Research Article/ Civil and Sanitary Engineering https://doi.org/10.15446/ing.investig.90023

Ensemble Kalman Filter for Hourly Streamflow Forecasting in Huaynamota River, Nayarit, México

Filtro de Kalman de Conjuntos para pronóstico de caudales horarios en el río Huaynamota, Nayarit, México

Ildefonso Narváez-Ortiz[1], Laura Ibáñez-Castillo[2], Ramon Arteaga-Ramírez[3], and Mario Vázquez-Peña[4]

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 flow forecasting using the Ensemble Kalman Filter, given its relative ease of implementation and lower computational effort in comparison with other techniques. The hourly streamflow of the Chapalagana station was forecasted based on that of the Platanitos station in northwestern México. 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 coefficients of the Discrete Kalman filter were 0,99 and 0,85 for steps 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 (EnKF) obtained 0,99 and 0,05 in steps one and six. However, it converges on the observed series with the limitation of considerable overestimation in higher steps. All three algorithms have equal statistical adjustment values in step one, and there are progressive differences in further steps, where ARX and DKF remain similar and EnKF is differentiated by the overestimation. EnKF enables capturing non-linearity in sudden streamflow changes but generates overestimation at the peaks.

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

RESUMEN

Los fenómenos hidrológicos se caracterizan por conformar un sistema dinámico no lineal, y 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, dada su relativa facilidad de implementación y menor esfuerzo computacional en contraste con otras técnicas. Se pronosticó el caudal horario de la estación Chapalagana en función del 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. Los coeficientes de Nash-Sutcliffe para el Filtro de Kalman Discreto fueron de 0,99 y 0,85 en los pasos 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 (EnKF) obtuvo 0,99 y 0,05 en los pasos uno y seis. No obstante, este converge sobre la serie observada con la limitante de sobrestimación considerable en pasos superiores. Los tres algoritmos tienen igual valor de ajuste estadístico en el paso uno, y se dan diferencias progresivas en pasos sucesivos, donde ARX y DKF se mantienen similares y EnKF se diferencia por la sobrestimación. EnKF permite captar la no linealidad en los cambios bruscos de caudal, pero genera sobrestimación en los picos.

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

Received: August 20th, 2020
Accepted: September 21th, 2021

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 on streamflow behavior becomes an indispensable tool for the administration of dams and disaster risk management (IPCC, 2012; Singh and Zommers, 2014). Different methods have been used for streamflow forecasting, such as autoregressive methods, neural networks (Box *et al*., 2016; Shmueli and Lichtendahl, 2016), and, more recently, data assimilation methods 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

[1] PhD in Agricultural Engineering and Integral Water Use at Universidad Autónoma de Chapingo, Texcoco, México. Affiliation: International University of American Tropics – Unitrópico, Casanare Colombia. Email: ildenarvaez@unitropico.edu.co

[2] PhD in Civil Engineering from Texas A&M University. Affiliation: Universidad Autónoma de Chapingo, Texcoco, México. Email: libacas@gmail.com

[3] PhD in Hydrosciences from Colegio de Postgraduados. Affiliation: Universidad Autónoma de Chapingo, Texcoco, México. Email: arteagar@correo.chapingo.mx

[4] PhD in Statistics from Colegio de Postgraduados. Affiliation: Universidad Autónoma de Chapingo, Texcoco, México. Email: mvazquezp@chapingo.mx

How to cite: Narváez-Ortiz, I. Ibáñez-Castillo, L. Arteaga-Ramírez, R., and Vázquez-Peña, M. (2022). Ensemble Kalman Filter for Hourly Streamflow Forecasting in Huaynamota River, Nayarit, México. *Ingeniería e Investigación, 42*(3), e90023. https://doi.org/10.15446/ing.investig.90023

Filter (EnKF) (Evensen, 1994, 2009; Gillijns *et al*., 2006) has been widely used as a method of assimilation (Liu and Gupta, 2007; Maxwell *et al*., 2018; Sun *et al*., 2016), with little evaluation in forecasting flows. EnKF is an extension of the Discrete Kalman Filter (DKF) (Kalman, 1960) and a computationally less demanding alternative to the Extended Kalman Filter (EKF) for treating 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*., 2017), and soil moisture (Brandhorst *et al*., 2017; Meng *et al*., 2017). Moreover, it has been evaluated while 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 a 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 for data assimilation favors the 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 with 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 the incorporation of registers from diverse sources, as well as continuous updates (Box *et al*., 2016; Welch and Bishop, 2006).

