueda transmitir para actualizar las estimaciones.

En una cuenca las interrelaciones entre las variables de entrada y la variable de salida se denomina función de respuesta de la cuenca o bien Hidrograma Unitario Instantáneo (HUI) (Valdés et al., 1980), donde cada barra del hidrograma equivale a la cantidad de datos previos o términos autorregresivos que ingresan como parte de cada variable al algoritmo. Mediante el filtro de Kalman se estiman los parámetros o variables de estado que representan la influencia de cada variable de entrada sobre la variable de salida. Esta configuración es la

frecuentemente utilizada en hidrología, pero, desaprovecha la característica de recursividad del Filtro de Kalman (Alvarado-Hernández et al., 2020; González-Leiva et al., 2015; Morales-Velázquez et al., 2014; Valdés et al., 1980).

### 2.14.2 Aplicaciones del Filtro del Kalman

En hidrología el filtro de Kalman se ha utilizado principalmente en la estimación de la humedad de suelo y pronostico climáticos de precipitación (Medina González et al., 2015), además, se han realizado evaluaciones sobre el desempeño del Filtro de Kalman discreto para pronóstico de caudales y avenidas con 1, 24 y 48 horas de anticipación. Concluyendo que hace predicciones aceptables y que podrían ser mejoradas si se tiene buena instrumentación en la cuenca (González-Leiva et al., 2015; Morales-Velázquez et al., 2014). No obstante, no es conveniente su utilización en conjunto con modelos distribuidos o semidistribuidos (Sun, Seidou, Nistor, & Liu, 2016).

Considerando que los fenómenos atmosféricos no son lineales (Cuo et al., 2011; Lorenz, 1969), se infiere que con métodos especializados en tratar la no linealidad se puede obtener resultados con mayor precisión y disminuir la resolución temporal a un valor inferior al tiempo de concentración de la cuenca. EnKF es frecuentemente utilizado en la estimación de parámetros hidrológicos (S. Wang et al., 2017) y se han realizado diversos avances para darle mayor precisión, la estimación dual de parámetros y estados (Moradkhani et al., 2005), estimación de varios estados simultáneos (D. Wang et al., 2009), estimación dual de parámetros en caudales y humedad del suelo (Samuel et al., 2014), detección de patrones temporales en caudales (Pathiraja et al., 2016), Integración de varias fuentes de datos para el pronóstico en cuencas urbanas (Leach et al., 2018), por mencionar algunos.

Por otra parte EnKF, no cuenta con evaluaciones en cuencas tropicales dominadas por lluvia escurrimiento, no obstante, se ha integrado con modelos de escorrentía como GR4J, HyMod, MAC-HBV, and SAC-SMA (Leach et al., 2018), estimación de la evapotranspiración con el modelo DTVGM-ET (Zou et al.,

2017b), identificación de la dinámica temporal de parámetros de modelos de simulación de caudales y evapotranspiración (Xiong et al., 2019), simulación de escurrimiento utilizando SWAT (Sun, Seidou, Nistor, & Liu, 2016; Sun, Seidou, Nistor, Goïta, et al., 2016), distribución de la humedad del suelo (Abaza et al., 2014; Brandhorst et al., 2017; Medina González et al., 2015; Zhu et al., 2017), evaluación de calidad de aguas de embalses, represas y lagos (Zhang et al., 2020), exploración 3D de aguas subterráneas con base en electro conductividad (Javaheri et al., 2019). Como se puede evidenciar, el EnKF es aplicado en diversas áreas, sin embargo, no se cuenta con trabajos sobre pronóstico de caudales sin integrar modelos distribuidos.

Al igual que un modelo estadístico o de simulación, para que el filtro de Kalman realice estimaciones con precisión es indispensable que se tenga variables de entrada que sean representativas de la cuenca y que guarden alta correlación o autocorrelación con la variable de salida. Las variables continuas son las más adecuadas para actualizar el algoritmo, es decir, para el pronóstico de caudales las variables como temperatura, presión atmosférica, vientos y los caudales en otras ubicaciones ofrecen mejores probabilidades de ajuste adecuado, en comparación con el uso de la precipitación.

La precipitación es una variable discreta con cambios pronunciados entre pasos de tiempo consecutivos, con predominancia de valores cero y cercanos a cero, resultando en promedios bajos. Los cambios bruscos entre cada registro de variables discretas dificultan la determinación de una tendencia que permita realizar pronósticos a corto plazo. En cambio, las variables continuas hacen cambios de manera progresiva definiendo una tendencia que se aprovecha para actualizar los modelos de simulación. En el caso del filtro de Kalman los cambios bruscos que caracterizan las series discretas impiden que haya convergencia y persisten notables diferencias entre los valores observados y pronosticados.

Por otra parte, para el pronóstico de caudales se ha probado otros métodos recursivos como el filtro de máxima verosimilitud en conjuntos (Maximun Likelihood Emsemble Filter – MLEF). Rafieeinasab et al. (2014) comparo el MLEF

con el EnKF y encontró que MLEF es superior al EnKF en rendimiento y confiabilidad, además, tiene menos costo computacional, debido a que EnKF requiere mayor cantidad de miembros.

## 2.15 Rutinas de código

El filtro de Kalman es un método recursivo de asimilación de datos lo que le confiere ventajas como flexibilidad para la implementación en sistemas de baja capacidad de almacenamiento y procesamiento. En este sentido y para promover la implementación del algoritmo las rutinas de código del DKF, EnKF y modelos autorregresivos se hizo con el lenguaje R (R Core Team, 2020) versión 4.0.2 apoyado con el editor RStudio (RStudio Team, 2020). El lenguaje R es gratuito y de código abierto y guarda cierta similitud con C++, lo que facilita la implementación en placas Arduino, o bien utilizar Python en placas Raspberry, con el objetivo de ganar adoptabilidad gracias a su bajo coste y facilidad relativa para implementación de aplicaciones prácticas en lugares con escaso acceso a sistemas de cómputo avanzados y financiamiento.

La codificación del filtro del Kalman discreto se hizo con base en Kalman (1960) y Welch & Bishop (2006), por su parte el filtro de Kalman de conjuntos se hizo con base en Evensen (1994) y Gillijns et al (2006). Los modelos autorregresivos se programaron con base a Ljung (1999).

## 2.16 Literatura citada


Abaza, M., Anctil, F., Fortin, V., & Turcotte, R. (2014). Sequential streamflow assimilation for short-term hydrological ensemble forecasting. Journal of Hydrology, 519(PD), 2692–2706. https://doi.org/10.1016/j.jhydrol.2014.08.038

Abaza, M., Anctil, F., Fortin, V., & Turcotte, R. (2015). Exploration of sequential streamflow assimilation in snow dominated watersheds. Advances in Water Resources, 86, 414–424. https://doi.org/10.1016/j.advwatres.2015.10.008

Alvarado-Hernández, L., Ibanez-Catillo, L., Ruiz-García, A., González-Leiva, F., & Vázquez-Peña, M. (2020). Pronóstico horario de caudales mediante filtro de kalman discreto en el Río Huaynamota, Nayarit, México. Agrociencia, 54, 295–312.

Aparicio, F. (2017). Fundamentos de hidrología de superficie. Limusa.

Archer, E., Park, I. M., Buesing, L., Cunningham, J., & Paninski, L. (2015). Black box variational inference for state space models. 6th International Conference on Learning Representations, 1–11. http://arxiv.org/abs/1511.07367

Bader, K., Lussier, B., & Schön, W. (2017). A fault tolerant architecture for data fusion: A real application of Kalman filters for mobile robot localization. Robotics and Autonomous Systems, 88, 11–23. https://doi.org/10.1016/j.robot.2016.11.015

Bishop, C. H., Etherton, B. J., & Majumdar, S. J. (2001). Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. Monthly Weather Review, 129(3), 420–436. https://doi.org/10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2

Box, G., Jenkins, G., Reinsel, G., & Ljung, G. (2016). Time series analysis: Forecasting and control (5th ed.). Wiley.

Brammer, R. (1974). Real-Time shipboard orbit determination using Kalman Filtering Techniques. IEEE Transactions on Aerospace and Electronic Systems, AES-10(4), 492–496. https://doi.org/10.1109/TAES.1974.307803

Brandhorst, N., Erdal, D., & Neuweiler, I. (2017). Soil moisture prediction with the ensemble Kalman filter: Handling uncertainty of soil hydraulic parameters. Advances in Water Resources, 110(February), 360–370. https://doi.org/10.1016/j.advwatres.2017.10.022

Bravo, L. (2016). Implementacion de redes neuronales para generar oleaje utilizando datos de viento en la costa de Sisal, Yucatan. Universidad Nacional Autonoma de Mexico.

Cloke, H. L., & Pappenberger, F. (2009). Ensemble flood forecasting: A review. Journal of Hydrology, 375(3–4), 613–626. https://doi.org/10.1016/j.jhydrol.2009.06.005

Cuo, L., Pagano, T. C., & Wang, Q. J. (2011). A Review of Quantitative Precipitation Forecasts and Their Use in Short- to Medium-Range Streamflow Forecasting. Journal of Hydrometeorology, 12(5), 713–728. https://doi.org/10.1175/2011JHM1347.1

DeChant, C. M., & Moradkhani, H. (2012). Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting. Water Resources Research, 48(4), 1–15. https://doi.org/10.1029/2011WR011011

Dumedah, G., & Coulibaly, P. (2013). Evaluating forecasting performance for data assimilation methods: The ensemble Kalman filter, the particle filter, and the evolutionary-based assimilation. Advances in Water Resources, 60, 47–63. https://doi.org/10.1016/j.advwatres.2013.07.007

Elsafi, S. H. (2014). Artificial Neural Networks (ANNs) for flood forecasting at Dongola Station in the River Nile, Sudan. Alexandria Engineering Journal, 53(3), 655–662. https://doi.org/10.1016/j.aej.2014.06.010

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. Journal of Geophysical Research, 99(C5), 10143. https://doi.org/10.1029/94jc00572

Evensen, G. (2003). The Ensemble Kalman Filter: theoretical formulation and practical implementation. Ocean Dynamics, 53(4), 343–367. https://doi.org/10.1007/s10236-003-0036-9

Evensen, G. (2009). Data Assimilation: The ensemble Kalman filter (2nd ed.). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-03711-5

Evensen, G., & van Leeuwen, P. J. (2000). An ensemble Kalman smoother for nonlinear dynamics. Monthly Weather Review, 128(6), 1852–1867. https://doi.org/10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2

Fan, Y. R., Huang, G. H., Baetz, B. W., Li, Y. P., Huang, K., Chen, X., & Gao, M. (2017). Development of integrated approaches for hydrological data assimilation through combination of ensemble Kalman filter and particle filter methods. Journal of Hydrology, 550, 412–426. https://doi.org/10.1016/j.jhydrol.2017.05.010

Garces, L., & Gomez, O. (2004). Inferencia estadística utilizando redes neuronales artificiales. Scientia et Technica, 10(26), 31–36.

Gillijns, S., Mendoza, O. B., Chandrasekar, J., De Moor, B. L. R., Bernstein, D. S., & Ridley, A. (2006). What is the ensemble Kalman filter and how well does it work? 2006 American Control Conference, 6 pp. https://doi.org/10.1109/ACC.2006.1657419

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. Tecnologia y Ciencias Del Agua, 6(4), 5–24.

Gouweleeuw, B. T., Thielen, J., Franchello, G., De Roo, A. P. J., & Buizza, R. (2005). Flood forecasting using medium-range probabilistic weather prediction. Hydrology and Earth System Sciences, 9(4), 365–380. https://doi.org/10.5194/hess-9-365-2005

Han, S., & Coulibaly, P. (2017). Bayesian flood forecasting methods: A review. Journal of Hydrology, 551, 340–351. https://doi.org/10.1016/j.jhydrol.2017.06.004

Hapuarachchi, H. A. P., Wang, Q. J., & Pagano, T. C. (2011). A review of advances in flash flood forecasting. Hydrological Processes, 25(18), 2771–2784. https://doi.org/10.1002/hyp.8040

Herr, H. D., & Krzysztofowicz, R. (2010). Bayesian ensemble forecast of river stages and ensemble size requirements. Journal of Hydrology, 387(3–4), 151–164. https://doi.org/10.1016/j.jhydrol.2010.02.024

Infante, S., & Zaraté, G. (2012). Metodos estadisticos: un enfoque interdisciplinario. La Gaya Ciencia.

IPCC. (2012). Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. In C. B. Field, V. Barros, T. F. Stocker, & Q. Dahe (Eds.), Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. Cambridge University Press. https://doi.org/10.1017/CBO9781139177245

Javaheri, A., Babbar-Sebens, M., Miller, R. N., Hallett, S. L., & Bartholomew, J. L. (2019). An adaptive ensemble Kalman filter for assimilation of multi-sensor, multi-modal water temperature observations into hydrodynamic model of shallow rivers. Journal of Hydrology, 572, 682–691. https://doi.org/10.1016/j.jhydrol.2019.03.036

Jonsson, G., & Palsson, O. P. (1994). An Application of Extended Kalman Filtering to Heat Exchanger Models. Journal of Dynamic Systems, Measurement, and Control, 116(2), 257. https://doi.org/10.1115/1.2899218

Julier, S. J., & Uhlmann, J. K. (1997). A new extension of the Kalman filter to nonlinear systems (I. Kadar (ed.); p. 182). https://doi.org/10.1117/12.280797

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. Journal of Basic Engineering, 82(Series D), 35–45.

Kaps, M., & Lamberson, W. (2004). Biostatistics for animal science. CABI Publishing.

Khan, E., & Khan, F. (2012). Summary for Policymakers. International Journal of Advanced Computer Science and Applications, 3(6), 12–15. https://www.cambridge.org/core/product/identifier/CBO9781107415324A009/type/book_part

Kotecha, J. H., & Djuric, P. M. (2003). Gaussian particle filtering. Proceedings of the 11th IEEE Signal Processing Workshop on Statistical Signal Processing (Cat. No.01TH8563), 429–432. https://doi.org/10.1109/SSP.2001.955314

Krener, A. J., & Duarte, A. (1996). A hybrid computational approach to nonlinear estimation. Proceedings of 35th IEEE Conference on Decision and Control, 2, 1815–1819. https://doi.org/10.1109/CDC.1996.572831

Kundzewicz, Z. W., & Napiórkowski, J. J. (1986). Nonlinear models of dynamic hydrology. Hydrological Sciences Journal, 31(2), 163–185. https://doi.org/10.1080/02626668609491038

Lawson, W. G., & Hansen, J. A. (2004). Implications of Stochastic and Deterministic Filters as Ensemble-Based Data Assimilation Methods in Varying Regimes of Error Growth. Monthly Weather Review, 132(8), 1966–

1981. https://doi.org/10.1175/1520-0493(2004)132<1966:IOSADF>2.0.CO;2

Leach, J. M., Kornelsen, K. C., & Coulibaly, P. (2018). Assimilation of near-real time data products into models of an urban basin. Journal of Hydrology, 563(April), 51–64. https://doi.org/10.1016/j.jhydrol.2018.05.064

Liu, Weerts, A. H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenberg, D., Smith, P., van Dijk, A. I. J. M., van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O., & Restrepo, P. (2012). Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. Hydrology and Earth System Sciences, 16(10), 3863–3887. https://doi.org/10.5194/hess-16-3863-2012

Ljung, L. (1999). System Identification: Theory for the User. Prentice Hall PTR. https://books.google.com.mx/books?id=nHFoQgAACAAJ

Lopez, J. A., & Caicedo, E. F. (2009). Una aproximacion practica a las redes neuronales artificiales. Universidad del Valle.

Lorenz, E. N. (1969). The predictability of a flow which possesses many scales of motion. Tellus, 21(3), 289–307. https://doi.org/10.3402/tellusa.v21i3.10086

Martin, B., & Sanz, A. (2017). Redes neuronales y sistemas borrosos (15th ed.). Alfaomega.

Maxwell, D. H., Jackson, B. M., & McGregor, J. (2018). Constraining the ensemble Kalman filter for improved streamflow forecasting. Journal of Hydrology, 560, 127–140. https://doi.org/10.1016/j.jhydrol.2018.03.015

Maybeck, P. (1979). Stochastic models , estimation, and control. Academic Press.

Medina González, H., Hernández Pereira, Y., Santiago Piloto, A. B., & Lau Quan, A. (2015). Modelación de perfil de humedad de suelos empleando un filtro de Kalman de Monte-Carlo. Revista Ciencias Técnicas Agropecuarias, 24(2), 31–37.

Mejía, D., & Eliana, T. (2004). Pronóstico de ventas usando redes neuronales. Scientia et Technica, 10(26), 25–30.

Moradkhani, H., Sorooshian, S., Gupta, H. V., & Houser, P. R. (2005). Dual state-parameter estimation of hydrological models using ensemble Kalman filter. Advances in Water Resources, 28(2), 135–147. https://doi.org/10.1016/j.advwatres.2004.09.002

Morales-Velázquez, M. I., Aparicio, J., & Valdes, J. B. (2014). Pronóstico de avenidas utilizando el filtro de kalman discreto. Tecnologia y Ciencias Del Agua, 5(2), 85–110.

Muluye, G. Y. (2011). Improving long-range hydrological forecasts with extended Kalman filters. Hydrological Sciences Journal, 56(7), 1118–1128. https://doi.org/10.1080/02626667.2011.608068

Naghettini, M., & De Andrade, É. (2007). Hidrologia estatística (1st ed.). CPRM - Serviço Geológico do Brasil.

Odry, Á., Fullér, R., Rudas, I. J., & Odry, P. (2018). Kalman filter for mobile-robot attitude estimation: Novel optimized and adaptive solutions. Mechanical Systems and Signal Processing, 110, 569–589. https://doi.org/10.1016/j.ymssp.2018.03.053

Pathiraja, S., Marshall, L., Sharma, A., & Moradkhani, H. (2016). Hydrologic modeling in dynamic catchments: A data assimilation approach. Water Resources Research, 52(5), 3350–3372. https://doi.org/10.1002/2015WR017192

Pujol Reig, L. (2009). Prediccion de caudales en tiempo real en grandes cuencas utilizando redes neuronales artificiales. Universidad Politecnica de Madrid.

Quadros, R. S. de, Harter, F. P., Buske, D., & Pereira, L. S. (2016). Data assimilation by ensemble Kalman filter with the Lorenz equations. Ciência e Natura, 38(July), 190. https://doi.org/10.5902/2179460X20158

Quiroz, K., Collischonn, W., & Paiva, R. C. D. de. (2019). Data assimilation using the ensemble Kalman filter in a distributed hydrological model on the Tocantins River, Brasil. RBRH, 24, 1–15. https://doi.org/10.1590/2318-0331.241920180031

R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. http://www.R-project.org/. https://www.r-project.org/

Rafieeinasab, A., Seo, D. J., Lee, H., & Kim, S. (2014). Comparative evaluation of maximum likelihood ensemble filter and ensemble Kalman filter for real-time assimilation of streamflow data into operational hydrologic models. Journal of Hydrology, 519(PD), 2663–2675. https://doi.org/10.1016/j.jhydrol.2014.06.052

Rajagopalan, B., Salas, J., & Lall, U. (2010). Stochastic methods for modeling precipitation and streamflow. In B. Sivakumar & R. Berndtsson (Eds.), Advances in Data-Based Approaches for Hydrologic Modeling and Forecasting (pp. 17–52). WORLD SCIENTIFIC. https://doi.org/10.1142/7783

Reichle, R. H. (2008). Data assimilation methods in the Earth sciences. Advances in Water Resources, 31(11), 1411–1418. https://doi.org/10.1016/j.advwatres.2008.01.001

Reichle, R. H., Walker, J. P., Koster, R. D., & Houser, P. R. (2002). Extended versus Ensemble Kalman Filtering for Land Data Assimilation. Journal of Hydrometeorology, 3(6), 728–740. https://doi.org/10.1175/1525-7541(2002)003<0728:EVEKFF>2.0.CO;2

RStudio Team. (2020). RStudio: Integrated Development Environment for R. http://www.rstudio.com/

Saba, I., Cedeño, F., & Ortega, J. (2010). Filtros con aprendizaje de parametros para optimizar modelos de redes nuronales en la prediccion de series de precipitaciones. Ingenieria Industiral. Actualidad y Nuevas Tendencias, 2(5), 37–56.

