



MARN

Ministerio de Medio Ambiente  
y Recursos Naturales

---

# Aplicación de Redes Neuronales Artificiales en Pronóstico hidrológico. Caso de estudio: Cuenca Rio Grande de San Miguel

José Rodolfo Valles León

Dirección Observatorio Ambiental, Gerencia de Hidrología

---

## RESUMEN

El presente estudio pretende generar modelos de pronóstico a corto plazo ocupando arquitectura de Redes Neuronales Artificiales, en la Cuenca hidrográfica del Rio Grande de San Miguel. Para completar el objetivo de dicho estudio, se recopilaron datos hidrometeorológicos de las estaciones telemétricas que se encuentra dentro y/o los alrededores de la cuenca. Posteriormente, estos datos fueron desagregados a valores horarios, debido a que el tiempo de concentración de la cuenca es menor a 1 día. Consecutivamente, se realizó un análisis de interdependencia entre variables, con el fin de determinar las variables de entrada de los distintos modelos de predicción. Finalmente, los resultados fueron evaluados ocupando funciones de error y comparados con un modelo de referencia "Naive", con el fin de evaluar los resultados de los mismos.

Los resultados de este estudio muestran un excelente resultado para pronósticos hidrológicos no mayores a 4 horas debido a que funciones de eficiencia tales como el Coeficiente Nash-Sutcliffe (NSE) brinda valores mayores a 80%. Asimismo, valores de pronóstico de hasta 9 horas de tiempo de anticipo brinda resultados buenos debido a que el valor de NRMSE es de 0.5. Por último, el pronóstico de 12 horas de tiempo de ventaja brinda resultados satisfactorios, ya que su desempeño en reproducir el caudal observado es cercano al 60%. Además, los modelos de predicción hidrológica a corto plazo muestran mejor reproducción comparada con un modelo de referencia Naive, lo cual indica que la técnica sofisticada de Redes Neuronales Artificiales se está beneficiando tanto de la precipitación en la cuenca, como en el caudal pasado para la predicción de caudales.

**Palabras claves:** Pronostico, Redes Neuronales Artificiales, tiempo anticipo, tiempo retraso, sigmoidea,

---

## INTRODUCCION

Redes Neuronales Artificiales son una de las herramientas de modelación hidrológica, las cuales han ayudado a resolver problemas de pronóstico a corto plazo en cuencas hidrográficas. Su aplicabilidad ha sido probada a nivel mundial, en donde centro de pronósticos hidrológico recurrentes a esta herramienta. Sin embargo, estos modelos no son muy buenos en pronósticos de mediano y largo plazo, por lo que se debe recurrir a modelos físicos o conceptuales para resolver el problema de representación de los caudales en la cuenca hidrográfica.

El objetivo de este estudio es de desarrollar un modelo de pronóstico hidrológico a corto plazo en una de las cuencas importantes de El Salvador. Adicionalmente, el tiempo de ventaja del pronóstico será determinado por las características físicas de la cuenca, tales como el tiempo de respuesta; la cual es el tiempo que transcurre entre la precipitación y el caudal máximo en un punto de control. Asimismo, la herramienta de modelación escogida en este estudio son las Redes Neuronales Artificiales (ANN) con funciones de activación en su capa escondida y funciones lineales en su capa de salida.

## CASO DE ESTUDIO

La cuenca del Rio Grande de San Miguel es una de las mayores cuencas hidrográficas del país. El área superficial es de 1,616 km<sup>2</sup> hasta el punto de control ubicado en el cantón La Canoa, (Ver Figura 1).

La información hidrometeorológica en la cuenca muestra información de precipitación en las estaciones telemétrica, las cuales, tienen un tiempo de transmisión de 10 o 15 minutos, dependiendo de la configuración. Asimismo, se tiene información histórica de niveles y aforo en las estaciones Villerías y La Canoa. Para este estudio, se pretende generar modelos de pronóstico hidrológico en la estación La Canoa, hasta un tiempo máximo de anticipo de 12 o 15 horas. Para lo anterior, se pretende ocupar información de lluvia promedio areal en la cuenca de las estaciones pluviométricas, así como de información de caudales pasados en la estación La Canoa.