To determine whether the algorithms for identifying non-linear dynamic systems allow forecasting short-term streamflow (6 h), this study evaluates the fit of the series predicted by algorithms of the EnKF, the DKF, and the first-order autoregressive model with first order exogenous variable (ARX(1, I)), 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.

Materials and methods

This study was conducted in a tributary of the Huaynamota River, also known as the Chapalagana or Atengo River, located in northwestern México 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 km^2. The altitude varies from 219 to 3 147 masl. The concentration time is 39,88 h, the mean annual precipitation is 707 mm, and the mean annual temperature is 17,9 °C (SMN, 2019).

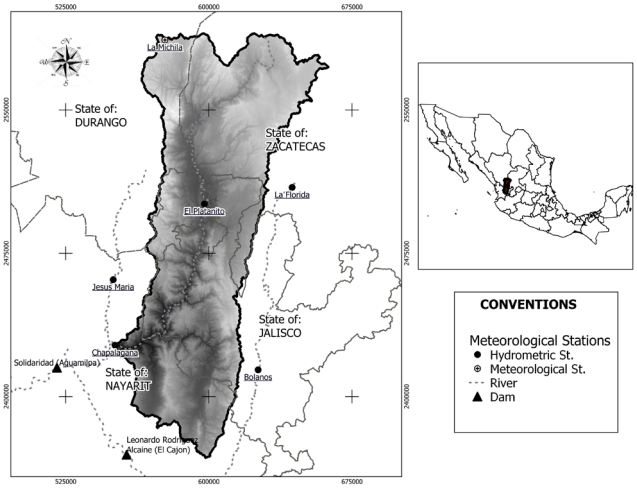

Figure 1. Location of the study area
Source: Authors

The Huaynamota River contributes to generating electricity. The Solidaridad dam (also known as Aguamilpa), located on the Lerma-Santiago River 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 EnKF (Evensen, 1994), DKF (Kalman, 1960), and ARX(1,1) (Bras and Rodríguez-Iturbe, 1985) algorithms. Forecasts were made at 1, 2, 3, 4, 5, and 6 h (L steps forward) of the flows at the Chapalagana station as a function of the flows at the Platanito station, located 100 km upstream from Chapalagana, as the exogenous variable. The hourly streamflow series between 9:00 hours on August 2 and 0:00 hours on September 28, 2017, were used, for a total of 1 360 registers supplied by Federal Electricity Comission (CFE).

EnKF, DKF, and ARX were implemented through R routines (R Core Team, 2021), 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 corresponding to the response function of the basin, or Instantaneous Unitary Hydrograph (IUH) (Valdés *et al*., 1980). Values were estimated which correspond to the columns of the IUH. Multiplying these values by those of the measured series results in an estimation. In the three algorithms, the last observations of each series were considered (Valdés *et al*., 1980). The ARX model was recursively implemented based on a fraction of a series with 100 registers.

By means of 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 square 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 the mvtnorm package (Multivariant Normal and t Distributions) (Genz and Bretz, 2009). In the evaluated Kalman Filter algorithms, the Q variance was assumed to be constant (Simon, 2001) with a value of zero (Morales-Velázquez *et al*., 2014), and the R was assumed to have a near-zero value (0,01) in order to confer flexibility to the convergence of the algorithm (Welch and Bishop, 2006). With these values, the covariance matrices were created.

The fit reached by each algorithm was evaluated using the Nash-Sutcliffe coefficient (NS) (Nash and Sutcliffe, 1970) and the RMSE (Morales-Velázquez *et al*., 2014), as expressed by Equations (1) and (2). The assumed normality of errors was verified using graphs (González-Leiva *et al*., 2015).

$$NS = 1 - \frac{\sum_{i=1}^{i=n}\left(\hat{y}_i - y_i\right)^2}{\sum_{i=1}^{i=n}\left(\overline{y}_i - \hat{y}_i\right)^2} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2}{n}} \quad (2)$$

where y_i is the forecasted data, y_i is the observed data, $\hat{y}_i$ is the average of the observed data, $\hat{y}_i$ is average of the forecasted data, and n is the amount of observations.