Samuel, J., Coulibaly, P., Dumedah, G., & Moradkhani, H. (2014). Assessing model state and forecasts variation in hydrologic data assimilation. Journal of Hydrology, 513, 127–141. https://doi.org/10.1016/j.jhydrol.2014.03.048

Shumway, R. H., & Stoffer, D. S. (2017). Time Series Analysis and Its Applications. Springer International Publishing. https://doi.org/10.1007/978-3-319-52452-8

Sun, L., Seidou, O., Nistor, I., Goïta, K., & Magagi, R. (2016). Simultaneous assimilation of in situ soil moisture and streamflow in the SWAT model using the Extended Kalman Filter. Journal of Hydrology, 543, 671–685. https://doi.org/10.1016/j.jhydrol.2016.10.040

Sun, L., Seidou, O., Nistor, I., & Liu, K. (2016). Review of the Kalman-type hydrological data assimilation. Hydrological Sciences Journal, 61(13), 2348–2366. https://doi.org/10.1080/02626667.2015.1127376

Te Chow, V., Maidment, D. R., Mays, L. W., & Saldarriaga, J. G. (1996). Hidrología aplicada. McGraw-Hill.

Torres, N., Hernandez, C., & Pedraza, L. (2011). Redes neuronales y predicción de tráfico. Tecnura, 15(29), 90–97.

Valdés, J., Mejía, J., & 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.

van der Merwe, R., & Wan, E. A. (2001). The Unscented Kalman Filter. In S. Haykin (Ed.), Kalman Filtering and Neural Networks (Vol. 6, pp. 221–279). JOHN WILEY & SONS, INC. https://doi.org/10.1002/0471221546.ch7

Vrugt, J. A., & Robinson, B. A. (2007). Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging. Water Resources Research, 43(1), 1–15. https://doi.org/10.1029/2005WR004838

Wang, D., Chen, Y., & Cai, X. (2009). State and parameter estimation of hydrologic models using the constrained ensemble Kalman filter. Water Resources Research, 45(11), 1–13. https://doi.org/10.1029/2008WR007401

Wang, S., Huang, G. H., Baetz, B. W., Cai, X. M., Ancell, B. C., & Fan, Y. R. (2017). Examining dynamic interactions among experimental factors influencing hydrologic data assimilation with the ensemble Kalman filter. Journal of Hydrology, 554, 743–757. https://doi.org/10.1016/j.jhydrol.2017.09.052

Welch, G., & Bishop, G. (2006). An Introduction to the Kalman Filter. In University of North Carolina at Chapel Hil (Issue 1).

Whitaker, J. S., & Hamill, T. M. (2002). Ensemble Data Assimilation without Perturbed Observations. Monthly Weather Review, 130(7), 1913–1924. https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2

Willner, D., Chang, C., & Dunn, K. (1976). Kalman filter algorithms for a multi-sensor system. 1976 IEEE Conference on Decision and Control Including the 15th Symposium on Adaptive Processes, 570–574. https://doi.org/10.1109/CDC.1976.267794

WMO. (2011). Manual on flood forecasting and warming - WMO-No. 1072. World Metereological Organization.

Wu, X., & Wang, Y. (2012). Extended and Unscented Kalman filtering based feedforward neural networks for time series prediction. Applied Mathematical Modelling, 36(3), 1123–1131. https://doi.org/10.1016/j.apm.2011.07.052

Xiong, M., Liu, P., Cheng, L., Deng, C., Gui, Z., Zhang, X., & Liu, Y. (2019). Identifying time-varying hydrological model parameters to improve simulation efficiency by the ensemble Kalman filter: A joint assimilation of streamflow and actual evapotranspiration. Journal of Hydrology, 568, 758–768. https://doi.org/10.1016/j.jhydrol.2018.11.038

Zealand, C. M., Burn, D. H., & Simonovic, S. P. (1999). Short term streamflow forecasting using artificial neural networks. Jounal of Hydrology, 214, 32–48.

Zhang, C., Yan, Q., Kuczyńska-Kippen, N., & Gao, X. (2020). An Ensemble Kalman Filter approach to assess the effects of hydrological variability, water diversion, and meteorological forcing on the total phosphorus concentration in a shallow reservoir. Science of The Total Environment, 724, 138215. https://doi.org/10.1016/j.scitotenv.2020.138215

Zhu, P., Shi, L., Zhu, Y., Zhang, Q., Huang, K., & Williams, M. (2017). Data assimilation of soil water flow via ensemble Kalman filter: Infusing soil moisture data at different scales. Journal of Hydrology, 555, 912–925. https://doi.org/10.1016/j.jhydrol.2017.10.078

Zou, L., Zhan, C., Xia, J., Wang, T., & Gippel, C. J. (2017a). Implementation of evapotranspiration data assimilation with catchment scale distributed hydrological model via an ensemble Kalman Filter. Journal of Hydrology, 549, 685–702. https://doi.org/10.1016/j.jhydrol.2017.04.036

Zou, L., Zhan, C., Xia, J., Wang, T., & Gippel, C. J. (2017b). Implementation of evapotranspiration data assimilation with catchment scale distributed hydrological model via an ensemble Kalman Filter. Journal of Hydrology, 549, 685–702. https://doi.org/10.1016/j.jhydrol.2017.04.036

# 3 COMPARACIÓN DEL FILTRO DE KALMAN DISCRETO CON EL FILTRO DE KALMAN DE CONJUNTOS PARA PRONÓSTICO DE CAUDALES HORARIOS EN EL RÍO HUAYNAMOTA, NAYARIT, MÉXICO

**Comparison of discrete and ensemble Kalman filter for hourly streamflow forecasting in Huaynamota river, Nayarit, México.**

Ildefonso Narvaez-Ortiz[1], Laura Ibáñez-Castillo[1], Ramon Arteaga-Ramírez[1], Mario Vázquez-Peña[1], Carlos Cintora-González[1]

[1] Posgrado en Ingeniería Agrícola y Uso Integral del Agua Universidad Autónoma Chapingo. (UACH). Km 38.5 Carretera México-Texcoco, Código postal. 56230. Estado de México.

Autor para correspondencia: libacas@gmail.com

## 3.1 Resumen

La asimilación de datos integrada para el pronóstico de caudales puede brindar flexibilidad y reducción de errores sistemáticos en los modelos. En este trabajo se evalúan la capacidad predictiva del filtro de Kalman discreto, filtro de Kalman de conjuntos y su integración, utilizando registros horarios de caudal de las estaciones Chapalagana y Platanitos ubicadas sobre el rio Huaynamota, región hidrológica 12. La cuenca se ubica al noroeste de la República mexicana y se comparte entre los estados de Durango, Nayarit, Zacatecas y Jalisco. Para el análisis se utilizaron series con 1360 registros horarios del año 2017 comprendidos entre el 02 de agosto a las 9:00 horas hasta el 28 de septiembre a las 0:00 horas. Se realizaron pronósticos a 1, 2, 3, 4, 5 y 6 pasos hacia adelante, combinados con tamaños de conjunto de 5, 8, 10, 20, 50 y 100 miembros utilizando caudales de la estación Platanitos como variable exógena. El ajuste entre la serie observada y las pronosticadas se estimó mediante el coeficiente de Nash-Sutcliffe y la Raíz del Cuadrado Medio del Error para determinar que el filtro de Kalman discreto alcanza mejor ajuste y actualización con base en el tiempo de retraso entre series. El filtro de Kalman de conjuntos genera un suavizado de la serie pronosticada, y al igual que la integración de filtros aumenta el efecto de desplazamiento de la serie pronosticada. El filtro de Kalman discreto alcanza ajuste superior a ARX y a la combinación ARX-DKF.

**Palabras clave**: Filtro de Kalman, conjuntos, modelos autorregresivos, pronósticos de caudales a corto plazo.

## 3.2 Abstract

Integrated data assimilation for flow forecasting can provide flexibility and reduce systematic errors in the models. In this work we evaluate the predictive capacity of the discrete Kalman filter, Ensemble Kalman filter and its integration, using hourly flow records from Chapalagana and Platanitos stations located on the Huaynamota river, hydrological region 12. The basin is located in the northwest of the Mexican Republic and is shared between the states of Durango, Nayarit, Zacatecas and Jalisco. For the analysis, series with 1360 data from 2017 were used, from August 2nd at 9:00 a.m. to September 28th at 0:00 a.m. Forecasts were evaluated at 1, 2, 3, 4, 5 and 6 steps forward, combined with set sizes of 5, 8, 10, 20, 50 and 100 members, using measurements at the Platanitos station as an exogenous variable. The fit between observed and predicted series was estimated using the Nash-Sutcliffe coefficient and the Mean Square Root of Error to determine that the discrete Kalman filter achieves better fit and update based on the time delay between series. The Ensemble Kalman filter generates smoothing of the predicted series, and like the integration of filters increases the displacement effect of the predicted series. The discrete Kalman filter achieves superior adjustment to ARX and the ARX-DKF combination

**Keywords**: Kalman filter, ensembles, autoregressive models, short-term streamflow forecasting.

## 3.3 Introducción

La asimilación de datos para la estimación de estados es esencial en las aplicaciones de pronóstico en hidrología (Liu et al., 2012), ya que, mediante la actualización cíclica que llevan a cabo se puede disminuir los errores que se acumulan en los modelos (Clark et al., 2008; Maxwell et al., 2018). El Filtro de Kalman de Conjuntos (EnKF) (Evensen, 1994) es uno de los algoritmos ampliamente utilizado como método recursivo en hidrología (Liu & Gupta, 2007; Maxwell et al., 2018; Sun, Seidou, Nistor, & Liu, 2016), es una extensión del Filtro de Kalman discreto (Kalman, 1960) para tratar sistemas dinámicos no lineales (Evensen, 1994, 2003), y se ha utilizado entre otros para el pronóstico de caudales (Maxwell et al., 2018), evapotranspiración (Zou et al., 2017b) y humedad del suelo (Brandhorst et al., 2017; Meng et al., 2017), además, se ha

integrado con modelos hidrológicos distribuidos como TopNet, Hydrotel y MGB-IPH (Abaza et al., 2015; Clark et al., 2008; Quiroz et al., 2019).

En los sistemas de cuenca hidrográfica se puede identificar la relación entre los caudales medidos en diferentes posiciones y un punto de interés, a lo que se denomina función de respuesta (Valdés et al., 1980). Con base en esta relación se pueden hacer pronósticos a corto plazo que se actualizan constantemente y de forma indefinida mientras el fenómeno bajo estudio persista.

En este trabajo se evalúa la capacidad de pronosticar del filtro de Kalman discreto (DKF) (Kalman, 1960), el filtro de Kalman de conjuntos (EnKF) (Evensen, 1994) y la integración del filtro de Kalman discreto (DKF) con el filtro de Kalman de conjuntos (EnKF) para el pronóstico de uno hasta seis pasos hacia delante de los caudales del río Huaynamota en la estación Chapalagana. El DKF es implementado para estimar el vector de estado (Hidrograma Unitario Instantáneo – HUI) y el EnKF hace estimaciones escalares de caudal. La integración se hace mediante la ecuación de estados del EnKF.

### 3.4 Materiales y métodos

El área de estudio corresponde a la delimitación aguas arriba a partir de la estación Chapalagana sobre el río Huaynamota (Río Chapalagana o Río Atengo), se ubica al noroeste de la República mexicana y se comparte entre los estados de Durango, Nayarit, Zacatecas y Jalisco (INEGI, 2010)(Figura 1), geográficamente se ubica entre los -104°33’34.16” y los -103°27’29.84” de longitud, y entre los 23°28’50.05” y los 21°23’57.62” de latitud, tiene un área de 12 075.7 $km^2$, la altitud máxima es de 3 147 msnm (metros sobre el nivel del mar) y la mínima en la estación Chapalagana de 219 msnm, el tiempo de concentración es de 39.88 horas, la precipitación media anual es de 707 mm y la temperatura media anual de 17.9 °C (SMN, 2019).

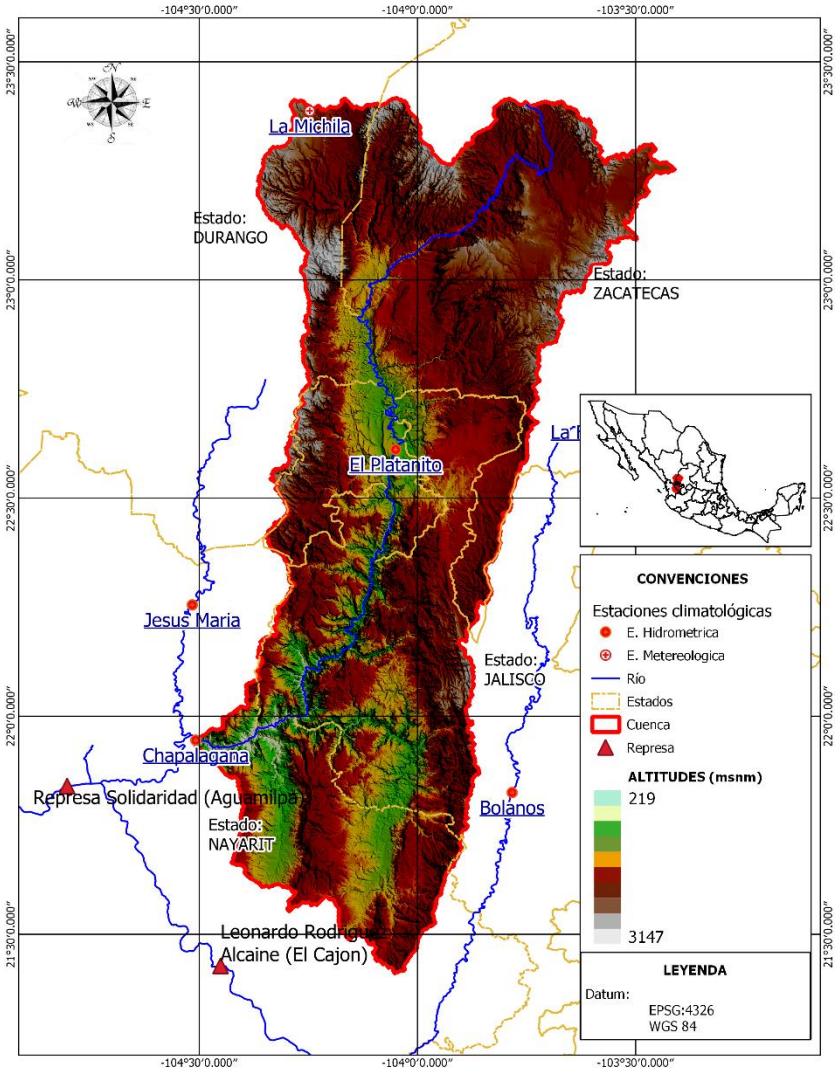



Figura 3.1. Localización del área de estudio.

La principal estructura hidráulica con fines de generación de electricidad que se encuentra aguas abajo del área de estudio y sobre el rio Lerma-Santiago Pacifico es la represa Solidaridad, comúnmente conocida como Aguamilpa, ubicada en los -104°48'10.55" de longitud y los 21°50'22.74" de latitud, y tiene una capacidad de 5 540 millones de $m^3$, para generar 960 MW de electricidad (CONAGUA, 2008). La presa Aguamilpa se encuentra aproximadamente a 90 km del Océano Pacífico, donde el río Santiago desemboca frente a las costas del Estado de Nayarit.

Se realizó la evaluación de la capacidad predictiva del algoritmo del Filtro de Kalman Discreto (DKF) (Kalman, 1960), el filtro de Kalman de conjuntos (EnKF)

(Evensen, 1994), la integración del DKF y EnKF (DKF-EnKF), y el modelo autorregresivo con variable exógena de primer orden (AR(1,1)), para realizar el pronóstico a 1, 2, 3, 4, 5 y 6 horas (pasos $L$) hacia adelante de caudales horarios en la estación Chapalagana, con base en los caudales de la estación Platanito, localizada aguas arriba, como variable exógena. Se utilizaron series horarias de caudal del año 2017 entre el 02 de agosto a las 9:00 horas hasta el 28 de septiembre a las 0:00 horas, para un total de 1360 registros suministrados por la CFE (Comisión Federal de Electricidad). Para el análisis de sensibilidad del tamaño del conjunto se utilizó 5, 8, 10, 20, 50 y 100 miembros que se combinaron con los seis pasos.

La implementación del DKF, EnKF, DKF-EnKF y ARX(1,1) se hizo mediante rutinas en R (R Core Team, 2020) que generan los pronósticos a seis pasos con DKF y ARX, y con 36 combinaciones entre pasos por tamaños de conjunto en EnKF y DKF-EnKF. El DKF se implementó para estimar el vector de estado que equivale a la función de respuesta de la cuenca o Hidrograma Unitario Instantáneo (HUI) (Valdés et al., 1980) y el EnKF realiza estimaciones escalares del caudal (Evensen, 2009), es decir, con el DKF se estiman los valores correspondientes a las columnas del HUI que se pasan a multiplicar por los valores de caudal, por su parte el EnKF estima directamente valores de caudal. En los tres casos con filtros se consideró las ultimas observaciones de cada variable, es decir, un retraso autorregresivo (Valdés et al., 1980). El modelo ARX se implementó de forma recursiva con base en una fracción de serie con 100 registros.

La cantidad adecuada de miembros en los conjuntos de EnKF y DKF-EnKF se determinó mediante análisis de sensibilidad, con base en la Raíz del Cuadrado Medio del Error (RMSE) (Quiroz et al., 2019). La generación del ruido blanco (simulación de Monte Carlo) se hizo con el paquete mvtnorm (Normal multivariante y distribución t) (Genz & Bretz, 2009). En los tres algoritmos evaluados la varianza $Q$ se supone constante (Simon, 2001), de valor cero

(Morales-Velázquez et al., 2014) y $R$ con valor cercano a cero (0.01) para dar flexibilidad a la convergencia del algoritmo (Welch & Bishop, 2006).

El ajuste de las series pronosticadas se evaluó mediante el coeficiente de Nash-Sutcliffe (Nash & Sutcliffe, 1970) y la Raíz del Cuadrado Medio del Error (RMSE) (Morales-Velázquez et al., 2014). Los supuestos de normalidad de errores del Filtro de Kalman se verificaron mediante gráficos comparados con la curva normal estandarizada (González-Leiva et al., 2015). Los valores atípicos y su ubicación se determinó con base en los residuales estandarizados (Cryer & Chan, 2008).

### 3.4.1 Filtro de Kalman Discreto

Es un algoritmo que permite la identificación de sistemas dinámicos lineales como un estimador óptimo de estados mediante un proceso recursivo (Kalman, 1960) (Figura 2). El DKF sirve de base para los algoritmos que tratan sistemas no lineales (Welch & Bishop, 2006)

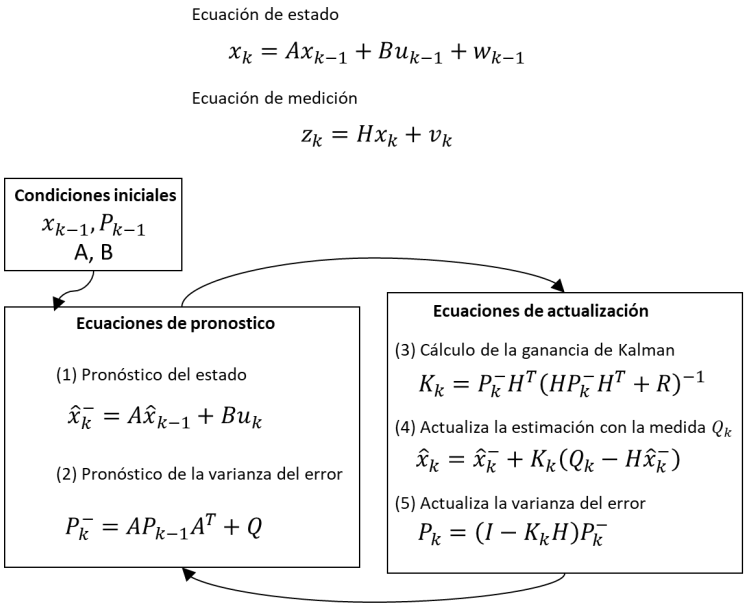



Figura 3.2. Algoritmo del filtro de Kalman discreto (Welch & Bishop, 2006).