La información de precipitación horaria de las estaciones telemétricas fue procesada en lluvia areal promedio mediante polígonos de Thiessen. Adicionalmente, los valores de evapotranspiración potencial fueron obtenidos mediante

relaciones de elevación y evapotranspiración brindadas por SNET, 2005

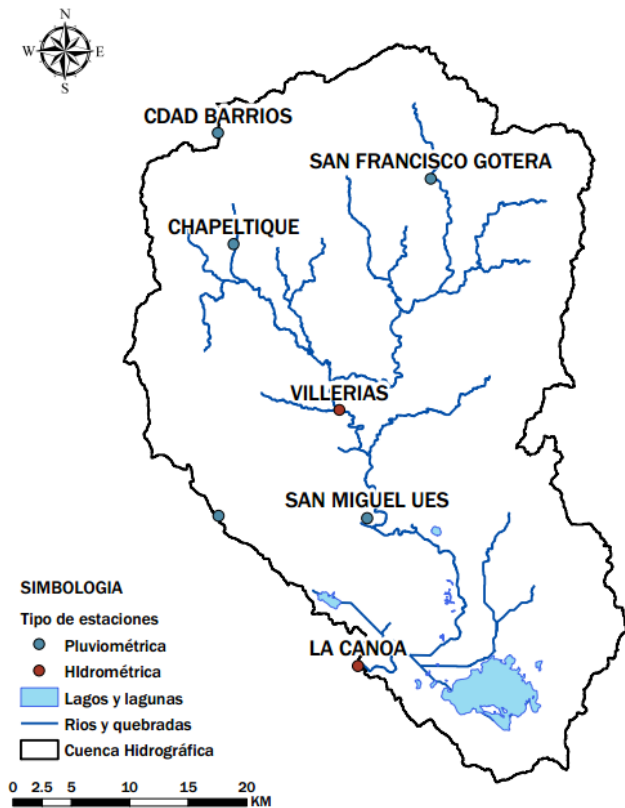


Figura 1: Cuenca hidrográfica del Río Grande de San Miguel y estaciones hidrometeorológicas con transmisión telemétrica.

## METODOLOGIA

La primera fase del estudio está basado en el análisis de los datos de lluvia, nivel y caudal disponible en el área de estudio. En esta fase, se eliminan los valores anormales los cuales pueden generar ruido en los modelos hidrológicos. Posteriormente se desagregaron los valores de nivel y lluvia en valores horarios, ya que el tiempo de respuesta de la cuenca es de horas. Finalmente, se convirtieron los valores de nivel en formato de caudal usando curvas de descarga.

La segunda fase contempla la definición del horizonte de pronóstico hidrológico y la preparación de los datos disponible. En esta fase se define el tiempo máximo con el cual se puede realizar pronóstico con lluvia y caudal observado. Posteriormente, los datos disponibles se separaron en tres tipos de datos, con el fin de entrenar y verificar el desempeño del modelo, el cual es una arquitectura de Red Neuronal Artificial con función de activación sigmoidea en la capa escondida y una función lineal en la capa de salida. Asimismo, estos modelos fueron comparados con un modelo de referencia "Naive", el cual permite evaluar técnicas sofisticadas de modelación y cuantificar cuanto es el beneficio que está aportando la precipitación en el cálculo del caudal.

La tercera fase contempla el análisis de resultados de los diversos modelos generados, para verificar que los datos obtenidos se mantienen dentro de un rango aceptable de error. Para este análisis, se ocuparon diversas funciones de

errores, así como inspección visual entre los datos observados y simulados.