Ensemble Kalman Filter

To extend the functionality of the DKF (Kalman, 1960) and deal with non-linear dynamic systems, the EKF, among others, has been proposed (Jazwinski, 2007; Welch and Bishop, 2006), which includes the EnKF (Evensen, 1994, 2009). The EnKF emerges as an alternative to the EKF, which has a high computational demand (Evensen, 1994, 2009) and is a sub-optimal estimator that, via 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 the random values that make up the first matrix $x_{k^{ai}}$, thus obtaining the first forecast via the measurement equation. The error matrices of the forecast were calculated against the new measurement, which is the input for the analysis equations, where the states are updated as new measurements are entered.

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, in the upstream from the Chapalagana station, there are no structures (e.g., dams) that have a direct impact on streamflow. The errors v_k^i and v_k^i correspond to the noise contained by the process 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, i.e., random values under normal distribution that correspond to $i = 1, 2, \ldots, \ldots, q$. 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 multiplying the h matrix and the x_{k-1}^{ai} and applying the measurement equation.

Discrete Kalman Filter

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

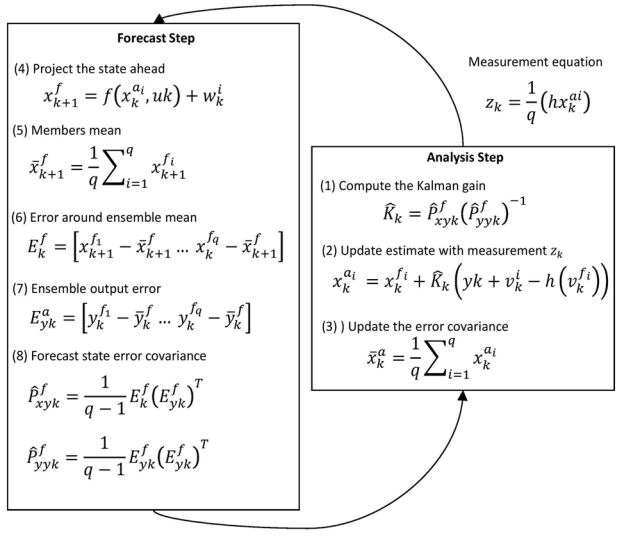

Figure 2. EnKF algorithm
Source: Gillijns *et al*. (2006)

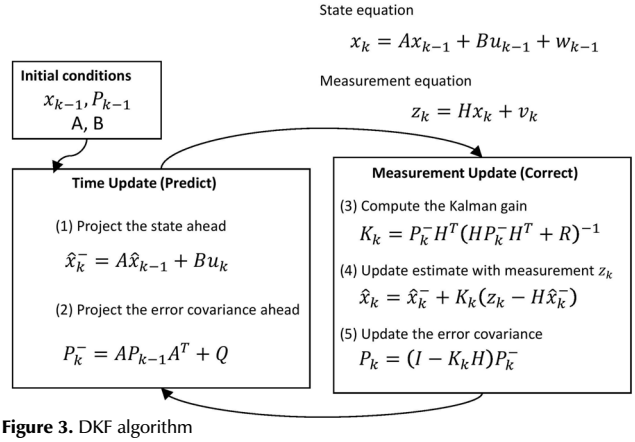

Figure 3. DKF algorithm
Source: Welch and Bishop (2006)

The state equation has two (n) hourly flow series from the Chapalagana and Platanitos stations. The matrices A (n x n) and B (n x 1) relate the state at time $k-1$ to the current state at time k. Like the EnKF, the control matrix B is not included, given that, in the 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 via 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 to be the 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.

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 and Rodríguez-Iturbe, 1985). Algebraically, it is described as follows:

$$y_{k+1} = \sum_{i=0}^{na} \alpha_i y_{k-i} + \sum_{j=0}^{nb} \beta_j \gamma_{k-j} + e_{k+1} \quad (3)$$

where y_{k+1} is the predicted value, α_i and β_j are the model parameters, and y_k and γ_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 and Lichtendahl, 2016). In the ARX(na, nb) model, na and nb represent the autoregressive delays that are used in each variable.