La ecuación de estado tiene como entrada dos series horarias ($n$) de caudal de la estación Chapalagana y Platanitos. La matriz A ($n\ x\ n$) y B ($n\ x\ 1$) relacionan el

estado del tiempo $k-1$ con el estado actual en el tiempo $k$. Debido a que en el curso del rio no se encuentran estructuras hidráulicas como represas que alteren de forma repentina los volúmenes de caudal, en este trabajo no se considera variable de control por tanto solo se tiene la matriz A, que se supone constante a lo largo del proceso. La matriz H en la ecuación de medición se compone por un vector fila de $1\,x\,n$ que contiene la última observación de cada variable de entrada. El valor pronosticado $z_k$ se obtiene mediante la ecuación de medición, multiplicando la matriz H por el vector de estado $x_k^-$ $(n\,x\,1)$. Las siguientes expresiones corresponden a la ecuación de estado, ecuación de medición y la matriz A que relaciona el estado anterior con el actual.

$$x_k = Ax_{k-1} + Bw_{k-1} \qquad (3.1)$$

$$z_k = Hx_k + v_k$$

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

El valor pronosticado $z_k$ (corresponde a un valor de caudal, tamaño $p$) lleva implícito el error de la medición $w_{k-1}$, y la ecuación de estado contiene el error del proceso $v_k$, los cuales deben cumplir el supuesto de normalidad.

$$w_k = N(0, Q) \qquad (3.2)$$

$$v_k = N(0, R)$$

Con las ecuaciones de estado y medición el algoritmo inicia un ciclo que se repite indefinidamente. En el tiempo $k-1$ hace la estimación a priori (pronóstico) de los estados, que son actualizados (estimación a posteriori) en el tiempo $k$. Los estados se toman como función de respuesta de la cuenca y con la estimación a posteriori se hace el pronóstico en el tiempo $k+1$. Este ciclo se repite de forma indefinida realizando el pronóstico del tiempo $k+1$ con base en la matriz $H$ y el vector de estado actualizados hasta el tiempo $k$.

### 3.4.2 Filtro de Kalman de Conjuntos

El filtro de Kalman de Conjuntos (EnKF) es un estimador subóptimo, que se basa en simulaciones de Monte Carlo para la estimación del error estadístico (Evensen, 1994; Gillijns et al., 2006; Rafieeinasab et al., 2014). Se supone distribución normal de los errores y las estimaciones se hacen con base en conjuntos que agrupan $q$ valores generados al azar bajo distribución normal. Al igual que el DKF tiene dos grupos de ecuaciones; análisis y predicción. En la Figura 3 las dos primeras ecuaciones tienen sus equivalentes en el DKF, y la última corresponde al promedio de los miembros que se supone como mejor pronóstico. El segundo grupo, de forma general, consta de cuatro ecuaciones donde se estiman errores y probabilidades que son insumo para calcular la ganancia de Kalman (Gillijns et al., 2006).

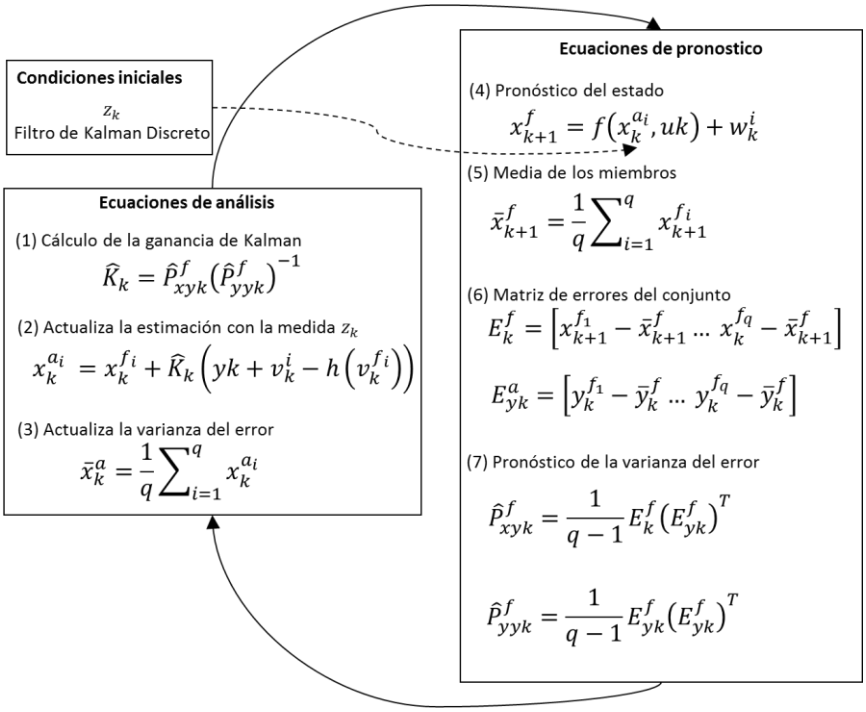



Figura 3.3. Algoritmo del filtro de Kalman de conjuntos (Gillijns et al., 2006).

Los errores $w_k^i$ y $v_k^i$ corresponden al ruido que contiene el proceso y la medición respectivamente, se suponen como ruido blanco con media cero y varianza $Q$ y $R$ (Figura 3), de igual forma que en la ecuación 2. Para el pronóstico de caudales

la función de respuesta ($x_{k+1}^{f}$) del EnKF se estableció incorporando ruido blanco $w_k^i$ a los estados previos $x_{k-1}^{ai}$:

$$x_{k+1}^{f} = x_k^{ai} + w_k^i \qquad (3.3)$$

El ruido en las mediciones se genera agregando a la medición en el tiempo $k$, $q$ desviaciones con distribución normal.

$$y_k^i = y_k + v_k^i \qquad (3.4)$$

El vector $y_k^i$ de $1\ x\ q$ corresponde a las $q$ mediciones ruidosas, $y_k$ es la medición en el tiempo $k$, y el subíndice $i$ representa la cantidad de miembros que corresponde a $i = 1,2,\dots,\dots,q$. Es decir, se generan $q$ valores al azar bajo distribución normal que se suman al valor observado y resulta en un vector de tamaño $q$ con el valor observado como media. Se supone que a mayor cantidad de miembros mejor será el ajuste del pronóstico, debido a que se puede obtener mejor estimación de la distribución de probabilidades (Leutbecher, 2019), sin embargo, aumentar la cantidad de miembros implica mayor esfuerzo computacional, por tanto se debe hacer una análisis de sensibilidad para determinar la cantidad que obtenga los mínimos errores con esfuerzo computacional aceptable. Para estudios hidrológicos se sugiere entre 50 y 300 miembros por conjunto (Gillijns et al., 2006; Quiroz et al., 2019).

### 3.4.3 Filtro de Kalman discreto y filtro de Kalman de Conjuntos

La integración del DKF y EnKF se hizo mediante la ecuación de estado del algoritmo del EnKF. El DKF genera un pronóstico para el tiempo $k+1$ ($\hat{Q}_{k+1}$), a partir del cual se determina la diferencia con respecto al estado previo de EnKF ($x_{k-1}^{ai}$) que finalmente es sumada al estado previo $x_{k-1}^{ai}$, además se le agrega ruido blanco mediante simulación de Monte Carlo $w_k^i$.

$$x_{k+1}^{f} = x_{k-1}^{ai} + \left(\hat{Q}_{k+1} - x_{k-1}^{ai}\right) + w_k^i \qquad (3.5)$$

### 3.4.4 Modelo autorregresivo de primer orden (ARX(1,1)) y DKF

El modelo autorregresivo de primer orden también conocido como modelo de Markov es una de las primeras aproximaciones para el estudio de series de tiempo, y se basa en la autocorrelación que se presenta dentro de la misma serie de datos (Box et al., 2016; Bras & Rodríguez-Iturbe, 1985), algebraicamente se describe de la siguiente forma:

$$y_{k+1} = \sum_{i=0}^{na} \alpha_i y_{k-i} + \sum_{j=0}^{nb} \beta_j \gamma_{k-j} + e_{k+1} \qquad (3.6)$$

Donde: $y_{k+1}$ corresponde al valor pronosticado, $\alpha_i$ y $\beta_j$ son los parámetros del modelo, $y_k$ y $\gamma_k$ corresponde a la variable de entrada y variable exógena respectivamente. La estimación de los parámetros se hace por mínimos cuadrados y se requiere una sección de serie de al menos 50 registros. En el modelo ARX(na,nb), na y nb representan los retrasos autorregresivos que se utilizan en cada variable. El modelo ARX estima los parámetros $\alpha$ y $\beta$ que son incorporados en la matriz A del algoritmo DKF (Figura 2), y la actualización se aplica al caudal para el pronóstico en el siguiente ciclo.

## 3.5 Resultados y Discusión

En total se evaluaron 36 combinaciones resultado de seis pasos ($L$) por seis tamaños de conjunto. Los tres algoritmos se ejecutaron con 1360 registros horarios de caudal provenientes de las estaciones Chapalagana y Platanitos, para obtener pronósticos de caudal en la estación Chapalagana.

En la Figura 4 se muestra el análisis de sensibilidad que permite identificar el tamaño adecuado de los conjuntos para los algoritmos de EnKF (línea punteada) y DKF-EnKF (línea continua), mediante la relación entre el valor de RMSE a medida que aumenta la cantidad de miembros por conjunto. El conjunto con cinco miembros presenta el RMSE más alto en cada uno de los pasos $L$, cuando se tiene 8 y 10 miembros se dan variaciones con ligera tendencia a disminución, y los conjuntos con 20, 50 y 100 miembros reflejan convergencia y estabilidad, por

tanto, en el caso de la serie bajo estudio para optimizar el ajuste y la carga computacional se puede usar 20 miembros. La combinación DKF-EnKF en los 6 pasos $L$ tienen mayores valores de RMSE que EnKF, las diferencias se hacen cada vez más notables en pasos más grandes, sin embargo, la convergencia se alcanza a partir de 20 miembros por conjunto.

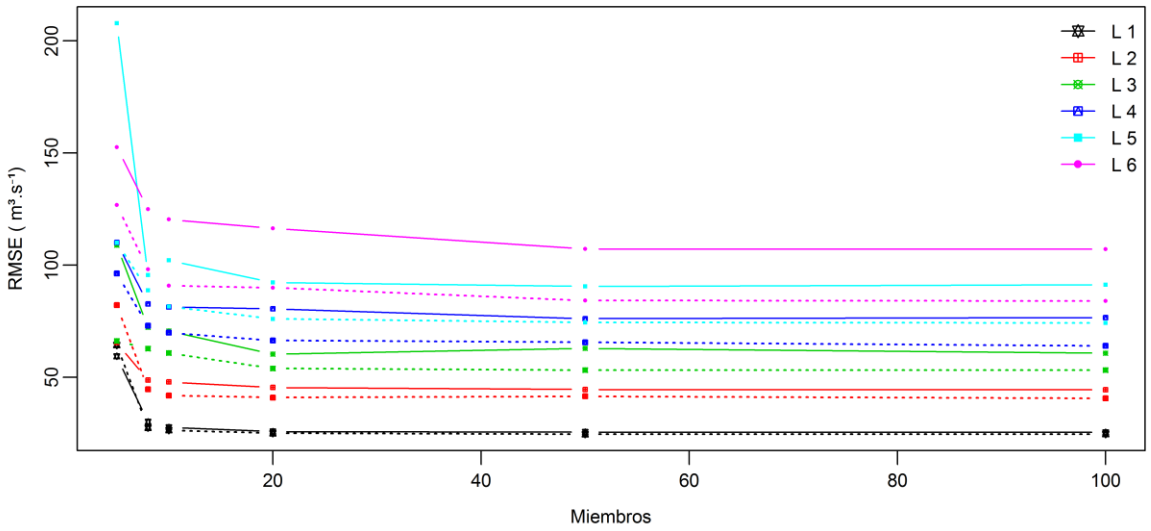



Figura 3.4. Raíz del cuadrado medio del error con diferentes conjuntos.

La importancia de la definición de la cantidad adecuada de miembros radica en obtener el mejor ajuste posible sin agregar carga computacional que no mejore el comportamiento del algoritmo. Cuando se tienen conjuntos grandes la carga computacional es alta, y por el contrario si se usan conjuntos pequeños se puede perder ajuste entre la serie observada y pronosticada (Gillijns et al., 2006; Quiroz et al., 2019).

Los resultados que se presentan son con base a 20 miembros por cada conjunto. En la Tabla 1 se presenta los indicadores estadísticos de ajuste de la serie observada contra las series pronosticadas.

Tabla 3.1. Estadísticos para la aplicación de DKF, EnKF, DKF-EnKF y ARX-DKF.

| | Índice | L1 | L2 | L3 | L4 | L5 | L6 | DEp |
|---|---|---|---|---|---|---|---|---|
| **DKF** | **RMSE** | 24.76 | 39.55 | 50.46 | 59.21 | 67.20 | 74.49 | 18.36 |
| | **Nash - Sutcliffe** | 0.9869 | 0.9666 | 0.9457 | 0.9252 | 0.9037 | 0.8816 | 0.0393 |
| | **Media** | 161.55 | 161.71 | 161.39 | 161.57 | 160.74 | 160.67 | 0.45 |
| | **DEs** | 217.58 | 216.33 | 213.69 | 212.47 | 209.96 | 208.30 | 3.58 |
| **EnKF** | **RMSE** | 25.08 | 41.00 | 53.99 | 66.40 | 76.04 | 89.93 | 23.67 |
| | **Nash - Sutcliffe** | 0.9866 | 0.9641 | 0.9378 | 0.9060 | 0.8767 | 0.8275 | 0.0587 |
| | **Media** | 160.65 | 160.55 | 160.88 | 160.15 | 160.93 | 161.27 | 0.38 |
| | **DEs** | 216.48 | 215.81 | 217.25 | 216.67 | 218.17 | 219.18 | 1.23 |
| **DKF-EnKF** | **RMSE** | 25.83 | 45.45 | 60.29 | 80.57 | 92.24 | 116.41 | 32.89 |
| | **Nash - Sutcliffe** | 0.9858 | 0.9559 | 0.9225 | 0.8616 | 0.8185 | 0.7110 | 0.1013 |
| | **Media** | 160.28 | 159.51 | 159.78 | 160.34 | 161.64 | 162.09 | 1.03 |
| | **DEs** | 216.21 | 217.95 | 219.67 | 226.07 | 229.10 | 239.75 | 8.84 |
| **ARX-DKF** | **RMSE** | 26.07 | 44.11 | 59.70 | 72.60 | 83.23 | 91.65 | 24.71 |
| | **Nash - Sutcliffe** | 0.9855 | 0.9586 | 0.9243 | 0.8882 | 0.8532 | 0.8222 | 0.0625 |
| | **Media** | 163.32 | 165.27 | 166.79 | 168.01 | 169.71 | 170.52 | 2.72 |
| | **DEs** | 219.90 | 222.08 | 224.70 | 226.56 | 228.20 | 228.34 | 3.42 |

Nota: RMSE: Raíz del error cuadrático medio, Nash-Sutcliffe: Índice de Nash, Media: Promedio, DEs: desviación estándar en cada serie pronosticada y DEp: desviación estándar entre los seis pasos por cada algoritmo; El promedio y desviación estándar de la serie observada son 160.95 y 216.67 respectivamente.

La implementación del modelo ARX(1,1) y el modelo ARX-DKF con actualización recursiva obtienen similares resultados debido a que los parámetros de ARX se multiplican por los valores de caudal en el tiempo $k$, operación que se ejecuta de manera matricial dentro de la ecuación de pronóstico de estado del algoritmo del DKF. Los resultados que se presenta se refieren únicamente al modelo ARX-DKF.

Para el paso $L = 1$ los algoritmos con Filtro de Kalman generan pronósticos con coeficientes de Nash - Sutcliffe (NS) superiores a 0.9855, que es ligeramente mejor a lo encontrado en los pronósticos de los caudales en la presa Ángel Albino Corzo (NS=0.9774) (Morales-Velázquez et al., 2014). Sin embargo, a medida que el paso $L$ se hace más grande se presentan diferencias crecientes entre los algoritmos. El DKF se mantiene con el mejor ajuste, seguido del EnKF y ARX-

DKF, y, por último, la integración DKF-EnKF con un descenso pronunciado hasta 0.7110 en el paso $L = 6$ (Tabla 1). El RMSE se puede interpretar con unidades en $m^3/s$ y presenta un comportamiento de ajuste similar al evidenciado por NS. Tanto NS como RMSE demuestran que el menor error se logra con DKF seguido de EnKF, ARX-DKF, y DKF-EnKF. En términos de NS y RMSE se puede advertir que los algoritmos DKF y EnKF generan pronósticos con mejor ajuste respecto al modelo ARX-DKF, además, la utilización del DKF no requiere de largos tramos de serie para el entrenamiento, lo que brinda mayor simplicidad del modelo.

De acuerdo con el resumen estadístico de la Tabla 1, la media y la desviación estándar en los seis pasos tienden a sobreestimarse con diferentes magnitudes respecto a la serie observada. La mayor estabilidad de la media la tiene el algoritmo de EnKF con desviación estándar (DEp) de 0.38, con leve diferencia con DKF que tiene desviación de 0.45 entre los seis pasos, mientras que ARX-DKF y DKF-EnKF tienen mayor desviación estándar de 1.03 y 2.72 respectivamente. Por otra parte, la dispersión representada por la desviación estándar entre las series pronosticadas (DEp) confirma que EnKF guarda mayor similitud con la serie observada en los seis pasos. El aparente mejor ajuste de los pronósticos de EnKF se debe al efecto de desplazamiento respecto a la observada sin mayor cambio en los picos máximos o mínimos de tal manera que no hay efecto notable en la media y desviación estándar.

Las diferencias entre los pronósticos generados se hacen notables de forma gráfica y para una comparación detallada en las Figuras 5 y 6 se presentan dos avenidas a diferentes pasos. Los pronósticos para el tiempo $k + L$ se desplazan un valor similar a $L$ debido a que la actualización incorpora la última medición correspondiente al tiempo $k$. No obstante, en el evento mostrado en la Figura 6 el pronóstico generado por DKF y ARX el desplazamiento es casi nulo, esto se debe a que el retraso entre el pico del 25/09/2017 18:00 en la estación Platanitos y el pico del 26/09/17 02:00 en la estación Chapalagana es de 8 horas. Es decir, DKF estima y actualiza el Hidrograma Unitario Instantáneo (HUI) con base en registros de caudal de la estación Platanitos del tiempo actual ($k$), que en la

estación Chapalagana se mostraran 8 horas después consecuencia al tiempo de concentración. La proximidad del tamaño del retraso y el paso $L = 6$ permite que la actualización se haga con registros correlacionados disminuyendo el desplazamiento y dando mejor ajuste en el pico máximo. Por su parte ARX genera el pronóstico con similar ajuste que DKF pero con un segmento de serie de mayor extensión, es decir, el modelo DKF representa mayor sencillez para implementarlo.