Entre las funciones de error evaluados en este documento se puede mencionar:

### Error Medio (ME)

Esta función de error calcula la diferencia media entre el caudal observado y el simulado.

$$ME = \frac{\sum(Q_{sim} - Q_{obs})}{n} \quad (1)$$

### Coefficiente de Eficiencia Nash-Sutcliffe (NSE)

Propuesto por Nash & Sutcliffe (1970), el cual varía entre  $-\infty$  hasta 1, indicando una reproducción perfecta entre el caudal observado y simulado cuando el NSE es igual a un valor de 1. Además, la literatura nos detalla que un valor mayor a 0.6 de coeficiente de eficiencia, brinda un resultado satisfactorio del desempeño del modelo. De igual forma, si el valor del NSE es menor o igual a 0, entonces el modelo no es mejor que usar el caudal observado promedio como predictor. (Gupta, Sorooshian, & Yapo, 1998)

$$NSE = 1 - \frac{\sum(Q_{sim} - Q_{obs})^2}{\sum(Q_{obs} - \bar{Q}_{obs})^2} \quad (2)$$

### Coefficiente de determinación ( $R^2$ )

Esta función permite evaluar la bondad de la predicción mediante la proporción de la variación del caudal simulado y el caudal observado. Si la función es igual a 0, significa que existe nula capacidad de predicción. Mientras que si la función es igual a 1 significa que la predicción no tiene error alguno.

$$R^2 = 1 - \frac{SSE}{SST} \quad (3)$$

El coeficiente de determinación se calcula mediante la fórmula detallada en ecuación 3, en donde las variables SSE y SST indican la suma cuadrática del error y de la regresión, respectivamente.

### Raíz del error cuadrático medio (RMSE)

Esta función de error es ampliamente ocupada para verificar el error en los modelos hidrológicos, debido a que brinda una máxima estimación de verosimilitud (MLE, por sus siglas en ingles).

$$RMSE = \sqrt{\frac{\sum(Q_{sim} - Q_{obs})^2}{n}} \quad (4)$$

### Valor Normalizado del RMSE (NRMSE)

Aunque es ampliamente conocido que un valor bajo de RMSE indica un mejor desempeño del modelo, solamente Singh et al, (2004) ha publicado unas guías para cuantificar que es un

valor bajo de RMSE, el cual está basado en la desviación estándar del caudal observado ( $\sigma_{obs}$ )

$$NRMSE = \frac{RMSE}{\sigma_{obs}} \quad (5)$$

Coefficiente de eficiencia Kling-Gupta (KGE)

$$KGE = 1 - \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (6)$$

El coeficiente de eficiencia Kling-Gupta, el cual ha sido propuesto por Gupta et al. (1998), el cual relaciona las variables  $r$  del coeficiente de correlación de Pearson,  $\alpha$  es la relación entre la varianza y  $\beta$  es la relación los caudales simulados y observados promedios. Un valor de KGE igual a 1 indica que existe una excelente representación entre el caudal observado y simulado.

### DETERMINACION DEL HORIZONTE DEL PRONÓSTICO

Los modelos basados en datos o “Data-Driven Models” como son conocidos en inglés, demuestran una reducción en su eficiencia conforme se aumentan el tiempo de ventaja del pronóstico, sin embargo se muestra menos incertidumbres en pronóstico a corto plazo debido a que brinda muy buenos resultados. Figura 2 muestra el valor del pronóstico hidrológico conforme se aumenta el tiempo de ventaja del mismo. En dicha figura se puede observar que el valor de los pronósticos a corto plazo es menor comparado con los pronósticos a largo plazo, debido a que mayor tiempo de anticipo permite tomar medidas de mitigación con mucha eficiencia. No obstante, la precisión del pronóstico es muy baja comparada con un pronóstico a corto plazo. Por lo tanto, un mayor tiempo de ventaja, brinda mayores incertidumbres.

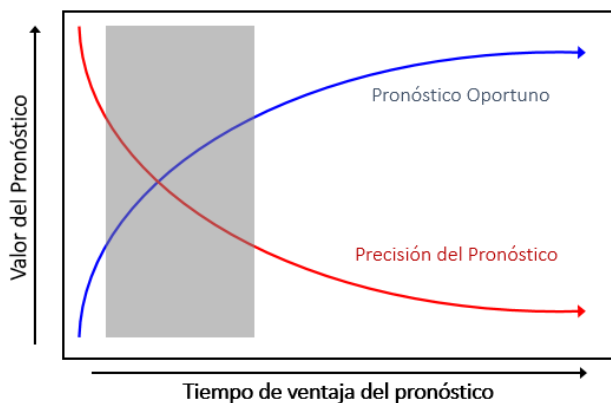


Figura 2: valor del pronóstico para diferentes tiempo de ventaja en modelos hidrológicos. Adaptación de Werner et al. (2005)