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 in order to estimate the error, which had 6, 10, 20, 30, 40, 50, and 100 members. Series with 1 360 registers were used, which included two main events.

As of 50 members per set, the EnKF algorithm showed a stable convergence for all steps. Consequently, under the conditions of this study, it is acceptable to use at least 50 members per set to have an adequate fit regarding the convergence of the algorithm and the stabilization of the error, aiming to minimize computational effort. As previously indicated, the results presented below have a base of 50 members per set. Table 1 presents the statistical indicators of fit for the observed series against the predicted one.

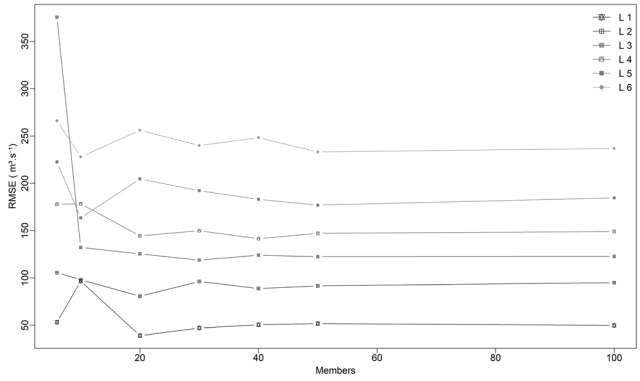

Figure 4. RMSE for different set sizes
Source: Authors

Table 1. Summary of statistics for application of EnKF, DKF, and ARX

Algorithm	Metric	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
	SD	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
	SD	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
	SD	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. The mean and standard deviation (SD) of the observed series are 198,7 and 239,77, respectively.

Source: Authors

According to the statistical indicators in Table 1, the NS shows similarities between DKF and ARX, with values of 0,99 and 0,83 in steps one and six, respectively, while EnKF obtained 0,95 and 0,05 in steps one and six. 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, whereas DKF has a downward trend.

Despite the low fit values, the EnKF algorithm expressed the changes with a non-linear trend and showed better convergence on the observed series once it was updated with new measurements. DKF and ARX generated forecasts in which the displacement of the predicted series was accentuated against the observed series, similar to the persistent forecast method (Aguado-Rodríguez *et al.*, 2016;

Kavasseri and Seetharaman, 2009). This behavior is also noticeable in the work of Alvarado-Hernández *et al*. (2020), González-Leiva *et al*. (2015), and Morales-Velázquez *et al*. (2014).

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

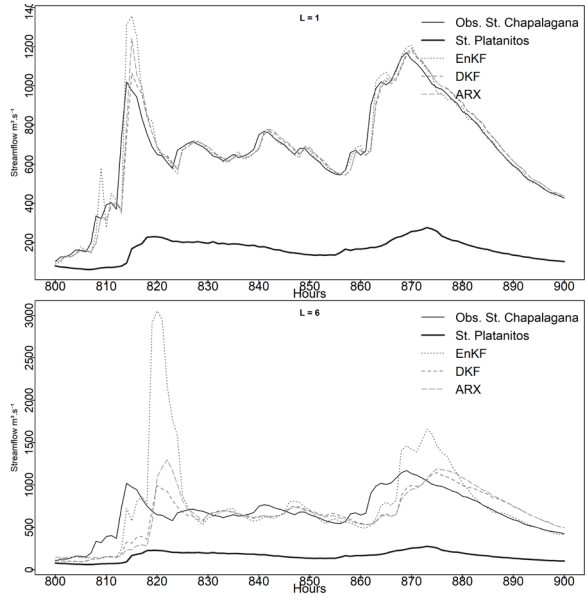

Figure 5. Observed flow and flows predicted with EnKF, DKF, and ARX (flood from 4/9/2017, 16:00 h, to 8/9/2017, 20:00 h)
Source: Authors

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 and DKF. This model considers the delay between series to be one, whereas, in our study, it was updated dynamically throughout the series, favoring the fit at peaks and reducing the effect of displacement of the predicted series. EnKF, in its six steps, obtained lower fits due to overestimation or underestimation at the peaks, with the difference that it achieved a better fit in ascents and descents of the observed series. The EnKF algorithm obtained a better temporal fit at peaks and converged more precisely on the observed series when the trend persisted in a number of hours higher or equal to the evaluated step.