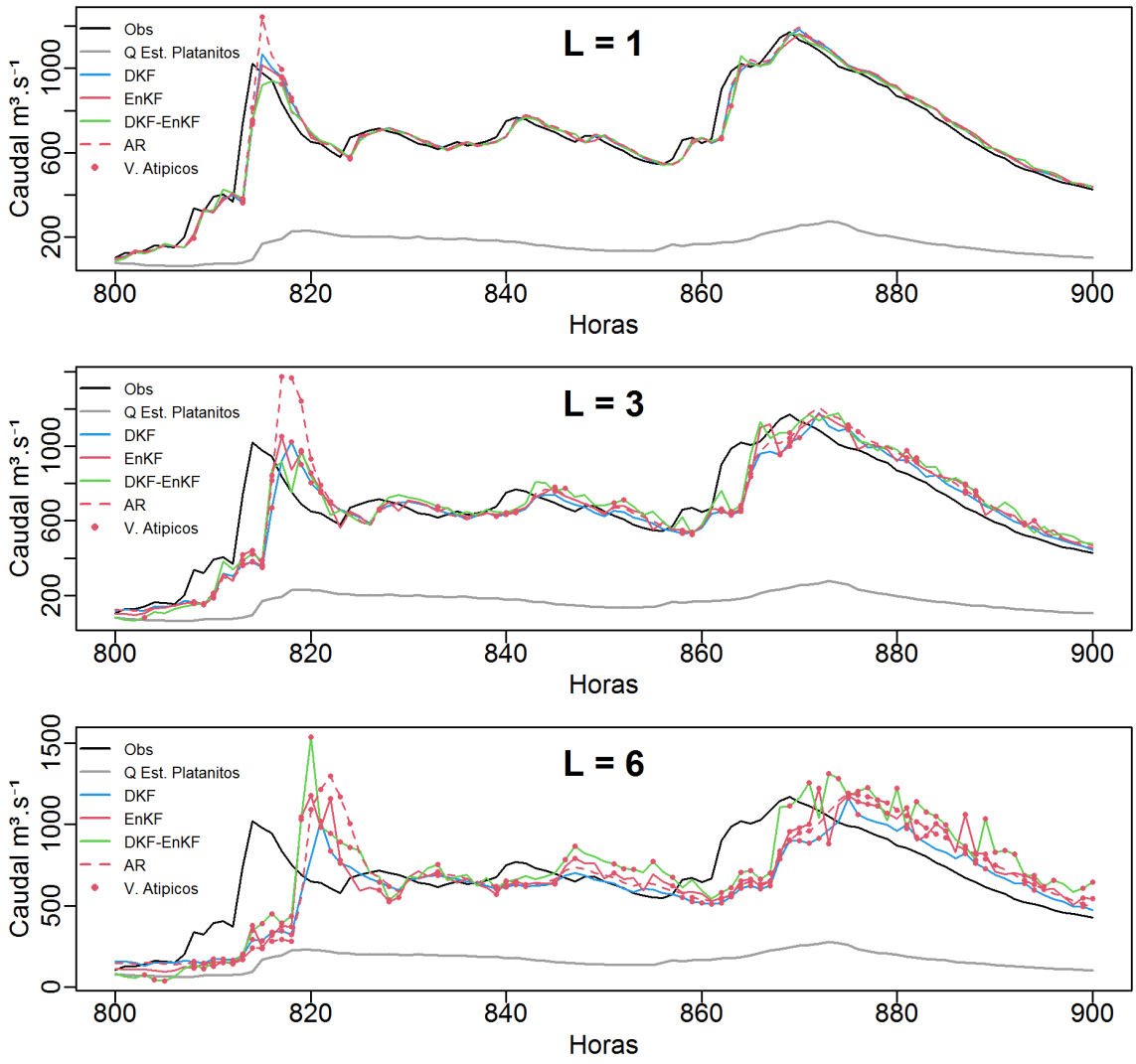



Figura 3.5. Pronósticos con DKF, ENKF y DKF-EnKF (Avenida del 4/9/2017 16:00 h al 8/9/2017 20:00 h)

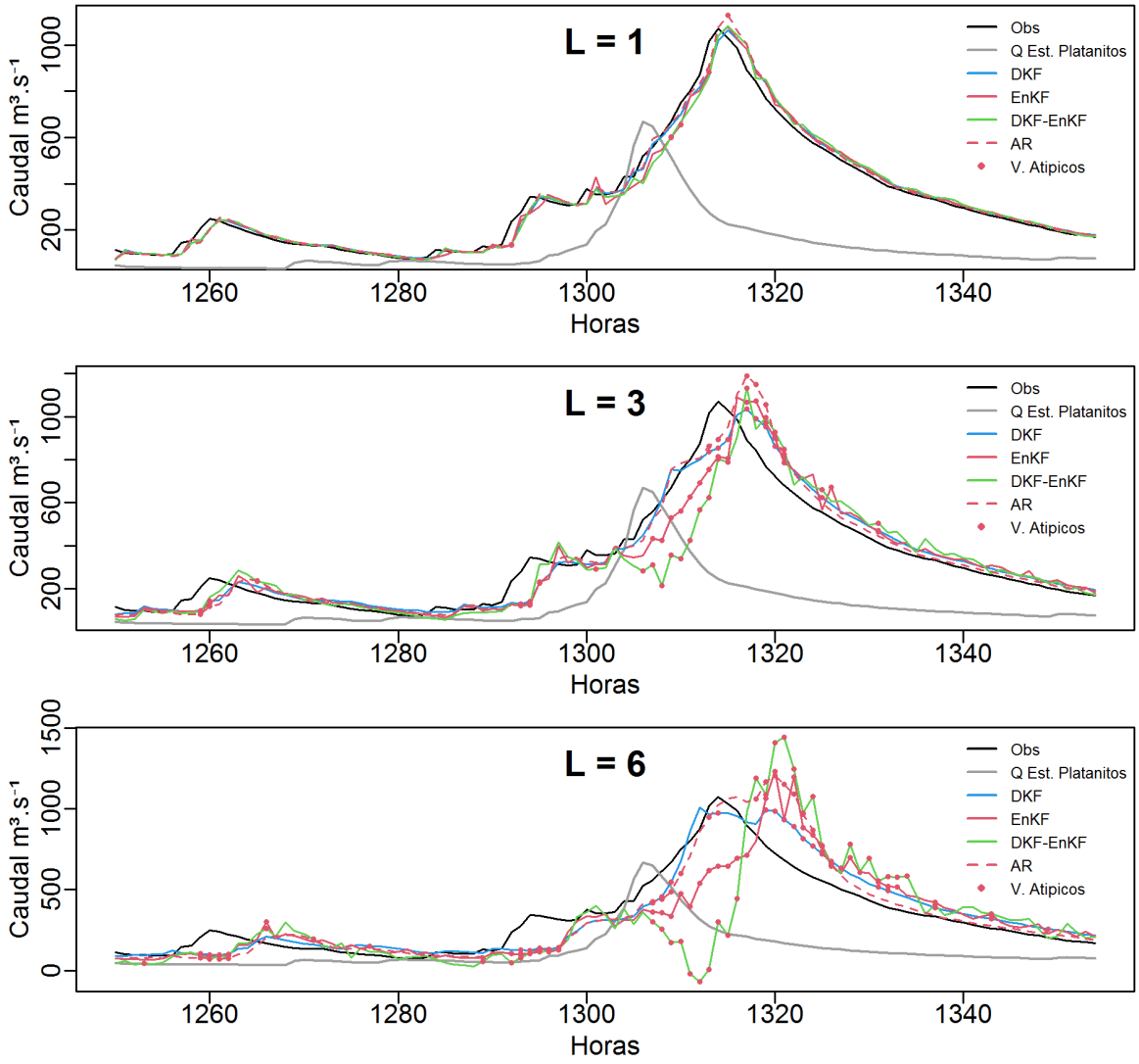



Figura 3.6. Pronósticos con DKF, EnKF y DKF-EnKF (Avenida del 23/09/17 10:00 h al 28/09/17 00:00 h)

Por otra parte, los algoritmos de EnKF y DKF-EnKF no obtienen mejoras debido a que EnKF hace estimaciones de escalares y no se incluye el retraso entre series. A diferencia de DKF, el pronóstico generado por DKF-EnKF se invierte de tal forma que genera un pico mínimo (Hora 1311, 25/09/2017 23:00) que se debe a la actualización con el pico máximo de caudal que presenta la estación Platanitos (Hora 1306, 25/09/2017 18:00) y que el algoritmo de EnKF trata de corregir. Una posible adecuación es la implementación de EnKF para la

estimación de estados en lugar de escalares que también puede agregar capacidad para tratamiento de no linealidad.

Los pronósticos pueden alcanzar mejor ajuste con la incorporación dinámica del tiempo de retraso entre series, tratamiento de variables con distintos tipos de observación y proporciones (Meng et al., 2017), e incorporación de las variables que afectan el tiempo de retraso como la ubicación y dirección de los eventos de precipitación y la condición de humedad antecedente del suelo de la cuenca. La precipitación, presión atmosférica, temperatura y humedad relativa son variables que pueden ayudar a tener mejores ajustes en el pronóstico. Asimismo, dado que los caudales que se miden en la parte alta de la cuenca son proporcionalmente menores debido a la menor área de captación disponible, es conveniente incluir en el modelo parámetros que ayuden a equiparar las magnitudes de los caudales. Además, es determinante tener registros de estaciones meteorológicas distribuidas en el área de la cuenca para definir y actualizar la función de respuesta de la cuenca (HUI) (González-Leiva et al., 2015).

La dispersión entre la serie observada y las series pronosticadas (Figura 7) muestran similitud entre los algoritmos. A medida que se aumenta $L$ el valor de la pendiente (en la ecuación lineal Figura 7) va disminuyendo, es decir, que en general los valores pronosticados son subestimados respecto al valor observado.

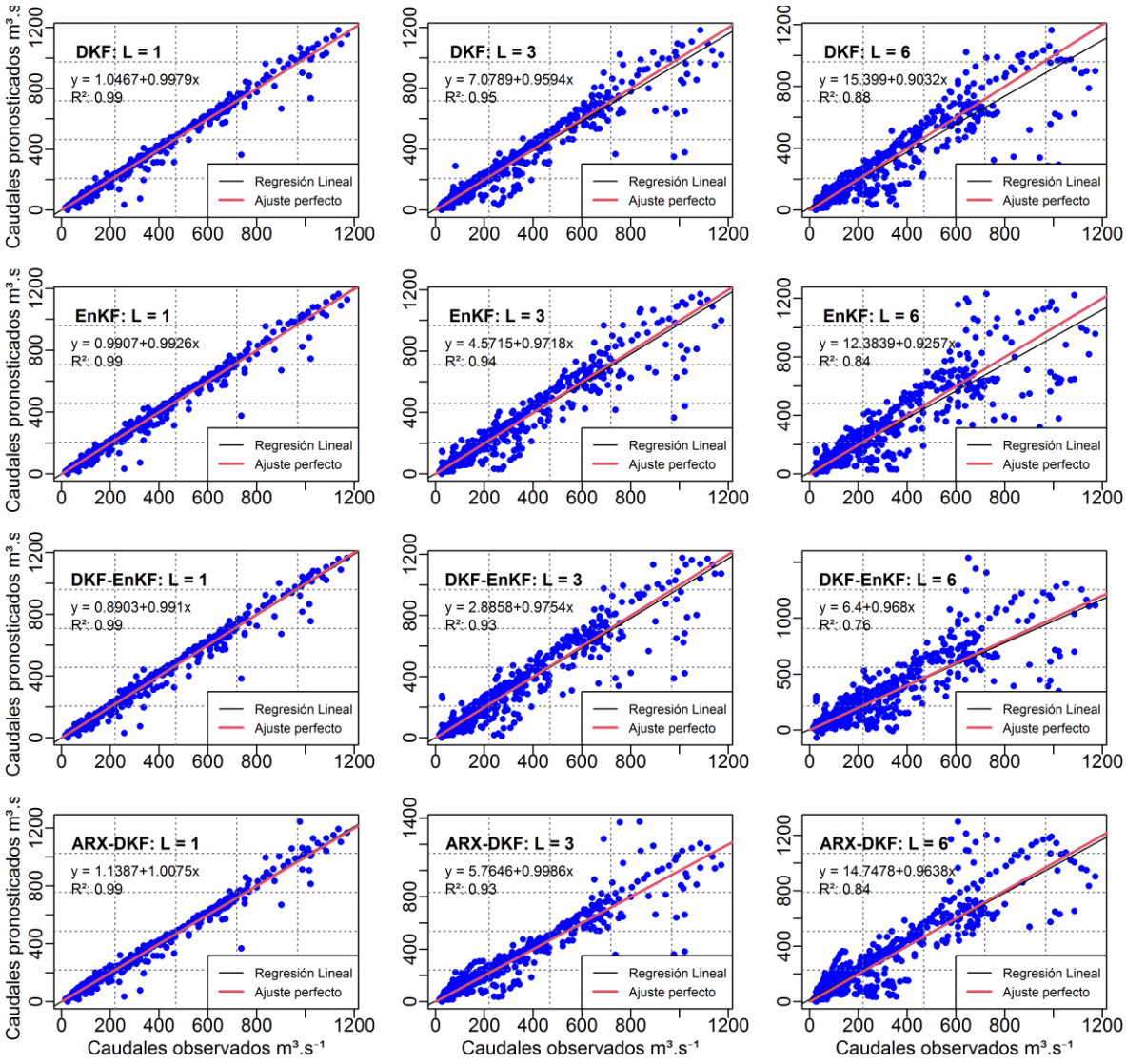



Figura 3.7. Caudales observados versus pronosticados.

Cuando $L = 1$ el ajuste lineal entre los valores observados y pronosticados tiene el coeficiente de determinación de 0.99, que es el más alto y similar entre los cuatro algoritmos (Tabla 2). En los pasos $L = 2$ en adelante, la dispersión aumenta debido al distanciamiento del conjunto de datos utilizado para el pronóstico. El DKF mantiene el mejor ajuste en los seis pasos, con marcada similitud con EnKF hasta $L = 3$. Cuando $L = 6$ DKF tiene el coeficiente de determinación de 0.88 que es el más alto. El comportamiento del ajuste lineal es congruente con el valor de NS mostrado en la Tabla 1. Los puntos alejados de la línea central corresponden a pronósticos donde el caudal de la estación Chapalagana presenta cambios bruscos para conformar una avenida. Dichos

puntos, generan pronósticos con magnitudes atípicas. En las horas $k = 813$ y $k = 814$ se presentan las mayores desviaciones respecto al valor observado y equivalen a la etapa inicial de conformación de la avenida que se muestra en la Figura 5.

Tabla 3.2. Coeficientes de determinación de las series pronosticadas.

| | **L1** | **L2** | **L3** | **L4** | **L5** | **L6** |
|---|---|---|---|---|---|---|
| **DKF** | 0.99 | 0.97 | 0.95 | 0.93 | 0.90 | 0.88 |
| **EnKF** | 0.99 | 0.96 | 0.94 | 0.91 | 0.88 | 0.84 |
| **DKF-EnKF** | 0.99 | 0.96 | 0.93 | 0.87 | 0.84 | 0.76 |
| **ARX-DKF** | 0.99 | 0.96 | 0.93 | 0.90 | 0.87 | 0.84 |

A fin de verificar la normalidad de los residuales generados a partir de los pronósticos del DKF, EnKF y DKF-EnKF, en la Figura 8 se muestran los histogramas y el ajuste de la curva normal para cada algoritmo en los diferentes pasos. Es importante considerar que es ideal que los residuales tengan distribución normal, pero en pocos casos se obtiene así, por tanto, se admite una aproximación (Chatfield, 2001)

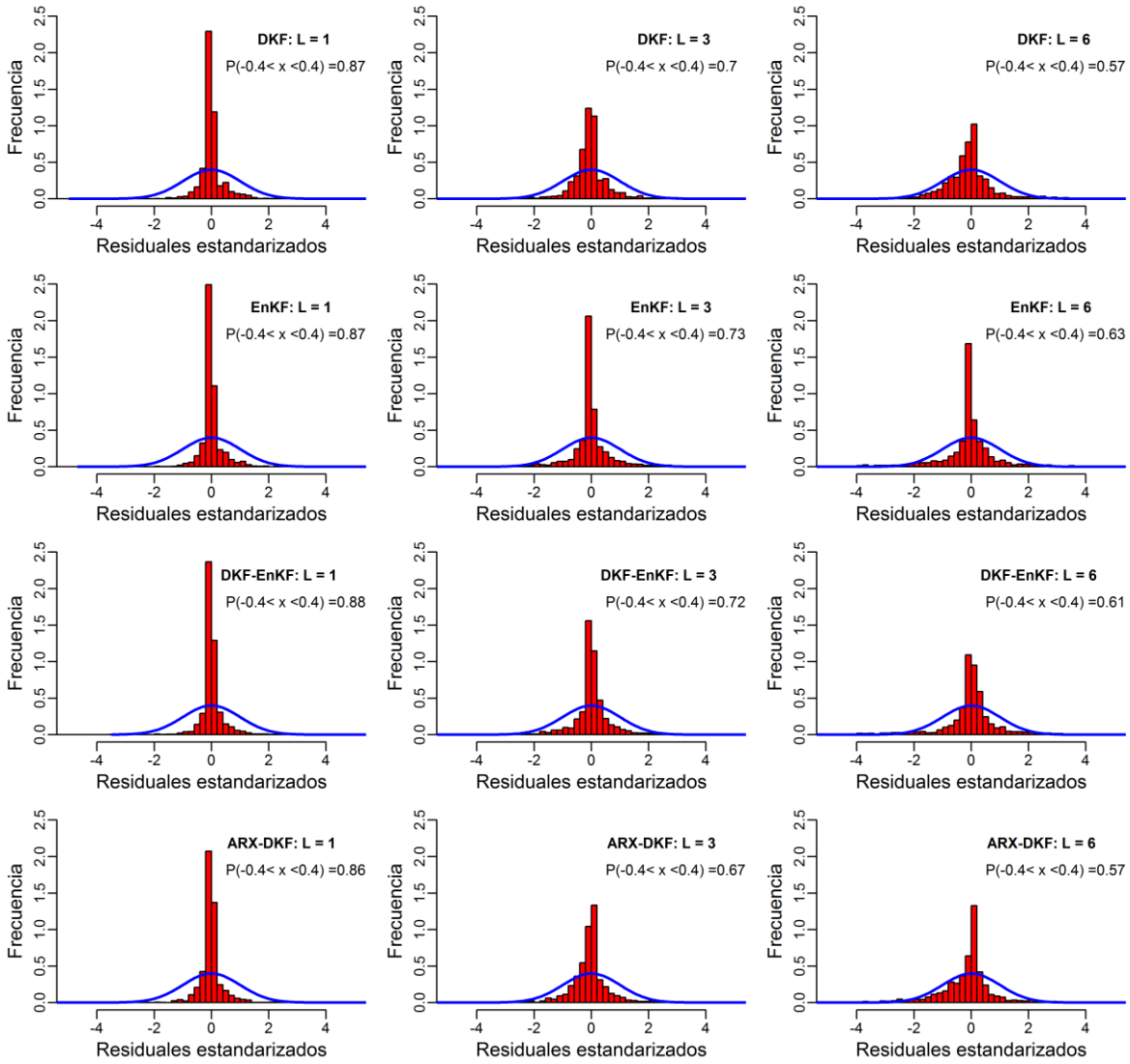



Figura 3.8. Histograma de residuales del DKF.

Se puede notar distribución simétrica con mayor concentración de pronósticos en la parte central. En el rango entre -0.4 y 0.4 de los residuales estandarizados, $L = 1$ agrupa entre 86.03 % y 87.3 % y en $L = 6$ se agrupan entre el 56.67 % y 63.02 % del total de las estimaciones. Asumiendo que se tiene distribución normal aproximada en los residuales, se hizo la estandarización para determinar la presencia de valores atípicos mayores y menores a dos desviaciones estándar a partir del cero. En la Figura 9 se agrupan los residuales estandarizados para 1 y 6 pasos resultado de los pronósticos realizados con DKF, EnKF y DKF-EnKF. Los valores atípicos se ubican en los ascensos y descensos de los picos, y se muestran como puntos rojos en las Figuras 5 y 6. En concordancia con los índices

estadísticos de la Tabla 1, a medida que el paso $L$ se hace más grande los valores atípicos tienen mayor frecuencia. En el paso $L = 1$ se presenta similitud que se va distorsionando hasta tener un rango de 55 entre DKF y DKF-EnKF en el paso $L = 6$.

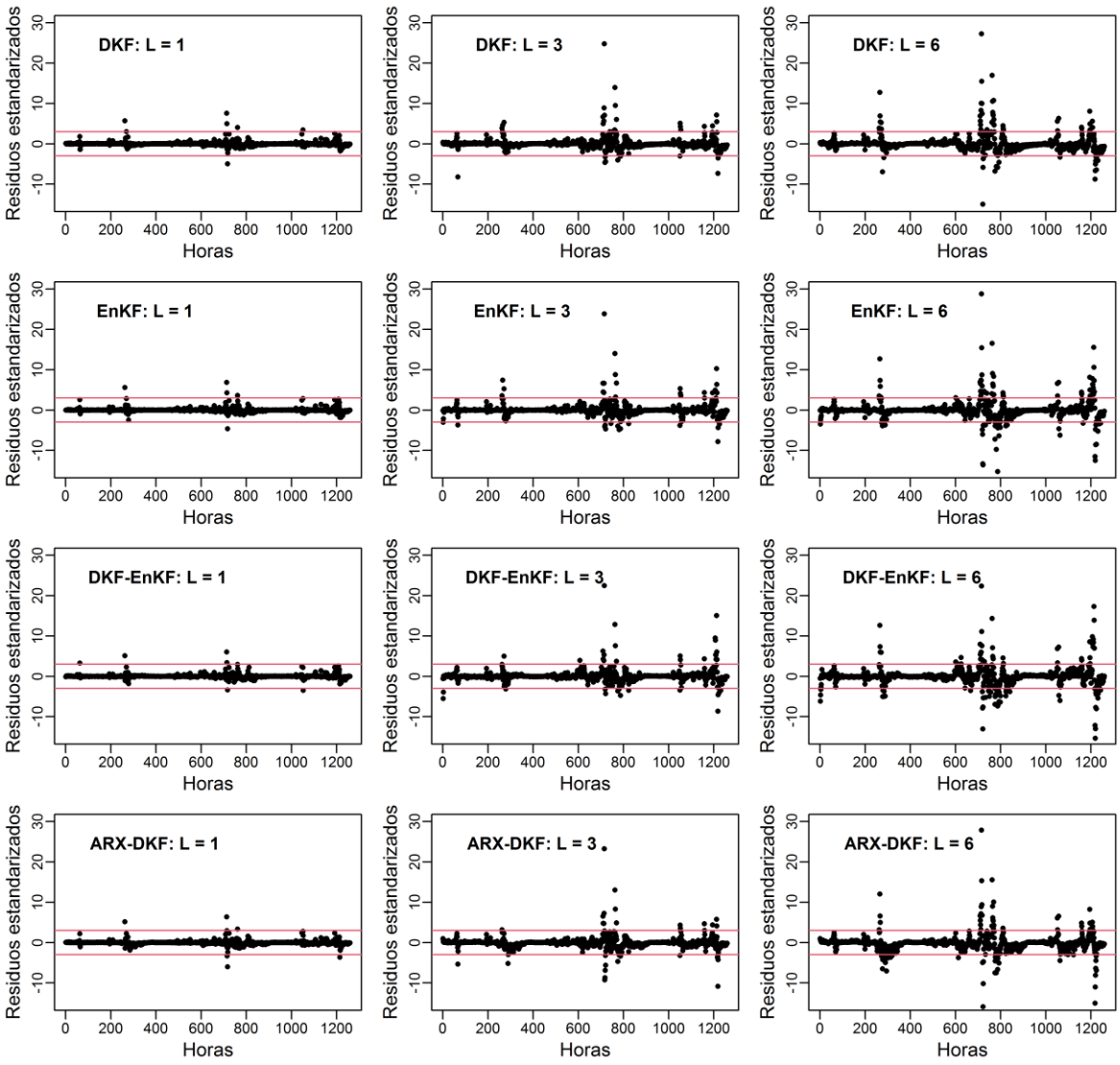



Figura 3.9. Residuales estandarizados.