Para el caso de estudio del Rio Grande de San Miguel, se revisaron las características físicas de la cuenca tales como el tiempo de respuesta de la cuenca. Como resultado, se puede observar en la Figura 3 que en promedio, el tiempo de respuesta de la cuenca es alrededor de 15 horas. Por supuesto que este tiempo puede ser menor, debido a la distribución espacial de la tormenta, pero esta información sirve para delimitar el tiempo de ventaja del pronóstico hidrológico con lluvia y caudal observado.

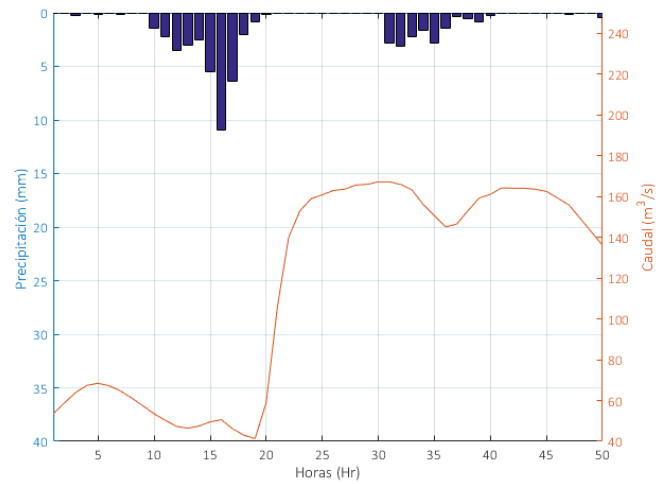


Figura 3: precipitación (barras azules) y caudal (línea roja) en la cuenca del Rio Grande de San Miguel, en donde se observa un tiempo de respuesta promedio de 15 horas.

### ANALISIS DE INTERDEPENDENCIA ENTRE VARIABLES

Entre las variables hidrometeorológicas que más influyen para pronóstico hidrológico se pueden mencionar la precipitación, evapotranspiración, temperatura y el caudal. Sin embargo, es importante saber que no todas estas variables influyen de manera sustancial en el modelo de predicción de caudales. Por lo mencionado anteriormente, es importante realizar un análisis de interdependencia entre variables para determinar cuáles variables afectan más en el caudal pronosticado. Para lo anterior, se ocupan herramientas tales como el Coeficiente de Correlación Lineal (CoC), el cual permite evaluar la relación lineal entre una variable dependiente y una independiente. Otra de las herramientas ocupadas es la Información Mutua Promedio AMI (Shannon, 1948), la cual permite evaluar la cantidad de información que es posible aprender de una variable con respecto a otra variable.

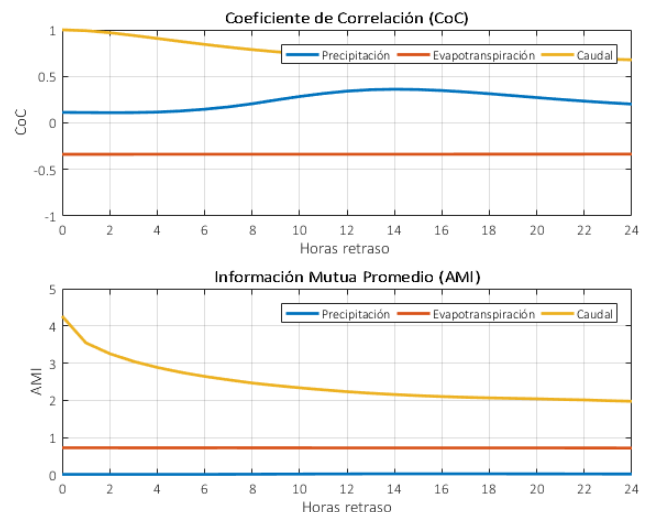


Figura 4: Coeficiente de Correlación (CoC) e Información Mutua Promedio (AMI) entre el caudal y las variables retrasadas de precipitación, evapotranspiración y el mismo caudal