The forecast with EnKF showed overestimations relative to the observed series. This occurred because we treated the measurements as a non-linear phenomenon (Evensen, 1994, 2009), a situation that, in step one, allows for an acceptable fit in the entire series. However, in step six, broad overestimations may be found which can affect the quality of the forecast. As the forecast step increases, the frequency of overestimations increases because the new register that serves to update the states also breaks away, and there may be changes in the L interval that are not considered in the initial forecast. The dynamic incorporation of the delay time between series (Meng *et al*., 2017) allows improving the fit, given that updating is performed with the equivalent event in the exogenous variable (Figure 6).

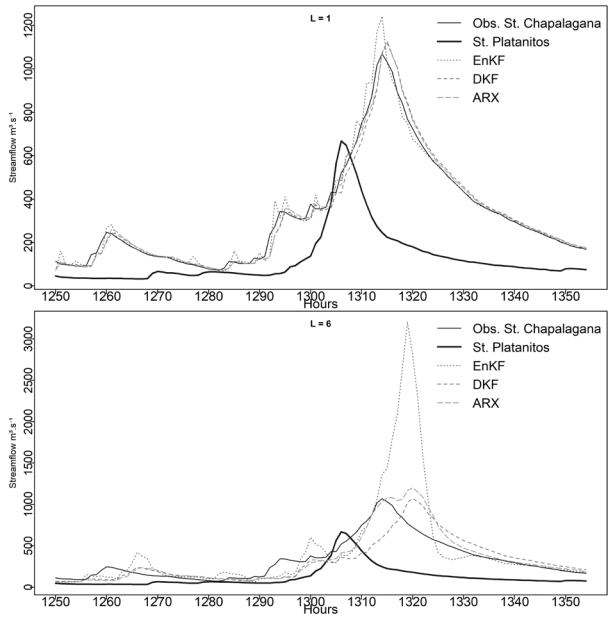

Figure 6. Observed flow and flows predicted with EnKF, DKF and ARX (flood from 23/09/17, 10:00 h, to 28/09/17, 00:00 h)
Source: Authors

The dispersion of the observed series against the predicted one was 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 tended to overestimate throughout the series, unlike DFK and ARX, which caused a slight tendency to underestimate. Step six with EnKF produced large overestimations that are reflected as isolated points above the diagonal in Figure 7.