La presencia de datos atípicos está estrechamente relacionada con el nivel de ajuste. Como se muestra en la Tabla 3, DKF presenta la menor cantidad de valores atípicos asociado a la capacidad que tiene para gestionar los retrasos entre las series (Tabla 1). Por el contrario, DKF-EnKF a excepción del paso $L = 1$ tiene la mayor presencia de valores atípicos, debido a la aplicación de ganancia

por el DKF y en secuencia por el EnKF que proyecta de forma ampliada los cambios bruscos en el caudal, asimismo, el pico mínimo que se muestra en la Figura 7 favorece al aumento de valores atípicos.

Tabla 3.3. Cantidad de valores atípicos por algoritmo y paso.

| | **L1** | **L2** | **L3** | **L4** | **L5** | **L6** |
|---|---|---|---|---|---|---|
| **DKF** | 13 | 41 | 63 | 84 | 96 | 112 |
| | 1.03% | 3.25% | 5.00% | 6.67% | 7.62% | 8.89% |
| **EnKF** | 15 | 41 | 70 | 99 | 112 | 143 |
| | 1.19% | 3.25% | 5.56% | 7.86% | 8.89% | 11.35% |
| **DKF-EnKF** | 14 | 46 | 66 | 105 | 132 | 167 |
| | 1.11% | 3.65% | 5.24% | 8.33% | 10.48% | 13.25% |
| **ARX-DKF** | 15 | 42 | 62 | 87 | 107 | 137 |
| | 1.19% | 3.33% | 4.92% | 6.90% | 8.49% | 10.87% |

## 3.6 Conclusiones

El DKF obtiene el mejor ajuste debido a la estimación de estados que permite el aprovechamiento del retraso entre series como base para la actualización. El EnKF estima valores escalares y genera suavizado con efecto de desplazamiento que mantiene estables la media y la desviación estándar, pero con ajuste inferior al DKF. De esta manera, la utilización del DKF donde se incorpore de forma recursiva el tiempo de retraso para actualización puede mejorar ajuste en los picos, además, la simplicidad para programar y operar el DKF los hacen viable para pronóstico de caudales a corto plazo.

Para mejorar el ajuste de los pronósticos es importante adelantar trabajos de investigación donde se incluya la actualización recursiva del tiempo de retraso entre las series de diferentes estaciones teniendo en cuenta que se presentan variaciones a lo largo del tiempo. Además, se puede evaluar los pronósticos con tamaños de paso de varias horas, por ejemplo, agrupar seis horas mediante valores como la media o máximo a fin de que cada paso equivale a seis horas y un pronóstico a 6 pasos equivale a 36 horas, de esta forma se da más flexibilidad al modelo y se puede ampliar el periodo de pronóstico.

## 3.7 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 durante el año 2018.

## 3.8 Referencias

Abaza, M., Anctil, F., Fortin, V., & Turcotte, R. (2015). Exploration of sequential streamflow assimilation in snow dominated watersheds. Advances in Water Resources, 86, 414–424. https://doi.org/10.1016/j.advwatres.2015.10.008

Box, G., Jenkins, G., Reinsel, G., & Ljung, G. (2016). Time series analysis: Forecasting and control (5th ed.). Wiley.

Brandhorst, N., Erdal, D., & Neuweiler, I. (2017). Soil moisture prediction with the ensemble Kalman filter: Handling uncertainty of soil hydraulic parameters. Advances in Water Resources, 110(February), 360–370. https://doi.org/10.1016/j.advwatres.2017.10.022

Bras, R., & Rodríguez-Iturbe, I. (1985). Random Functions and Hydrology. Addison-Wesley Publishing Company.

Chatfield, C. (2001). Prediction Intervals for Time-Series Forecasting. In J. S. Armstrong (Ed.), A Handbook for Researchers and Practitioners (Vol. 30, pp. 475–494). Springer US. https://doi.org/10.1007/978-0-306-47630-3

Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., Schmidt, J., & Uddstrom, M. J. (2008). Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model. Advances in Water Resources, 31(10), 1309–1324. https://doi.org/10.1016/j.advwatres.2008.06.005

CONAGUA. (2008). Estadisticas del agua en México. SEMARNAT.

Cryer, J. D., & Chan, K.-S. (2008). Time Series Analysis with Applications in R (2nd ed.). Springer.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. Journal of Geophysical Research, 99(C5), 10143. https://doi.org/10.1029/94jc00572

Evensen, G. (2003). The Ensemble Kalman Filter: theoretical formulation and practical implementation. Ocean Dynamics, 53(4), 343–367. https://doi.org/10.1007/s10236-003-0036-9

Evensen, G. (2009). Data Assimilation: The ensemble Kalman filter (2nd ed.). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-03711-5

Genz, A., & Bretz, F. (2009). Computation of Multivariate Normal and t Probabilities. Springer-Verlag.

Gillijns, S., Mendoza, O. B., Chandrasekar, J., De Moor, B. L. R., Bernstein, D. S., & Ridley, A. (2006). What is the ensemble Kalman filter and how well does it work? 2006 American Control Conference, 6 pp. https://doi.org/10.1109/ACC.2006.1657419

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. Tecnologia y Ciencias Del Agua, 6(4), 5–24.

INEGI. (2010). Hidrografía. https://www.inegi.org.mx/temas/hidrografia/default.html#Descargas

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. Journal of Basic Engineering, 82(Series D), 35–45.

Leutbecher, M. (2019). Ensemble size: How suboptimal is less than infinity? Quarterly Journal of the Royal Meteorological Society, 145(S1), 107–128. https://doi.org/10.1002/qj.3387

Liu, Weerts, A. H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenberg, D., Smith, P., van Dijk, A. I. J. M., van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O., & Restrepo, P. (2012). Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. Hydrology and Earth System Sciences, 16(10), 3863–3887. https://doi.org/10.5194/hess-16-3863-2012

Liu, Y., & Gupta, H. V. (2007). Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. Water Resources Research, 43(7), 1–18. https://doi.org/10.1029/2006WR005756

Maxwell, D. H., Jackson, B. M., & McGregor, J. (2018). Constraining the ensemble Kalman filter for improved streamflow forecasting. Journal of Hydrology, 560, 127–140. https://doi.org/10.1016/j.jhydrol.2018.03.015

Meng, S., Xie, X., & Liang, S. (2017). Assimilation of soil moisture and streamflow observations to improve flood forecasting with considering runoff routing lags. Journal of Hydrology, 550, 568–579. https://doi.org/10.1016/j.jhydrol.2017.05.024

Morales-Velázquez, M. I., Aparicio, J., & Valdes, J. B. (2014). Pronóstico de avenidas utilizando el filtro de kalman discreto. Tecnologia y Ciencias Del Agua, 5(2), 85–110.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. Journal of Hydrology, 10(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Quiroz, K., Collischonn, W., & Paiva, R. C. D. de. (2019). Data assimilation using the ensemble Kalman filter in a distributed hydrological model on the

Tocantins River, Brasil. RBRH, 24, 1–15. https://doi.org/10.1590/2318-0331.241920180031

R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. http://www.R-project.org/. https://www.r-project.org/

Rafieeinasab, A., Seo, D. J., Lee, H., & Kim, S. (2014). Comparative evaluation of maximum likelihood ensemble filter and ensemble Kalman filter for real-time assimilation of streamflow data into operational hydrologic models. Journal of Hydrology, 519(PD), 2663–2675. https://doi.org/10.1016/j.jhydrol.2014.06.052

Simon, D. (2001). Kalman Filtering. Embedded Systems Programming, June, 72–79.

SMN. (2019). Sistema de Información Climática Computarizada CLICOM. Servicio Metereologico Nacional. http://clicom-mex.cicese.mx/malla/index.php

Sun, L., Seidou, O., Nistor, I., & Liu, K. (2016). Review of the Kalman-type hydrological data assimilation. Hydrological Sciences Journal, 61(13), 2348–2366. https://doi.org/10.1080/02626667.2015.1127376

Valdés, J., Mejía, J., & 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.

Welch, G., & Bishop, G. (2006). An Introduction to the Kalman Filter. In University of North Carolina at Chapel Hil (Issue 1).

Zou, L., Zhan, C., Xia, J., Wang, T., & Gippel, C. J. (2017). Implementation of evapotranspiration data assimilation with catchment scale distributed hydrological model via an ensemble Kalman Filter. Journal of Hydrology, 549, 685–702. https://doi.org/10.1016/j.jhydrol.2017.04.036

# 4 ENSEMBLE KALMAN FILTER (ENKF) FOR HOURLY STREAMFLOW FORECASTING IN HUAYNAMOTA RIVER, NAYARIT, MÉXICO

## Filtro de Kalman de Conjuntos (EnKF) para pronóstico de caudales horarios en el Río Huaynamota, Nayarit, México

Ildefonso Narvaez-Ortiz[1], Laura Ibáñez-Castillo[1], Ramon Arteaga-Ramírez[1], Mario Vázquez-Peña[1]

[1] Posgrado en Ingeniería Agrícola y Uso Integral del Agua Universidad Autónoma Chapingo. (UACH). Km 38.5 Carretera México-Texcoco, Código postal. 56230. Estado de México.

Autor para correspondencia: libacas@gmail.com

### 4.1 Abstract

Hydrological phenomena are characterized by the formation of a non-linear dynamic system, and streamflows are not unrelated to this premise. Data assimilation offers an alternative for streamflow forecasting through the Ensemble Kalman Filter, due to its relative ease of implementation and lower computational effort compared with other techniques. The hourly streamflow forecast of the Chapalagana station was made based on the streamflow of the Platanitos station in northwest Mexico. The forecasts were made from one to six steps forward, combined with set sizes of 5, 10, 20, 30, 50 and 100 members. The Nash-Sutcliffe coefficient for the discrete Kalman filter is 0.99 and 0.85 in step one and six, respectively, achieving the best fit with a tendency to shift the predicted series, similar to the persistent forecast. The Ensemble Kalman Filter obtains 0.99 and 0.05 in steps one and six; however, it converges on the observed series with the limitation of considerable overestimates in higher steps. All three algorithms have equal statistical adjustment value in step one, and as it increases there are progressive differences, where ARX and DKF remain similar and Ensemble Kalman Filter is differentiated by the overestimates. The Ensemble Kalman Filter enables capturing non-linearity in sudden changes of streamflow but generates overestimates at the peaks.

**Keywords:** Ensemble Kalman Filter, autoregressive models, short-term streamflow forecasting, data assimilation.

## 4.2 Resumen

Los fenómenos hidrológicos se caracterizan por conformar un sistema dinámico no lineal, los caudales no son ajenos a esta premisa. La asimilación de datos ofrece una alternativa para el pronóstico de caudales mediante el Filtro de Kalman de Conjuntos, por su relativa facilidad de implementación y menor esfuerzo computacional respecto a otras técnicas. Se realizó el pronóstico del caudal horario de la estación Chapalagana en función de los caudales de la estación Platanitos, en el noroeste de México. Los pronósticos se realizaron de uno a seis pasos hacia adelante, combinados con tamaños de conjunto de 5, 10, 20, 30, 50 y 100 miembros. El coeficiente de Nash-Sutcliffe para el filtro de Kalman discreto es de 0.99 y 0.85 en el paso uno y seis respectivamente, logrando el mejor ajuste con tendencia a desplazar la serie pronosticada, similar al pronóstico persistente. El Filtro de Kalman de Conjuntos obtiene 0.99 y 0.05 en los pasos uno y seis, no obstante, converge sobre la serie observada con la limitante de sobrestimaciones considerables en pasos superiores. Los tres algoritmos tienen igual valor de ajuste estadístico en el paso uno, y a medida que aumenta se dan diferencias progresivas, donde ARX y DKF se mantienen similares y EnKF se diferencia por las sobrestimaciones. EnKF permite captar la no linealidad en los cambios bruscos de caudal, pero genera sobrestimaciones en los picos.

**Palabras clave:** Filtro de Kalman de conjuntos, modelos autorregresivos, pronósticos de caudales a corto plazo, asimilación de datos.

## 4.3 Introduction

Climate variability has intensified the incidence of extreme precipitation events that can generate sudden changes in streamflow and lead to floods and landslides (IPCC, 2012). Having advance information of streamflow behavior becomes an indispensable tool for administration of dams and disaster risk management (IPCC, 2012; Singh & Zommers, 2014). For streamflow forecasting, different methods have been used, such as autoregressive methods, neuronal networks (Box et al., 2016; Shmueli & Lichtendahl, 2016), and more recently, methods of data assimilation such as Kalman filters (Abaza et al., 2015; Alvarado-Hernández et al., 2020; González-Leiva et al., 2015; Morales-Velázquez et al., 2014). In hydrological studies, the Ensemble Kalman Filter (EnKF) (Evensen, 1994, 2009; Gillijns et al., 2006) has been widely used as a method of assimilation (Liu & Gupta, 2007; Maxwell et al., 2018; Sun, Seidou, Nistor, & Liu, 2016), with little

evaluation in forecasting flows. EnKF is an extension of the discrete Kalman Filter (Kalman, 1960) and a computationally less demanding alternative to the Extended Kalman filter for treatment of non-linear dynamic systems (Evensen, 1994, 2003). Among the applications of EnKF are streamflow forecasting in basins dominated by melting snow and ice (Abaza et al., 2015), evapotranspiration(Zou et al., 2017b) and soil moisture (Brandhorst et al., 2017; Meng et al., 2017). Moreover, it has been evaluated integrated with distributed hydrological models such as TopNet, Hydrotel and MGB-IPH (Abaza et al., 2015; Clark et al., 2008; Quiroz et al., 2019).

Hydrological phenomena, such as streamflow, have non-linear behavior (Bai et al., 2016; Xu et al., 2009), especially when there are sudden changes in river levels. For this reason, the use of non-linear algorithms of data assimilation favors fit of forecasts (Brandhorst et al., 2017; Medina González et al., 2015). In addition, systematic errors can be reduced by recursive updating based on each new available measurement (Clark et al., 2008; Maxwell et al., 2018).

According to Valdés et al. (1980) and Winkler et al. (2010), in hydrographic basins, different representative measurements and a point of interest have a dynamic relationship to the predominant physical and biological characteristics of the area. Based on this information, the behavior of a given phenomenon is modeled to obtain short-term forecasts. The Kalman filter enables incorporation of registers from diverse sources and constant updating (Box et al., 2016; Welch & Bishop, 2006).

To determine whether the algorithms of identification of non-linear dynamic systems allow forecasting short-term streamflow (6 h), this study evaluates the fit of the series predicted by algorithms of the Ensemble Kalman Filter (EnKF), the Discrete Kalman Filter (DKF) and the first-order autoregressive model, AR(1), in flows measured at the Chapalagana station on the Huaynamota River. The Kalman Filter algorithms estimate the system dynamic states and correspond to the response function of the basin.

## 4.4 Materials and Methods

The study was conducted in a tributary of the Huaynamota River, the Chapalagana or Atengo River, located in northwestern Mexico, between the states of Durango, Nayarit, Zacatecas, and Jalisco (INEGI, 2010) (Figure 1), between -104°33’34.16” and -103°27’29.84” W and between 23°28’50.05” and 21°23’57.62” N, with an area of 12 075.7 km2. Altitude varies from 219 to 3,147 masl. Concentration time is 39.88 hours, mean annual precipitation is 707 mm, and mean annual temperature is 17.9 °C (SMN, 2019).

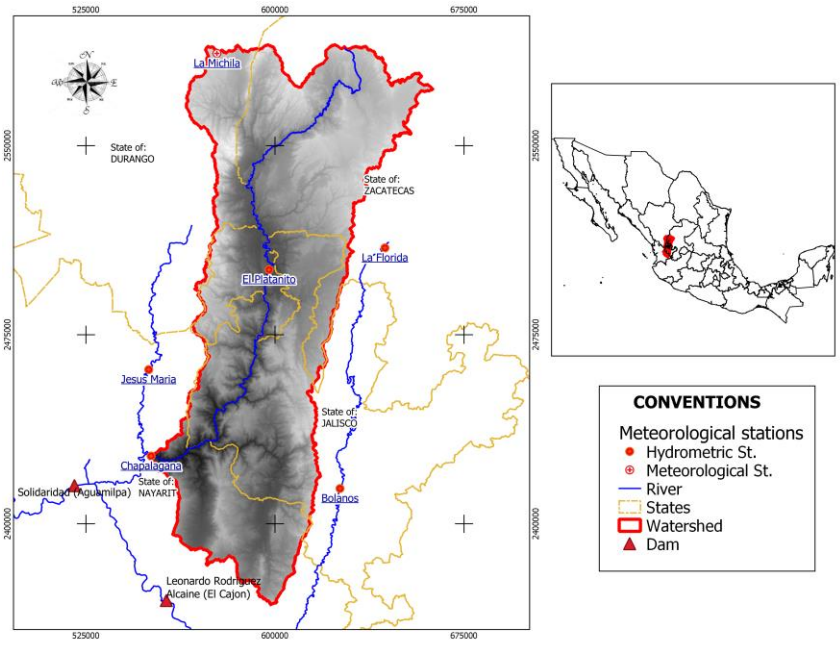



Figure 4.1. Location of the study area.
Source: Authors

The Huaynamota River contributes to generating electricity. The dam Solidaridad (known as Aguamilpa), located on the River Lerma-Santiago and geographically at 104°48’10.55” W and 21°50’22.74” N, produces 960 MW of electricity and has a maximum capacity of 5 540 million $m^3$ of water (CONAGUA, 2008). The Huaynamota River discharges into the Santiago River where the Aguamilpa dam is located, approximately 90 km upstream from the Pacific Ocean into which the Santiago River empties on the coast of the Mexican state of Nayarit.

We implemented the algorithm of the ensemble Kalman filter (EnKF) (Evensen, 1994), of the discrete Kalman filter (DKF) (Kalman, 1960), and the autoregressive model with first order exogenous variable (ARX(1,1)) (Bras & Rodríguez-Iturbe, 1985). Forecasts were made at 1, 2, 3, 4, 5 and 6 hours (L steps forward) of the flows at the Chapalagana station in function of the flows at the Platanito station, located 100 km upstream from Chapalagana, as the exogenous variable. Hourly streamflow series between 9:00 hours on August 2 to 0:00 hours on September 28, 2017 were used, a total of 1360 registers supplied by the Comisión Federal de Electricidad (CFE).

EnKF, DKF and ARX were implemented through R routines (R Core Team, 2020), which generate forecasts in six steps with DKF and ARX, and with 42 combinations between steps by set size in EnKF. Both EnKF and DKF were implemented to estimate the state vector that corresponds to the response function of the basin, or Instantaneous Unitary Hydrograph (IUH) (Valdés et al., 1980). Values are estimated that correspond to the columns of IUH, which when multiplied by the values of the measured series result in an estimation. In the three algorithms, the last observations of each series were considered (Valdés et al., 1980). The ARX model was implemented recursively based on a fraction of a series with 100 registers.

With the sensitivity analysis with 5, 10, 20, 30, 40, 50 and 100 members that were combined with the six steps and, based on the Root Mean Squared Error (RMSE), the adequate number of members in the EnKF sets was determined (Quiroz et al., 2019). White noise that is integrated in the EnKF members was generated with mvtnorm software (Multivariant Normal and t Distributions) (Genz & Bretz, 2009). In the algorithms of the Kalman filters evaluated, Q variance was assumed constant (Simon, 2001) with value of zero (Morales-Velázquez et al., 2014) and $R$ with a near zero value (0.01) to confer flexibility to convergence of the algorithm (Welch & Bishop, 2006).

The fit reached by each algorithm was evaluated using the Nash-Sutcliffe coefficient (Nash & Sutcliffe, 1970) and the Root Mean Square Error (RMSE) (Morales-Velázquez et al., 2014). The assumed normality of errors was verified using graphs (González-Leiva et al., 2015).

### 4.5 Ensemble Kalman Filter

To extend the functionality of the Discrete Kalman Filter (Kalman, 1960) and deal with non-linear dynamic systems, the Extended Kalman Filter, among others, has been proposed (Jazwinski, 2007; Welch & Bishop, 2006), as well as the Ensemble Kalman Filter (Evensen, 1994, 2009). The Ensemble Kalman Filter emerges as an alternative to the extended Kalman Filter, which has a high computational demand (Evensen, 1994, 2009) and is a sub-optimal estimator that, using Monte Carlo simulations, estimates the statistical error (Evensen, 1994; Gillijns et al., 2006; Rafieeinasab et al., 2014). Errors should satisfy the normal distribution assumption and are estimated based on sets of q values.

The algorithm is based on two groups of equations: forecast and analysis (Figure 2). In this study, the cycle began with the forecast equations using random values that make up the first matrix $\boldsymbol{x_k^{ai}}$, obtaining the first forecast using the measurement equation. The error matrixes of the forecast are calculated against the new measurement that is input for the analysis equations, where the states are updated in the measure that new measurements enter.

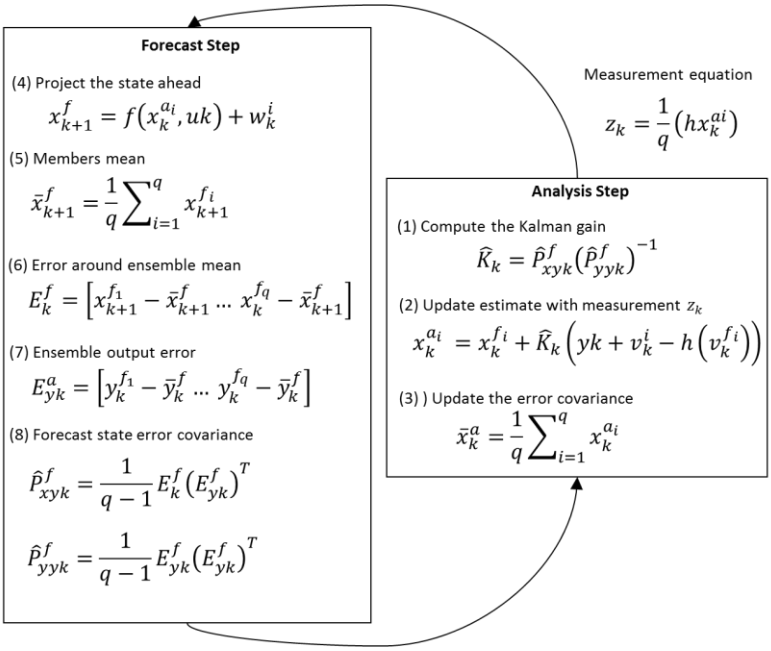



Figure 4.2. Algorithm of the ensemble Kalman filter.
Source: (Gillijns et al., 2006)

The h matrix is formed with the last register of two (n) hourly flow series of the Chapalagana and Platanitos stations. The uk parameter is absent because upstream from the Chapalagana station there are no structures, such as dams, that have a direct impact on streamflow. The errors $w_k^i$ and $v_k^i$ correspond to the noise the process contains and the measurement, respectively. They are assumed to be white noise with a mean of zero and variance $Q$ and $R$ (Figure 2). The noise in the measurements is generated by adding q deviations with normal distribution to the measurement in k time.

In the second analysis equation, the component $y_k + v_k^i$ represents the noisy measurements, $y_k$ is the measurement in time k, and the superscript i represents the number of members, that is, random values under normal distribution that correspond to $i = 1,2, \dots, \dots, q$. The adequate number of members in a set in hydrological studies is between 50 and 300 (Gillijns et al., 2006; Quiroz et al., 2019). The predicted value $z_k$ is obtained by averaging the vector resulting from multiplication of the h matrix and the $x_{k-1}^{ai}\mathrm{matrix}$, applying the measurement equation.

## 4.6 Discrete Kalman Filter

The discrete Kalman filter is an optimal recursive estimator of states in linear dynamic systems (Kalman, 1960) (Figure 2).

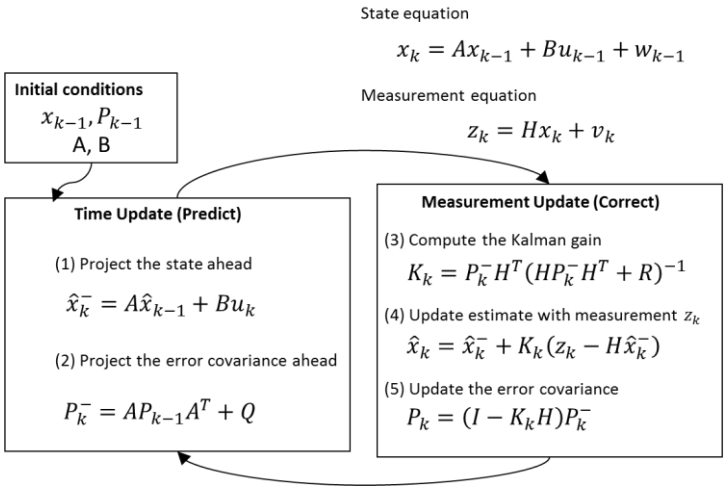



Figure 4.3. Discrete Kalman Filter.
Source: (Welch & Bishop, 2006)

The state equation has two (n) hourly flow series from the Chapalagana and Platanitos stations. The matrixes A ($n\ x\ n$) and B ($n\ x\ 1$) relate the state at time $k-1$ to the current state at time $k$. Like EnKF, the control matrix B is not included, because upriver from the Chapalagana station there are no structures that have a direct impact on streamflow. Matrix A is assumed constant throughout the process and matrix H is composed of a vector row of $1\ x\ n$ that contains the last observation of each entry series. The predicted value $z_k$ is obtained by the measurement equation by multiplying the H matrix by the state vector $x_k$ ($n\ x\ 1$).

Implicit in the predicted value $z_k$ is the measurement error $w_{k-1}$, and likewise, the process error $v_k$ is contained in the state equation. In both cases the normality assumption must be satisfied.

The state and measurement equations maintain a cycle that is repeated indefinitely. At time $k-1$, it makes the a priori estimation (forecast) of the states. And at time $k$, they are updated (a posteriori estimation). The states are assumed

as response function of the basin and the a posteriori estimation corresponds to the forecast for time $k+1$. This cycle is repeated indefinitely predicting time $k+1$ based on the H matrix and the state vector updated to time $k$.

## 4.7 First-order autoregressive model

One of the first approximations for the forecast is the first-order autoregressive model, which is based on the autocorrelation that occurs within the same series of data (Box et al., 2016; Bras & Rodríguez-Iturbe, 1985). Algebraically, it is described in the following way:

$$y_{k+1} = \sum_{i=0}^{na} \alpha_i y_{k-i} + \sum_{j=0}^{nb} \beta_j \gamma_{k-j} + e_{k+1} \quad (4.1)$$

where $y_{k+1}$ is the predicted value, $\alpha_i$ and $\beta_j$ are the model parameters, $y_k$ and $\gamma_k$ correspond to the entry variable and the exogenous variable, respectively. The parameters are estimated by the method of least squares, which requires a series section of at least 50 registers (Box et al., 2016; Shmueli & Lichtendahl, 2016). In the model $ARX(na, nb)$, $na$ and $nb$ represent the autoregressive delays that are used in each variable.

## 4.8 Results and discussion

Forecasts with six-hour steps were made of the flows at the Chapalagana station. For the case of EnKF, seven set sizes were evaluated to estimate the error: 6, 10, 20, 30, 40, 50 and 100 members. Series with 1360 registers, which included two main events, were used.

The EnKF algorithm shows stable convergence for all the steps, as of 50 members per set. Consequently, under the conditions of this study, it is acceptable to use at least 50 members per set to have adequate fit in the convergence of the algorithm and stabilization of the error, with the aim of minimizing computational effort. The results presented below have a base of 50 members per set. Table 1 presents the statistical indicators of fit of the observed series against the predicted series.

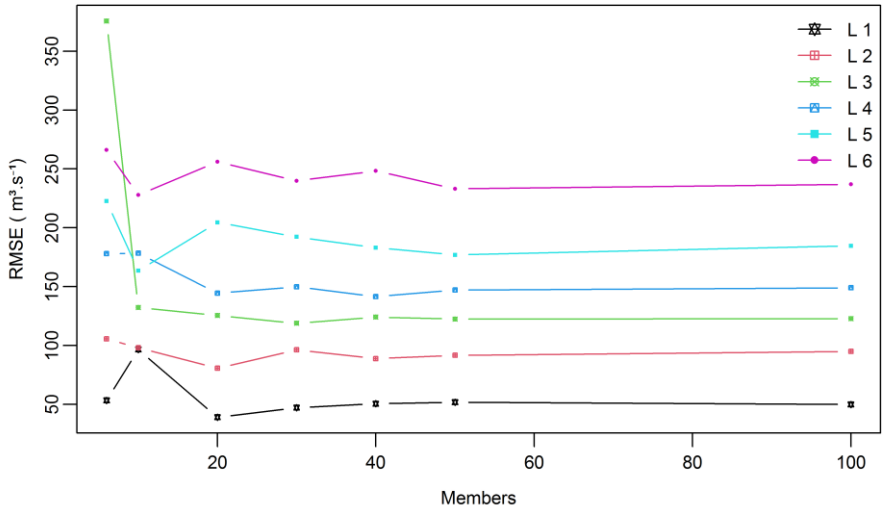



Figure 4.4. Root mean square error with different set sizes.
Source: Authors

**Table 4.1. Summary of statistics for application of EnKF, DKF, and ARX**

| Algorithm | Index | L1 | L2 | L3 | L4 | L5 | L6 |
|---|---|---|---|---|---|---|---|
| DKF | RMSE | 27,27 | 44,99 | 58,84 | 70,84 | 81,69 | 92,09 |
| | Nash - Sutcliffe | 0,99 | 0,96 | 0,94 | 0,91 | 0,88 | 0,85 |
| | Mean | 199,39 | 198,22 | 197,37 | 196,43 | 195,44 | 194,39 |
| | DE | 241,28 | 239,47 | 238,41 | 237,39 | 236,42 | 235,42 |
| EnKF | RMSE | 51,77 | 91,61 | 122,43 | 147,15 | 176,96 | 233,10 |
| | Nash - Sutcliffe | 0,95 | 0,85 | 0,74 | 0,62 | 0,45 | 0,05 |
| | Mean | 202,29 | 206,91 | 211,78 | 216,71 | 222,32 | 230,85 |
| | DE | 250,35 | 268,34 | 287,56 | 308,41 | 330,24 | 370,35 |
| ARX(1,1) | RMSE | 27,84 | 48,34 | 65,83 | 79,89 | 91,14 | 99,74 |
| | Nash - Sutcliffe | 0,99 | 0,96 | 0,92 | 0,89 | 0,86 | 0,83 |
| | Mean | 200,88 | 202,51 | 203,83 | 204,81 | 205,93 | 206,42 |
| | DE | 243,54 | 245,72 | 248,35 | 250,15 | 251,53 | 251,21 |

Note: RMSE: Root mean square error, Nash-Sutcliffe: Nash-Sutcliffe index, Mean: Average; mean and standard deviation of the observed series are 198.7 and 239.77, respectively.

Source: Authors

According to the statistical indicators in Table 1, the Nash-Sutcliffe coefficient (NS) shows similarities between DKF and ARX, with values of 0.99 and 0.83 in steps one and six, respectively, while EnKF obtains 0.95 and 0.05 in steps one and six, respectively. According to the NS, the fit of all the steps is less with EnKF; in step one it is 0.04 less and there is a marked change up to step six where the difference

is around 0.78. The mean of the predicted series is more stable with DKF, followed by ARX and, finally, EnKF. EnKF and ARX show an upward trend in the mean value for each step, while DKF has a downward trend.

Despite the low fit values, the EnKF algorithm expresses the changes with a non-linear trend and shows better convergence on the observed series once it is updated with new measurements. DKF and ARX generate forecasts in which displacement of the predicted series was accentuated against the observed series, similar to the persistent forecast method (Aguado-Rodríguez et al., 2016; Kavasseri & Seetharaman, 2009). This behavior is also notable in the work of Alvarado-Hernández et al. (2020), González-Leiva et al. (2015), and Morales-Velázquez et al. ((Morales-Velázquez et al., 2014).

In the flood that begins at 810 h, EnKF assumes the abrupt change in flow and generates a forecast with a steeper slope than DKF and ARX. Also, in the flow descents between 870 h and 900 h, EnKF converges more precisely on the observed series (Figure 5).

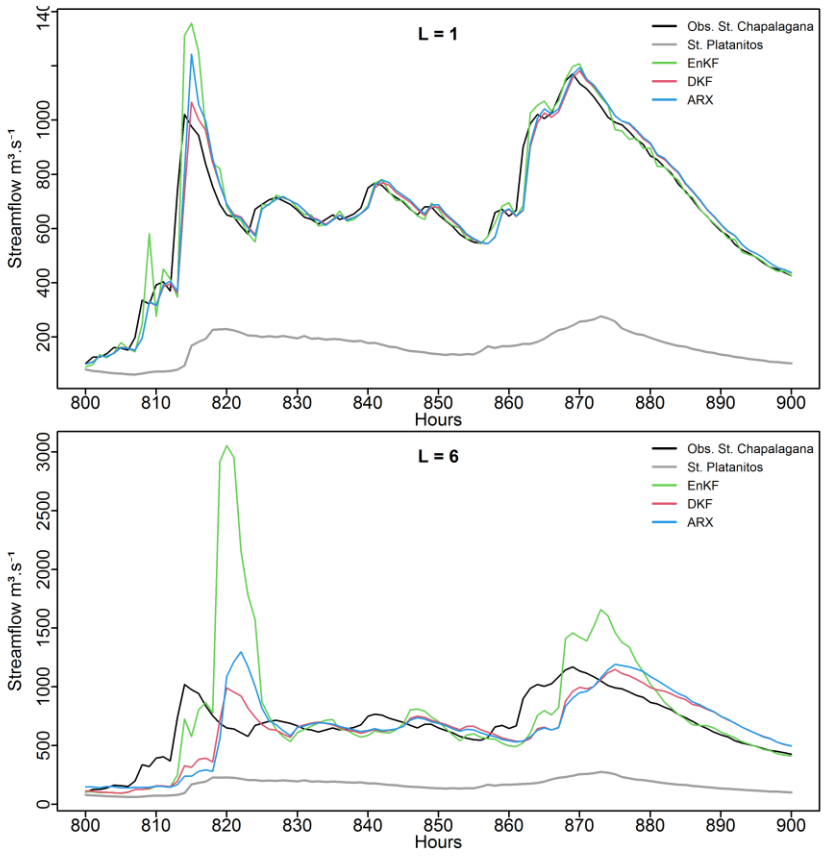



Figure 4.5. Observed flow and flows predicted with EnKF, DKF and ARX (Flood of 4/9/2017 16:00 h to 8/9/2017 20:00 h).
Source: Authors

The DKF fits are similar to those obtained by Alvarado-Hernández et al (2020), who used the same series and a model that integrates ARX-DKF. This model considers that the delay between series is one, while in our study it was updated dynamically throughout the series, favoring fit at peaks and decreases the effect of displacement of the predicted series. EnKF in the six steps obtains lower fits due to overestimation or underestimation at the peaks, with the difference that it achieves better fit in ascents and descents of the observed series. The EnKF algorithm obtains better temporal fit at peaks and converges more precisely on the observed series when the trend persists in a number of hours higher or equal to the evaluated step.

The forecast with EnKF shows overestimations, relative to the observed series. This occurs because we are treating the measurements as a non-linear

phenomenon (Evensen, 1994, 2009), a situation that in step one allows acceptable fit in the entire series, but at step six broad overestimations are found that can affect the quality of the forecast. In the measure that the forecast step increases, frequency of overestimations increases because the new register that serves to update the states also breaks away, and in the L interval there may be changes that are not considered in the initial forecast. The dynamic incorporation of delay time between series (Meng et al., 2017) allows improvement of the fit, given that updating is done with the equivalent event in the exogenous variable (Figure 6).

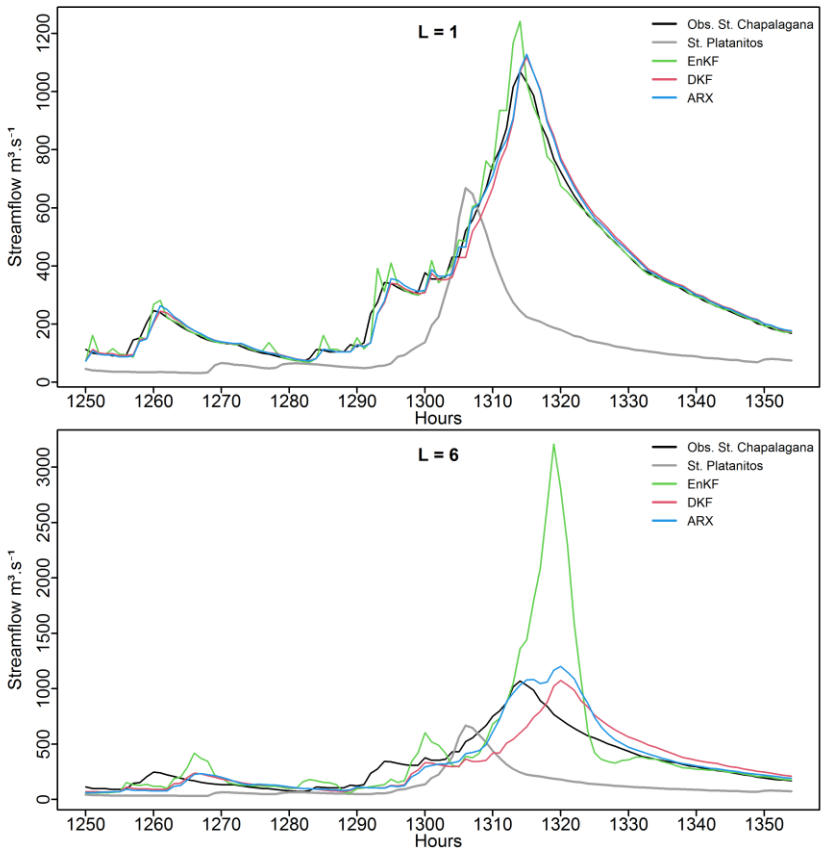



Figure 4.6. Observed flow and flows predicted with EnKF, DKF and ARX (Flood from 23/09/17 10:00 h al 28/09/17 00:00 h).
Source: Authors

Dispersion of the observed series against the predicted series is congruent with the NS index, highlighting the similarities between DKF and ARX. The difference exhibited by EnKF is due to the peaks associated with abrupt changes in flow.

The EnKF algorithm tends to overestimate throughout the series, unlike DFK and ARX that cause a slight tendency to underestimate. Step six, with EnKF, produces large overestimations that are reflected as isolated points above the diagonal in Figure 7.