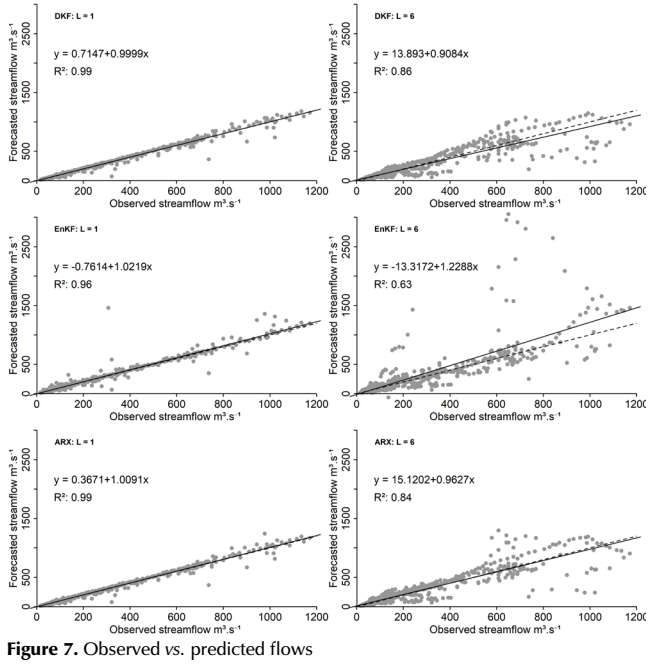

Figure 7. Observed *vs*. predicted flows
Source: Authors

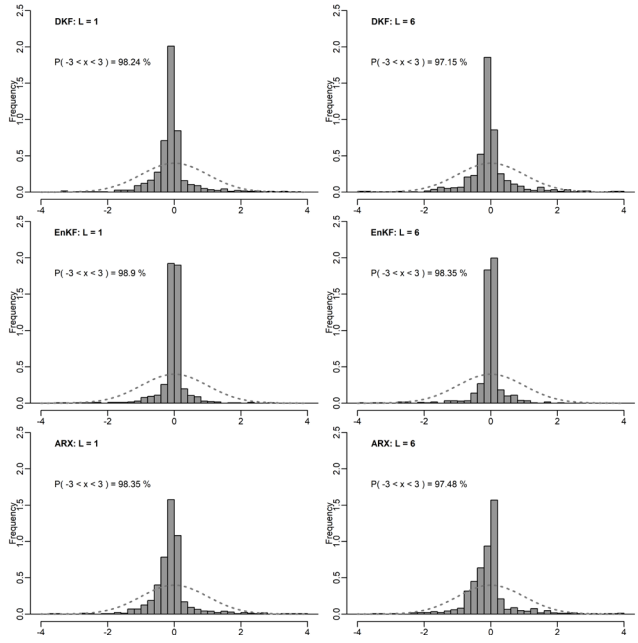

Figure 8. Histographs of residuals
Source: Authors

The standardized errors of the forecasts have a symmetrical distribution around zero (Figure 8) (Cryer and Chan, 2008; Martínez *et al*., 2012; Wang *et al*., 2017). According to Chatfield (2001), a 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 steps.

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

Conclusions

The dynamic updating of delay time relative to the exogenous variable allowed improving the fit in the evaluated algorithms. The EnKF algorithm achieved a better convergence on the observed series but generated overestimations of greater magnitude as the step increased, which resulted in a lower degree of fit, as demonstrated with the Nash-Sutcliffe index. The potential of EnKF lies in its convergence and non-linear treatment of abrupt changes in flow. Basically, EnKF helps capture the non-linearity in some parts of the of the hydrograph and accurately represents the timing or times of occurrence of maximum flows, even though they are overestimated.

In future studies, the use of EnKF for streamflow forecasting can be a viable alternative when integrated with autocorrelation analysis, so that stationality and stationarity become part of the model, thus allowing to represent the changes associated with daytime, nighttime, or the months of the year. Together, the quantity of registers for updating the states can be increased in order to enable the detection of changes in trends 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 (*e.g.*, 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 six steps equals 36 hours.

Acknowledgements

We would like thank the Federal Electricity Commission for allowing access to the hydrometeorological database, through the web site administered by the National Institute for Electricity and Clean Energies during 2018.

References

Abaza, M., Anctil, F., Fortin, V., and 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., and Zamora-Morales, P. (2016). Predicción de variables meteorológicas por medio de modelos ARIMA. *Agrociencia*, *50*(1). http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-31952016000100001

Alvarado-Hernández, L., Ibáñez-Castillo, L., Ruiz-García, A., González-Leiva, F., and Vázquez-Peña, M. (2020). Pronóstico horario de caudales mediante filtro de kalman discreto en el Río Huaynamota, Nayarit, México. *Agrociencia*, *54*, 295-312. https://doi.org/10.47163/agrociencia.v54i3.1907

Bai, L., Chen, Z., Xu, J., and 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., and Ljung, G. (2016). *Time series analysis: Forecasting and control* (5th ed.). Wiley.

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

Bras, R., and 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. 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., and 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., and Chan, K.-S. (2008). *Time series analysis with applications in R* (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-75959-3

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. https://doi.org/10.1007/978-3-642-03711-5

Field, C. B., Barros, V., Stocker, T. F., and Dahe, Q. (Eds.) (2012). *Managing the risks of extreme events and disasters to advance climate change adaptation*. Cambridge University Press. https://doi.org/10.1017/CBO9781139177245

Genz, A., and Bretz, F. (2009). *Computation of multivariate normal and t probabilities*. Springer. https://doi.org/10.1007/978-3-642-01689-9

Gillijns, S., Mendoza, O. B., Chandrasekar, J., De Moor, B. L. R., Bernstein, D. S., and Ridley, A. (2006, June 14-16). *What is the ensemble Kalman filter and how well does it work?* [Conference presentation]. 2006 American Control Conference, Minneapolis, MN, USA. https://doi.org/10.1109/ACC.2006.1657419