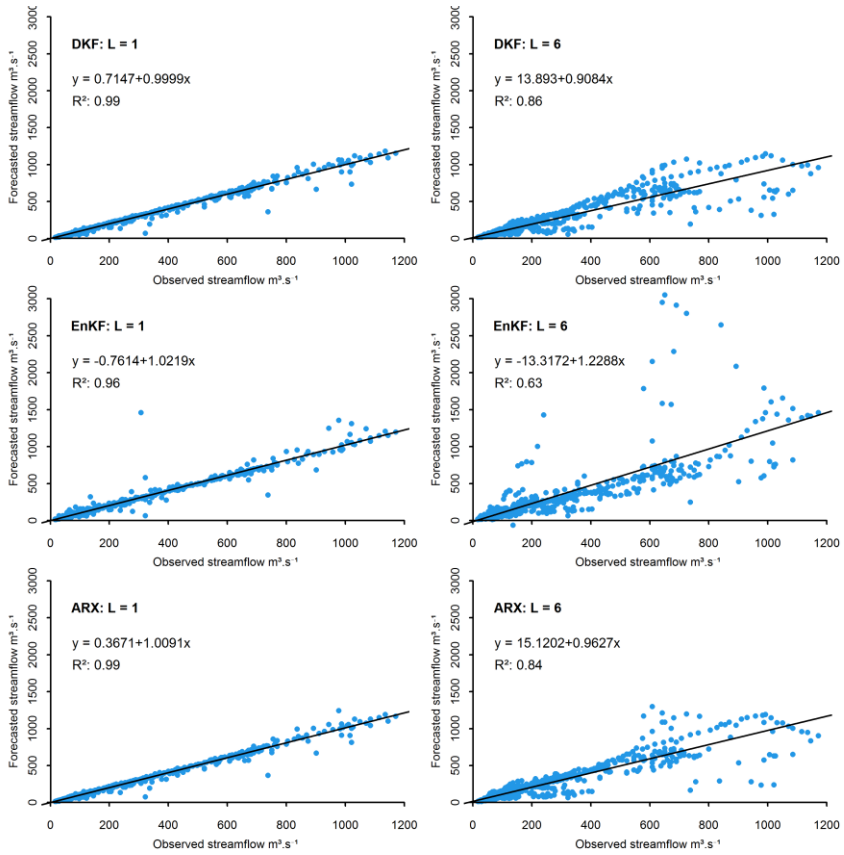



Figure 4.7. Observed versus predicted flows.
Source: Authors

The standardized errors of the forecasts have a symmetrical distribution around zero (Figure 8) (Cryer & Chan, 2008). According to Chatfield (2001), behavior approaching normality is accepted. As shown in Figure 8, there is symmetry and higher concentration of registers in the central area of the Gauss bell curve. The proportion of registers between three standard deviations above and below the mean is more than 97%, and EnKF had the highest values in all the steps.

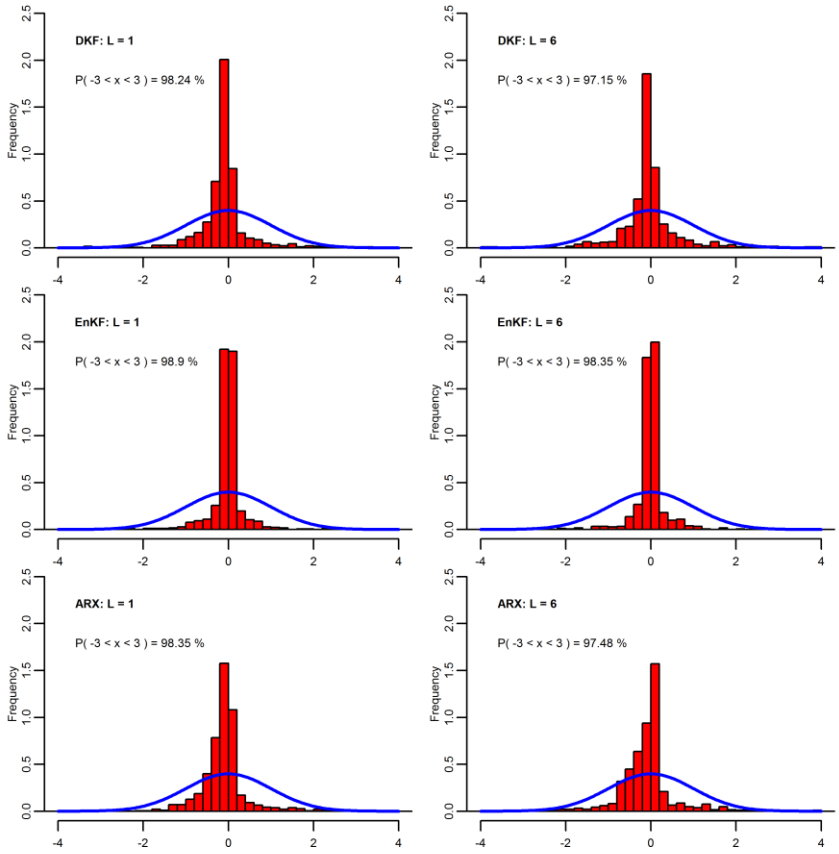



Figure 4.8. Histographs of residuals.
Source: Authors

In the measure that the steps of the forecast increase, the fit decreases, and for EnKF, overestimations and underestimations become more frequent. Introducing coefficients that represent stationality and autocorrelation into the transference function can reduce the over-dimensioned estimations and achieve better fit.

## 4.9 Conclusions

The dynamic updating of delay time, relative to the exogenous variable, enables improving fit in the evaluated algorithms. The EnKF algorithm achieves better convergence on the observed series but generates overestimations of larger magnitude as the step increases and results in a lower degree of fit, as demonstrated with the Nash-Sutcliffe index. The potential of EnKF centers in the convergence and non-linear treatment of abrupt changes in flow. Basically, EnKF helps capture the non-linearity in some parts of the hydrograph and represents

well the timing, or times of occurrence of maximum flows, although they are overestimated.

The use of EnKF for stream flow forecast can be a viable alternative when it is integrated with the autocorrelation analysis so that stationality/stationarity become part of the model and can thus represent the changes associated with day, night, or months of the year. Together, the quantity of registers for updating the states can be increased and, in this way, enable detection of changes in trend during the last registered hours.

To improve the fit of the forecast, it is important to advance in research with step sizes of several hours, for example, groups of six hours using the mean or maximum, in such a way that each step is equivalent to six hours and the forecast at 6 steps equals 36 hours.

## 4.10 Acknowledgements

We thank the Comisión Federal de Electricidad for allowing access to the hydrometeorological database, through the web site administered by the Instituto Nacional de Electricidad y Energías Limpias, during the year 2018.

## 4.11 References

Abaza, M., Anctil, F., Fortin, V., & Turcotte, R. (2015). Exploration of sequential streamflow assimilation in snow dominated watersheds. Advances in Water Resources, 86, 414–424. https://doi.org/10.1016/j.advwatres.2015.10.008

Aguado-Rodríguez, J., Quevedo-Nolasco, A., Castro-Popoca, M., Arteaga-Ramírez, R., Vázquez-Peña, M., & Zamora-Morales, P. (2016). Predicción de variables meteorológicas por medio de modelos ARIMA. Agrociencia, 50(1), 1–13.

Alvarado-Hernández, L., Ibanez-Catillo, L., Ruiz-García, A., González-Leiva, F., & Vázquez-Peña, M. (2020). Pronóstico horario de caudales mediante filtro de kalman discreto en el Río Huaynamota, Nayarit, México. Agrociencia, 54, 295–312.

Bai, L., Chen, Z., Xu, J., & Li, W. (2016). Multi-scale response of runoff to climate fluctuation in the headwater region of Kaidu River in Xinjiang of China. Theoretical and Applied Climatology, 125(3–4), 703–712. https://doi.org/10.1007/s00704-015-1539-2

Box, G., Jenkins, G., Reinsel, G., & Ljung, G. (2016). Time series analysis: Forecasting and control (5th ed.). Wiley.

Brandhorst, N., Erdal, D., & Neuweiler, I. (2017). Soil moisture prediction with the ensemble Kalman filter: Handling uncertainty of soil hydraulic parameters. Advances in Water Resources, 110(February), 360–370. https://doi.org/10.1016/j.advwatres.2017.10.022

Bras, R., & Rodríguez-Iturbe, I. (1985). Random Functions and Hydrology. Addison-Wesley Publishing Company.

Chatfield, C. (2001). Prediction Intervals for Time-Series Forecasting. In J. S. Armstrong (Ed.), A Handbook for Researchers and Practitioners (Vol. 30, pp. 475–494). Springer US. https://doi.org/10.1007/978-0-306-47630-3

Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., Schmidt, J., & Uddstrom, M. J. (2008). Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model. Advances in Water Resources, 31(10), 1309–1324. https://doi.org/10.1016/j.advwatres.2008.06.005

CONAGUA. (2008). Estadisticas del agua en México. SEMARNAT.

Cryer, J. D., & Chan, K.-S. (2008). Time Series Analysis with Applications in R (2nd ed.). Springer.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. Journal of Geophysical Research, 99(C5), 10143. https://doi.org/10.1029/94jc00572

Evensen, G. (2003). The Ensemble Kalman Filter: theoretical formulation and practical implementation. Ocean Dynamics, 53(4), 343–367. https://doi.org/10.1007/s10236-003-0036-9

Evensen, G. (2009). Data Assimilation: The ensemble Kalman filter (2nd ed.). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-03711-5

Genz, A., & Bretz, F. (2009). Computation of Multivariate Normal and t Probabilities. Springer-Verlag.

Gillijns, S., Mendoza, O. B., Chandrasekar, J., De Moor, B. L. R., Bernstein, D. S., & Ridley, A. (2006). What is the ensemble Kalman filter and how well does it work? 2006 American Control Conference, 6 pp. https://doi.org/10.1109/ACC.2006.1657419

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. Tecnologia y Ciencias Del Agua, 6(4), 5–24.

INEGI. (2010). Hidrografía. https://www.inegi.org.mx/temas/hidrografia/default.html#Descargas

IPCC. (2012). Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. In C. B. Field, V. Barros, T. F. Stocker, & Q.

Dahe (Eds.), Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. Cambridge University Press. https://doi.org/10.1017/CBO9781139177245

Jazwinski, A. H. (2007). Stochastic Processes and Filtering Theory. Dover Publications.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. Journal of Basic Engineering, 82(Series D), 35–45.

Kavasseri, R. G., & Seetharaman, K. (2009). Day-ahead wind speed forecasting using f-ARIMA models. Renewable Energy, 34(5), 1388–1393. https://doi.org/10.1016/j.renene.2008.09.006

Liu, Y., & Gupta, H. V. (2007). Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. Water Resources Research, 43(7), 1–18. https://doi.org/10.1029/2006WR005756

Maxwell, D. H., Jackson, B. M., & McGregor, J. (2018). Constraining the ensemble Kalman filter for improved streamflow forecasting. Journal of Hydrology, 560, 127–140. https://doi.org/10.1016/j.jhydrol.2018.03.015

Medina González, H., Hernández Pereira, Y., Santiago Piloto, A. B., & Lau Quan, A. (2015). Modelación de perfil de humedad de suelos empleando un filtro de Kalman de Monte-Carlo. Revista Ciencias Técnicas Agropecuarias, 24(2), 31–37.

Meng, S., Xie, X., & Liang, S. (2017). Assimilation of soil moisture and streamflow observations to improve flood forecasting with considering runoff routing lags. Journal of Hydrology, 550, 568–579. https://doi.org/10.1016/j.jhydrol.2017.05.024

Morales-Velázquez, M. I., Aparicio, J., & Valdes, J. B. (2014). Pronóstico de avenidas utilizando el filtro de kalman discreto. Tecnologia y Ciencias Del Agua, 5(2), 85–110.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. Journal of Hydrology, 10(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Quiroz, K., Collischonn, W., & Paiva, R. C. D. de. (2019). Data assimilation using the ensemble Kalman filter in a distributed hydrological model on the Tocantins River, Brasil. RBRH, 24, 1–15. https://doi.org/10.1590/2318-0331.241920180031

R Core Team. (2020). R: A Language and Environment for Statistical Computing. https://www.r-project.org/

Rafieeinasab, A., Seo, D. J., Lee, H., & Kim, S. (2014). Comparative evaluation of maximum likelihood ensemble filter and ensemble Kalman filter for real-time assimilation of streamflow data into operational hydrologic models. Journal of Hydrology, 519(PD), 2663–2675. https://doi.org/10.1016/j.jhydrol.2014.06.052

Shmueli, G., & Lichtendahl, K. (2016). Practical Time Series Forecasting with R: A Hands On Guide. Axelrod Schnall Publishers.

Simon, D. (2001). Kalman Filtering. Embedded Systems Programming, June, 72–79.

Singh, A., & Zommers, Z. (Eds.). (2014). Reducing Disaster: Early Warning Systems For Climate Change. Springer Netherlands. https://doi.org/10.1007/978-94-017-8598-3

SMN. (2019). Sistema de Información Climática Computarizada CLICOM. Servicio Metereologico Nacional. http://clicom-mex.cicese.mx/malla/index.php

Sun, L., Seidou, O., Nistor, I., & Liu, K. (2016). Review of the Kalman-type hydrological data assimilation. Hydrological Sciences Journal, 61(13), 2348–2366. https://doi.org/10.1080/02626667.2015.1127376

Valdés, J., Mejía, J., & 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.

Welch, G., & Bishop, G. (2006). An Introduction to the Kalman Filter. In University of North Carolina at Chapel Hil (Issue 1). http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.79.6578&rep=rep1&type=pdf

Winkler, R. D., Moore, R. D. D., Redding, T. E., Spittlehouse, D. L., Carlyle-moses, D. E., & Smerdon, B. D. (2010). Hydrologic Processes and Watershed Response. In R. . Pike, T. . Redding, R. . Moore, R. . Winkler, & K. . Bladon (Eds.), Compendium of forest hydrology and geomorphology in British Columbia. BC Ministry of Forests and Range, Forest Science Program, Victoria, BC, and FORREX Forum for Research and Extension in Natural Resources, Kamloops, BC. Land Management Handbook (pp. 133–178). Min. For. Range, For. Sci. Prog., Victoria, B.C. and FORREX Forum for Research and Extension in Natural Resources, Kamloops,.

Xu, J., Li, W., Ji, M., Lu, F., & Dong, S. (2009). A comprehensive approach to characterization of the nonlinearity of runoff in the headwaters of the Tarim River, western China. Hydrological Processes, n/a-n/a. https://doi.org/10.1002/hyp.7484

Zou, L., Zhan, C., Xia, J., Wang, T., & Gippel, C. J. (2017). Implementation of evapotranspiration data assimilation with catchment scale distributed hydrological model via an ensemble Kalman Filter. Journal of Hydrology, 549, 685–702. https://doi.org/10.1016/j.jhydrol.2017.04.036

# 5 CONCLUSIONES GENERALES

El pronóstico de caudales horarios mediante filtros de Kalman permite asimilar datos ruidosos simultáneos, provenientes de varias fuentes. Las estaciones de medición corresponden a las fuentes de datos, su representatividad está en función de la cantidad y distribución en la cuenca. En este trabajo se utilizaron dos estaciones de medición aplicando el filtro de Kalman de manera análoga a un modelo autorregresivo con una variable exógena.

El DKF presenta un comportamiento con tendencia a desplazar la serie pronosticada sobre la serie observada, este desplazamiento tiene una magnitud similar al paso en horas que se busca pronosticar. Cuando se actualiza el tiempo de retraso de la serie de la estación Platanitos frente a la estación Chapalagana se obtiene mejores ajustes, dado que se actualiza con el evento equivalente en diferentes partes de la cuenca que sucede antes de que se evidencie en la estación Chapalagana. Tanto DKF y EnKF obtienen mejor ajuste cuando se hace estimaciones de estados en lugar de estimar valores escalares de caudal, ya que permite aprovechar el retraso entre series como base para la actualización.

El algoritmo de EnKF logra mejor convergencia sobre la serie observada, pero genera sobrestimaciones de mayor magnitud a medida que el paso aumenta y resulta en menor grado de ajuste demostrado con el índice de Nash – Sutcliffe. El potencial de EnKF se centra en la convergencia y tratamiento no lineal de los cambios bruscos de caudal, básicamente EnKF ayuda a captar la no linealidad en algunas partes del hidrograma y a representar bien “el timing” o tiempos de ocurrencias de los caudales máximos, aunque estos son sobreestimados. El no tener suficientes variables de entrada que permitan actualizar el algoritmo puede ser causante de sobrestimaciones. EnKF es el algoritmo de mayor adecuación para estimar estados en sistemas no lineales como los hidrológicos donde suceden los caudales, la ejecución del algoritmo debe hacerse con datos en tiempo real tomados en ubicaciones representativas de la cuenca.

La utilización de EnKF para el pronóstico de caudales es una alternativa que puede ser viable al integrarse con el análisis de la autocorrelación para que la estacionalidad/estacionariedad hagan parte del modelo y así poder representar los cambios asociados al día, noche o meses de año. En este sentido, aumentar la cantidad de registros o términos autorregresivos e incluir mayor número de variables exógenas provenientes de estaciones de medición distribuidas en la cuenca puede permitir detectar los cambios de tendencia de las últimas horas y limitar las sobrestimaciones. La integración del EnKF con modelos hidrológicos distribuidos puede permitir simular y comprender la dinámica de los sistemas físicos que intervienen en la conformación de escorrentías. Por otra parte, se puede adelantar trabajos de investigación con tamaños de paso de varias horas, por ejemplo, grupos de seis horas mediante la media o máximo, de tal forma que cada paso equivale a seis horas y el pronóstico a 6 pasos equivale a 36 horas

# 6 LITERATURA CITADA GENERAL


Abaza, M., Anctil, F., Fortin, V., & Turcotte, R. (2014). Sequential streamflow assimilation for short-term hydrological ensemble forecasting. *Journal of Hydrology*, *519*(PD), 2692–2706. https://doi.org/10.1016/j.jhydrol.2014.08.038

Abaza, M., Anctil, F., Fortin, V., & Turcotte, R. (2015). Exploration of sequential streamflow assimilation in snow dominated watersheds. *Advances in Water Resources*, *86*, 414–424. https://doi.org/10.1016/j.advwatres.2015.10.008

Aguado-Rodríguez, J., Quevedo-Nolasco, A., Castro-Popoca, M., Arteaga-Ramírez, R., Vázquez-Peña, M., & Zamora-Morales, P. (2016). Predicción de variables meteorológicas por medio de modelos ARIMA. *Agrociencia*, *50*(1), 1–13.

Alvarado-Hernández, L., Ibanez-Catillo, L., Ruiz-García, A., González-Leiva, F., & Vázquez-Peña, M. (2020). Pronóstico horario de caudales mediante filtro de kalman discreto en el Río Huaynamota, Nayarit, México. *Agrociencia*, *54*, 295–312.

Aparicio, F. (2017). *Fundamentos de hidrología de superficie*. Limusa.

Archer, E., Park, I. M., Buesing, L., Cunningham, J., & Paninski, L. (2015). Black box variational inference for state space models. *6th International Conference on Learning Representations*, 1–11. http://arxiv.org/abs/1511.07367

Bader, K., Lussier, B., & Schön, W. (2017). A fault tolerant architecture for data fusion: A real application of Kalman filters for mobile robot localization. *Robotics and Autonomous Systems*, *88*, 11–23. https://doi.org/10.1016/j.robot.2016.11.015

Bai, L., Chen, Z., Xu, J., & Li, W. (2016). Multi-scale response of runoff to climate

fluctuation in the headwater region of Kaidu River in Xinjiang of China. *Theoretical and Applied Climatology*, *125*(3–4), 703–712. https://doi.org/10.1007/s00704-015-1539-2

Bishop, C. H., Etherton, B. J., & Majumdar, S. J. (2001). Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. *Monthly Weather Review*, *129*(3), 420–436. https://doi.org/10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2

Box, G., Jenkins, G., Reinsel, G., & Ljung, G. (2016). *Time series analysis: Forecasting and control* (5th ed.). Wiley.

Brammer, R. (1974). Real-Time shipboard orbit determination using Kalman Filtering Techniques. *IEEE Transactions on Aerospace and Electronic Systems*, *AES-10*(4), 492–496. https://doi.org/10.1109/TAES.1974.307803

Brandhorst, N., Erdal, D., & Neuweiler, I. (2017). Soil moisture prediction with the ensemble Kalman filter: Handling uncertainty of soil hydraulic parameters. *Advances in Water Resources*, *110*(February), 360–370. https://doi.org/10.1016/j.advwatres.2017.10.022

Bras, R., & Rodríguez-Iturbe, I. (1985). *Random Functions and Hydrology*. Addison-Wesley Publishing Company.

Bravo, L. (2016). *Implementacion de redes neuronales para generar oleaje utilizando datos de viento en la costa de Sisal, Yucatan*. Universidad Nacional Autonoma de Mexico.