González-Leiva, F., Ibáñez-Castillo, L. A., Valdés, J. B., Vázquez-Peña, M. A., and 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. http://revistatyca.org.mx/ojs/index.php/tyca/article/view/1176

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

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. https://doi.org/10.1115/1.3662552

Kavasseri, R. G., and 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., and Gupta, H. V. (2007). Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resources Research*, *43*(7), W07401. https://doi.org/10.1029/2006WR005756

Martínez, J., Domínguez, E., and Rivera, H. (2012). Uncertainty regarding instantaneous discharge obtained from stage-discharge rating curves built with low density discharge measurements. *Ingenieria e Investigación*, *32*(1), 30-35. https://revistas.unal.edu.co/index.php/ingeinv/article/view/28517

Maxwell, D. H., Jackson, B. M., and 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., and 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. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2071-00542015000200005

Meng, S., Xie, X., and 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., and Valdés, J. B. (2014). Pronóstico de avenidas utilizando el filtro de Kalman discreto. *Tecnologia y Ciencias Del Agua*, *5*(2), 85-110. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2007-24222014000200006

Nash, J. E., and 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., and de Paiva, R. C. D. (2019). Data assimilation using the ensemble Kalman filter in a distributed hydrological model on the Tocantins River, Brasil. *RBRH*, *24*, e14. https://doi.org/10.1590/2318-0331.241920180031

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/

Rafieeinasab, A., Seo, D. J., Lee, H., and 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., and Lichtendahl, K. (2016). *Practical time series forecasting with R: A hands on guide*. Axelrod Schnall Publishers.

Simon, D. (2001, June). Kalman filtering. *Embedded Systems Programming*, 72-79. https://abel.math.harvard.edu/archive/116_fall_03/handouts/kalman.pdf

Singh, A., and Zommers, Z. (Eds.). (2014). *Reducing disaster: Early warning systems for climate change*. Springer. https://doi.org/10.1007/978-94-017-8598-3

Servicio Metereológico Nacional (SMN) (2019). *Sistema de información climática computarizada CLICOM*. Servicio Metereológico Nacional. http://clicom-mex.cicese.mx/malla/index.php

Sun, L., Seidou, O., Nistor, I., and 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., and Rodríguez-Iturbe, I. (1980). *Filtros de Kalman en la hidrología: predicción de descargas fluviales para la operación óptima de embalses. Informe Técnico No. 80-2*. https://n9.cl/hd6y9

Wang, S., Huang, G. H., Baetz, B. W., Cai, X. M., Ancell, B. C., and 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., and Bishop, G. (2006). *An introduction to the Kalman filter*. https://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf

Winkler, R. D., Moore, R. D. D., Redding, T. E., Spittlehouse, D. L., Carlyle-Moses, D. E., and Smerdon, B. D. (2010). Hydrologic processes and watershed response. In R. Pike, T. Redding, R. Moore, R. Winkler, and K. Bladon (Eds.), *Compendium of forest hydrology and geomorphology in British Columbia* (pp. 133-178). B.C. Ministry of Forests and Range. https://www.for.gov.bc.ca/hfd/pubs/docs/lmh/Lmh66/Lmh66_ch06.pdf

Xu, J., Li, W., Ji, M., Lu, F., and Dong, S. (2009). A comprehensive approach to characterization of the nonlinearity of runoff in the headwaters of the Tarim River, western China. *Hydrological Processes*, *24*(2), 136-146. https://doi.org/10.1002/hyp.7484

Zou, L., Zhan, C., Xia, J., Wang, T., and 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

Ildefonso Narváez-Ortiz, Laura Ibánez-Castillo,
Ramon Arteaga-Ramírez, Mario Vázquez-Pena

Ensemble Kalman Filter for Hourly Streamflow Forecasting in Huaynamota River, Nayarit, México
Filtro de Kalman de Conjuntos para pronóstico de caudales horarios en el río Huaynamota, Nayarit, México

Ingeniería e Investigación
vol. 42, no. 3, e208, 2022
Facultad de Ingeniería, Universidad Nacional de Colombia.,
ISSN: 0120-5609
ISSN-E: 2248-8723

DOI: https://doi.org/10.15446/ing.investig.90023