Chatfield, C. (2001). Prediction Intervals for Time-Series Forecasting. In J. S. Armstrong (Ed.), *A Handbook for Researchers and Practitioners* (Vol. 30, pp. 475–494). Springer US. https://doi.org/10.1007/978-0-306-47630-3

Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., Schmidt, J., & Uddstrom, M. J. (2008). Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in

a distributed hydrological model. *Advances in Water Resources*, *31*(10), 1309–1324. https://doi.org/10.1016/j.advwatres.2008.06.005

Cloke, H. L., & Pappenberger, F. (2009). Ensemble flood forecasting: A review. *Journal of Hydrology*, *375*(3–4), 613–626. https://doi.org/10.1016/j.jhydrol.2009.06.005

CONAGUA. (2008). *Estadisticas del agua en México*. SEMARNAT.

Cryer, J. D., & Chan, K.-S. (2008). *Time Series Analysis with Applications in R* (2nd ed.). Springer.

Cuo, L., Pagano, T. C., & Wang, Q. J. (2011). A Review of Quantitative Precipitation Forecasts and Their Use in Short- to Medium-Range Streamflow Forecasting. *Journal of Hydrometeorology*, *12*(5), 713–728. https://doi.org/10.1175/2011JHM1347.1

DeChant, C. M., & Moradkhani, H. (2012). Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting. *Water Resources Research*, *48*(4), 1–15. https://doi.org/10.1029/2011WR011011

Dumedah, G., & Coulibaly, P. (2013). Evaluating forecasting performance for data assimilation methods: The ensemble Kalman filter, the particle filter, and the evolutionary-based assimilation. *Advances in Water Resources*, *60*, 47–63. https://doi.org/10.1016/j.advwatres.2013.07.007

Elsafi, S. H. (2014). Artificial Neural Networks (ANNs) for flood forecasting at Dongola Station in the River Nile, Sudan. *Alexandria Engineering Journal*, *53*(3), 655–662. https://doi.org/10.1016/j.aej.2014.06.010

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, *99*(C5), 10143. https://doi.org/10.1029/94jc00572

Evensen, G. (2003). The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics*, *53*(4), 343–367. https://doi.org/10.1007/s10236-003-0036-9

Evensen, G. (2009). *Data Assimilation: The ensemble Kalman filter* (2nd ed.). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-03711-5

Evensen, G., & van Leeuwen, P. J. (2000). An ensemble Kalman smoother for nonlinear dynamics. *Monthly Weather Review*, *128*(6), 1852–1867. https://doi.org/10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2

Fan, Y. R., Huang, G. H., Baetz, B. W., Li, Y. P., Huang, K., Chen, X., & Gao, M. (2017). Development of integrated approaches for hydrological data assimilation through combination of ensemble Kalman filter and particle filter methods. *Journal of Hydrology*, *550*, 412–426. https://doi.org/10.1016/j.jhydrol.2017.05.010

Garces, L., & Gomez, O. (2004). Inferencia estadística utilizando redes neuronales artificiales. *Scientia et Technica*, *10*(26), 31–36.

Genz, A., & Bretz, F. (2009). *Computation of Multivariate Normal and t Probabilities*. Springer-Verlag.

Gillijns, S., Mendoza, O. B., Chandrasekar, J., De Moor, B. L. R., Bernstein, D. S., & Ridley, A. (2006). What is the ensemble Kalman filter and how well does it work? *2006 American Control Conference*, 6 pp. https://doi.org/10.1109/ACC.2006.1657419

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. *Tecnologia y Ciencias Del Agua*, *6*(4), 5–24.

Gouweleeuw, B. T., Thielen, J., Franchello, G., De Roo, A. P. J., & Buizza, R. (2005). Flood forecasting using medium-range probabilistic weather prediction. *Hydrology and Earth System Sciences*, *9*(4), 365–380.

https://doi.org/10.5194/hess-9-365-2005

Han, S., & Coulibaly, P. (2017). Bayesian flood forecasting methods: A review. *Journal of Hydrology*, *551*, 340–351. https://doi.org/10.1016/j.jhydrol.2017.06.004

Hapuarachchi, H. A. P., Wang, Q. J., & Pagano, T. C. (2011). A review of advances in flash flood forecasting. *Hydrological Processes*, *25*(18), 2771–2784. https://doi.org/10.1002/hyp.8040

Herr, H. D., & Krzysztofowicz, R. (2010). Bayesian ensemble forecast of river stages and ensemble size requirements. *Journal of Hydrology*, *387*(3–4), 151–164. https://doi.org/10.1016/j.jhydrol.2010.02.024

INEGI. (2010). *Hidrografía*. https://www.inegi.org.mx/temas/hidrografia/default.html#Descargas

Infante, S., & Zaraté, G. (2012). *Metodos estadisticos: un enfoque interdisciplinario*. La Gaya Ciencia.

IPCC. (2012). Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. In C. B. Field, V. Barros, T. F. Stocker, & Q. Dahe (Eds.), *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*. Cambridge University Press. https://doi.org/10.1017/CBO9781139177245

IPCC. (2013). *Cambio climático 2013: Bases físicas.* IPCC.

Javaheri, A., Babbar-Sebens, M., Miller, R. N., Hallett, S. L., & Bartholomew, J. L. (2019). An adaptive ensemble Kalman filter for assimilation of multi-sensor, multi-modal water temperature observations into hydrodynamic model of shallow rivers. *Journal of Hydrology*, *572*, 682–691. https://doi.org/10.1016/j.jhydrol.2019.03.036

Jazwinski, A. H. (2007). *Stochastic Processes and Filtering Theory*. Dover

Publications.

Jonsson, G., & Palsson, O. P. (1994). An Application of Extended Kalman Filtering to Heat Exchanger Models. *Journal of Dynamic Systems, Measurement, and Control*, *116*(2), 257. https://doi.org/10.1115/1.2899218

Julier, S. J., & Uhlmann, J. K. (1997). *A new extension of the Kalman filter to nonlinear systems* (I. Kadar (Ed.); p. 182). https://doi.org/10.1117/12.280797

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, *82*(Series D), 35–45.

Kaps, M., & Lamberson, W. (2004). *Biostatistics for animal science*. CABI Publishing.

Kavasseri, R. G., & Seetharaman, K. (2009). Day-ahead wind speed forecasting using f-ARIMA models. *Renewable Energy*, *34*(5), 1388–1393. https://doi.org/10.1016/j.renene.2008.09.006

Khan, E., & Khan, F. (2012). Summary for Policymakers. *International Journal of Advanced Computer Science and Applications*, *3*(6), 12–15. https://www.cambridge.org/core/product/identifier/CBO9781107415324A009/type/book_part

Kotecha, J. H., & Djuric, P. M. (2003). Gaussian particle filtering. *Proceedings of the 11th IEEE Signal Processing Workshop on Statistical Signal Processing (Cat. No.01TH8563)*, 429–432. https://doi.org/10.1109/SSP.2001.955314

Krener, A. J., & Duarte, A. (1996). A hybrid computational approach to nonlinear estimation. *Proceedings of 35th IEEE Conference on Decision and Control*, *2*, 1815–1819. https://doi.org/10.1109/CDC.1996.572831

Kundzewicz, Z. W., & Napiórkowski, J. J. (1986). Nonlinear models of dynamic hydrology. *Hydrological Sciences Journal*, *31*(2), 163–185. https://doi.org/10.1080/02626668609491038

Lawson, W. G., & Hansen, J. A. (2004). Implications of Stochastic and Deterministic Filters as Ensemble-Based Data Assimilation Methods in Varying Regimes of Error Growth. *Monthly Weather Review*, *132*(8), 1966–1981. https://doi.org/10.1175/1520-0493(2004)132<1966:IOSADF>2.0.CO;2

Leach, J. M., Kornelsen, K. C., & Coulibaly, P. (2018). Assimilation of near-real time data products into models of an urban basin. *Journal of Hydrology*, *563*(April), 51–64. https://doi.org/10.1016/j.jhydrol.2018.05.064

Leutbecher, M. (2019). Ensemble size: How suboptimal is less than infinity? *Quarterly Journal of the Royal Meteorological Society*, *145*(S1), 107–128. https://doi.org/10.1002/qj.3387

Liu, Weerts, A. H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenberg, D., Smith, P., van Dijk, A. I. J. M., van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O., & Restrepo, P. (2012). Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. *Hydrology and Earth System Sciences*, *16*(10), 3863–3887. https://doi.org/10.5194/hess-16-3863-2012

Liu, Y., & Gupta, H. V. (2007). Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resources Research*, *43*(7), 1–18. https://doi.org/10.1029/2006WR005756

Ljung, L. (1999). *System Identification: Theory for the User*. Prentice Hall PTR. https://books.google.com.mx/books?id=nHFoQgAACAAJ

Lopez, J. A., & Caicedo, E. F. (2009). *Una aproximacion practica a las redes neuronales artificiales*. Universidad del Valle.

Lorenz, E. N. (1969). The predictability of a flow which possesses many scales of motion. *Tellus*, *21*(3), 289–307. https://doi.org/10.3402/tellusa.v21i3.10086

Martin, B., & Sanz, A. (2017). *Redes neuronales y sistemas borrosos* (15th ed.). Alfaomega.

Maxwell, D. H., Jackson, B. M., & McGregor, J. (2018). Constraining the ensemble Kalman filter for improved streamflow forecasting. *Journal of Hydrology*, *560*, 127–140. https://doi.org/10.1016/j.jhydrol.2018.03.015

Maybeck, P. (1979). *Stochastic models , estimation, and control*. Academic Press.

Medina González, H., Hernández Pereira, Y., Santiago Piloto, A. B., & Lau Quan, A. (2015). Modelación de perfil de humedad de suelos empleando un filtro de Kalman de Monte-Carlo. *Revista Ciencias Técnicas Agropecuarias*, *24*(2), 31–37.

Mejía, D., & Eliana, T. (2004). Pronóstico de ventas usando redes neuronales. *Scientia et Technica*, *10*(26), 25–30.

Meng, S., Xie, X., & Liang, S. (2017). Assimilation of soil moisture and streamflow observations to improve flood forecasting with considering runoff routing lags. *Journal of Hydrology*, *550*, 568–579. https://doi.org/10.1016/j.jhydrol.2017.05.024

Moradkhani, H., Sorooshian, S., Gupta, H. V., & Houser, P. R. (2005). Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources*, *28*(2), 135–147. https://doi.org/10.1016/j.advwatres.2004.09.002

Morales-Velázquez, M. I., Aparicio, J., & Valdes, J. B. (2014). Pronóstico de avenidas utilizando el filtro de kalman discreto. *Tecnologia y Ciencias Del Agua*, *5*(2), 85–110.

Muluye, G. Y. (2011). Improving long-range hydrological forecasts with extended Kalman filters. *Hydrological Sciences Journal*, *56*(7), 1118–1128. https://doi.org/10.1080/02626667.2011.608068

Naghettini, M., & De Andrade, É. (2007). *Hidrologia estatística* (1st ed.). CPRM - Serviço Geológico do Brasil.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Odry, Á., Fullér, R., Rudas, I. J., & Odry, P. (2018). Kalman filter for mobile-robot attitude estimation: Novel optimized and adaptive solutions. *Mechanical Systems and Signal Processing*, *110*, 569–589. https://doi.org/10.1016/j.ymssp.2018.03.053

Pathiraja, S., Marshall, L., Sharma, A., & Moradkhani, H. (2016). Hydrologic modeling in dynamic catchments: A data assimilation approach. *Water Resources Research*, *52*(5), 3350–3372. https://doi.org/10.1002/2015WR017192

Pujol Reig, L. (2009). *Prediccion de caudales en tiempo real en grandes cuencas utilizando redes neuronales artificiales*. Universidad Politecnica de Madrid.

Quadros, R. S. de, Harter, F. P., Buske, D., & Pereira, L. S. (2016). Data assimilation by ensemble Kalman filter with the Lorenz equations. *Ciência e Natura*, *38*(July), 190. https://doi.org/10.5902/2179460X20158

Quiroz, K., Collischonn, W., & Paiva, R. C. D. de. (2019). Data assimilation using the ensemble Kalman filter in a distributed hydrological model on the Tocantins River, Brasil. *RBRH*, *24*, 1–15. https://doi.org/10.1590/2318-0331.241920180031

R Core Team. (2020). *R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. http://www.R-project.org/*. https://www.r-project.org/

Rafieeinasab, A., Seo, D. J., Lee, H., & Kim, S. (2014). Comparative evaluation of maximum likelihood ensemble filter and ensemble Kalman filter for real-

time assimilation of streamflow data into operational hydrologic models. *Journal of Hydrology*, *519*(PD), 2663–2675. https://doi.org/10.1016/j.jhydrol.2014.06.052

Rajagopalan, B., Salas, J., & Lall, U. (2010). Stochastic methods for modeling precipitation and streamflow. In B. Sivakumar & R. Berndtsson (Eds.), *Advances in Data-Based Approaches for Hydrologic Modeling and Forecasting* (pp. 17–52). WORLD SCIENTIFIC. https://doi.org/10.1142/7783

Reichle, R. H. (2008). Data assimilation methods in the Earth sciences. *Advances in Water Resources*, *31*(11), 1411–1418. https://doi.org/10.1016/j.advwatres.2008.01.001

Reichle, R. H., Walker, J. P., Koster, R. D., & Houser, P. R. (2002). Extended versus Ensemble Kalman Filtering for Land Data Assimilation. *Journal of Hydrometeorology*, *3*(6), 728–740. https://doi.org/10.1175/1525-7541(2002)003<0728:EVEKFF>2.0.CO;2

RStudio Team. (2020). *RStudio: Integrated Development Environment for R*. http://www.rstudio.com/

Saba, I., Cedeño, F., & Ortega, J. (2010). Filtros con aprendizaje de parametros para optimizar modelos de redes nuronales en la prediccion de series de precipitaciones. *Ingenieria Industiral. Actualidad y Nuevas Tendencias*, *2*(5), 37–56.

Samuel, J., Coulibaly, P., Dumedah, G., & Moradkhani, H. (2014). Assessing model state and forecasts variation in hydrologic data assimilation. *Journal of Hydrology*, *513*, 127–141. https://doi.org/10.1016/j.jhydrol.2014.03.048

Shmueli, G., & Lichtendahl, K. (2016). *Practical Time Series Forecasting with R: A Hands On Guide*. Axelrod Schnall Publishers.

Shumway, R. H., & Stoffer, D. S. (2017). *Time Series Analysis and Its Applications*. Springer International Publishing. https://doi.org/10.1007/978-

3-319-52452-8

Simon, D. (2001). Kalman Filtering. *Embedded Systems Programming*, *June*, 72–79.

Singh, A., & Zommers, Z. (Eds.). (2014). *Reducing Disaster: Early Warning Systems For Climate Change*. Springer Netherlands. https://doi.org/10.1007/978-94-017-8598-3

SMN. (2019). *Sistema de Información Climática Computarizada CLICOM*. Servicio Metereologico Nacional. http://clicom-mex.cicese.mx/malla/index.php

Sun, L., Seidou, O., Nistor, I., Goïta, K., & Magagi, R. (2016). Simultaneous assimilation of in situ soil moisture and streamflow in the SWAT model using the Extended Kalman Filter. *Journal of Hydrology*, *543*, 671–685. https://doi.org/10.1016/j.jhydrol.2016.10.040

Sun, L., Seidou, O., Nistor, I., & Liu, K. (2016). Review of the Kalman-type hydrological data assimilation. *Hydrological Sciences Journal*, *61*(13), 2348–2366. https://doi.org/10.1080/02626667.2015.1127376

Te Chow, V., Maidment, D. R., Mays, L. W., & Saldarriaga, J. G. (1996). *Hidrología aplicada*. McGraw-Hill.

Torres, N., Hernandez, C., & Pedraza, L. (2011). Redes neuronales y predicción de tráfico. *Tecnura*, *15*(29), 90–97.

Valdés, J., Mejía, J., & 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.*

van der Merwe, R., & Wan, E. A. (2001). The Unscented Kalman Filter. In S. Haykin (Ed.), *Kalman Filtering and Neural Networks* (Vol. 6, pp. 221–279). JOHN WILEY & SONS, INC. https://doi.org/10.1002/0471221546.ch7

Vrugt, J. A., & Robinson, B. A. (2007). Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging. *Water Resources Research*, *43*(1), 1–15. https://doi.org/10.1029/2005WR004838

Wang, D., Chen, Y., & Cai, X. (2009). State and parameter estimation of hydrologic models using the constrained ensemble Kalman filter. *Water Resources Research*, *45*(11), 1–13. https://doi.org/10.1029/2008WR007401

Wang, S., Huang, G. H., Baetz, B. W., Cai, X. M., Ancell, B. C., & Fan, Y. R. (2017). Examining dynamic interactions among experimental factors influencing hydrologic data assimilation with the ensemble Kalman filter. *Journal of Hydrology*, *554*, 743–757. https://doi.org/10.1016/j.jhydrol.2017.09.052

Welch, G., & Bishop, G. (2006). An Introduction to the Kalman Filter. In *University of North Carolina at Chapel Hil* (Issue 1).

Whitaker, J. S., & Hamill, T. M. (2002). Ensemble Data Assimilation without Perturbed Observations. *Monthly Weather Review*, *130*(7), 1913–1924. https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2

Willner, D., Chang, C., & Dunn, K. (1976). Kalman filter algorithms for a multi-sensor system. *1976 IEEE Conference on Decision and Control Including the 15th Symposium on Adaptive Processes*, 570–574. https://doi.org/10.1109/CDC.1976.267794

Winkler, R. D., Moore, R. D. D., Redding, T. E., Spittlehouse, D. L., Carlyle-moses, D. E., & Smerdon, B. D. (2010). Hydrologic Processes and Watershed Response. In R. . Pike, T. . Redding, R. . Moore, R. . Winkler, & K. . Bladon (Eds.), *Compendium of forest hydrology and geomorphology in British Columbia. BC Ministry of Forests and Range, Forest Science Program, Victoria, BC, and FORREX Forum for Research and Extension in Natural Resources, Kamloops, BC. Land Management Handbook* (pp. 133–

178). Min. For. Range, For. Sci. Prog., Victoria, B.C. and FORREX Forum for Research and Extension in Natural Resources, Kamloops,.

WMO. (2011). *Manual on flood forecasting and warming - WMO-No. 1072*. World Metereological Organization.

Wu, X., & Wang, Y. (2012). Extended and Unscented Kalman filtering based feedforward neural networks for time series prediction. *Applied Mathematical Modelling*, *36*(3), 1123–1131. https://doi.org/10.1016/j.apm.2011.07.052

Xiong, M., Liu, P., Cheng, L., Deng, C., Gui, Z., Zhang, X., & Liu, Y. (2019). Identifying time-varying hydrological model parameters to improve simulation efficiency by the ensemble Kalman filter: A joint assimilation of streamflow and actual evapotranspiration. *Journal of Hydrology*, *568*, 758–768. https://doi.org/10.1016/j.jhydrol.2018.11.038

Xu, J., Li, W., Ji, M., Lu, F., & Dong, S. (2009). A comprehensive approach to characterization of the nonlinearity of runoff in the headwaters of the Tarim River, western China. *Hydrological Processes*, n/a-n/a. https://doi.org/10.1002/hyp.7484

Zealand, C. M., Burn, D. H., & Simonovic, S. P. (1999). Short term streamflow forecasting using artificial neural networks. *Jounal of Hydrology*, *214*, 32–48.

Zhang, C., Yan, Q., Kuczyńska-Kippen, N., & Gao, X. (2020). An Ensemble Kalman Filter approach to assess the effects of hydrological variability, water diversion, and meteorological forcing on the total phosphorus concentration in a shallow reservoir. *Science of The Total Environment*, *724*, 138215. https://doi.org/10.1016/j.scitotenv.2020.138215

Zhu, P., Shi, L., Zhu, Y., Zhang, Q., Huang, K., & Williams, M. (2017). Data assimilation of soil water flow via ensemble Kalman filter: Infusing soil moisture data at different scales. *Journal of Hydrology*, *555*, 912–925. https://doi.org/10.1016/j.jhydrol.2017.10.078

Zou, L., Zhan, C., Xia, J., Wang, T., & Gippel, C. J. (2017a). Implementation of evapotranspiration data assimilation with catchment scale distributed hydrological model via an ensemble Kalman Filter. *Journal of Hydrology*, *549*, 685–702. https://doi.org/10.1016/j.jhydrol.2017.04.036

Zou, L., Zhan, C., Xia, J., Wang, T., & Gippel, C. J. (2017b). Implementation of evapotranspiration data assimilation with catchment scale distributed hydrological model via an ensemble Kalman Filter. *Journal of Hydrology*, *549*, 685–702. https://doi.org/10.1016/j.jhydrol.2017.04.036