Figura 4 muestra los Coeficiente de Correlación e Información Mutua Promedio (AMI) de las variables de entrada tales como Precipitación, evapotranspiración y

caudal. Se puede observar que las variables pasadas de caudal brinda los mayores valores de AMI y correlación, Asimismo, la precipitación que más afecta en el cálculo de caudal es la que se produce entre 10 y 14 horas de tiempo de retraso, no obstante debe mencionarse que su bajo valor de AMI y CoC podría deberse a los valores de 0 en precipitación, ya que físicamente la lluvia tiene gran influencia en el cálculo de caudal. Por último, no se observa gran aporte de la evapotranspiración en el cálculo del caudal, lo anterior podría deberse al método de cálculo del mismo, por lo que se podría explorar otro método de cálculo.

En base a los resultados anteriormente, es acertado pensar que la precipitación y el caudal, son las variables que más información brinda en los modelos de pronósticos y, por ende, deberían ser incluidos como datos de entrada.

Un similar análisis se realizó para los diferentes tiempos de ventajas evaluadas en este estudio, estos valores se muestran en Figura 5, en donde se muestra el Coeficiente de Correlación entre la precipitación y diferentes tiempos de ventaja y retraso. Se puede observar que aumentando las horas de tiempo de ventaja, el valor máximo de correlación tiene menos tiempo de retraso. Lo anterior nos indica que con mayor tiempo de ventaja o anticipo, tendremos menor tiempo de retraso en el valor óptimo de coeficiente de correlación, lo cual justifica el comportamiento físico del tiempo de respuesta de la cuenca. Adicionalmente, no se observa una disminución considerable en el valor máximo del coeficiente de correlación.

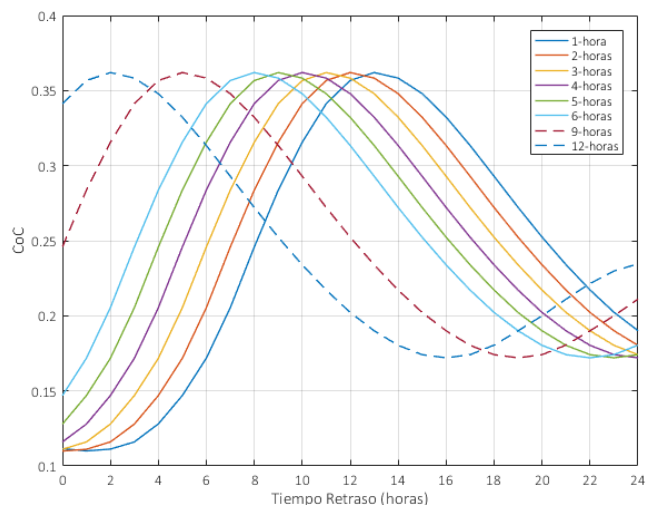


Figura 5: Coeficiente de Correlación para cada tiempo de retraso y tiempo de anticipo en la cuenca del Río Grande de San Miguel.

En base a lo discutido anteriormente, para cada uno de los modelos de pronóstico hidrológico, se definieron las variables de entrada, los cuales son mostrados en Tabla 1. Estas variables son las que transfieren la mayor cantidad de información para el cálculo del caudal para cada uno de los modelos de predicción hidrológica en la cuenca hidrográfica del Río Grande de San Miguel.

ID	Variable Entrada	Variable Salida
1	$RF_{t-14}, RF_{t-13}, RF_{t-12}, RF_{t-11}, RF_{t-10}, Q_{t-2}, Q_{t-1}, Q_t$	$Q_{t+1}$
2	$RF_{t-13}, RF_{t-12}, RF_{t-11}, RF_{t-10}, RF_{t-9}, Q_{t-2}, Q_{t-1}, Q_t$	$Q_{t+2}$
3	$RF_{t-12}, RF_{t-11}, RF_{t-10}, RF_{t-9}, RF_{t-8}, Q_{t-2}, Q_{t-1}, Q_t$	$Q_{t+3}$
4	$RF_{t-11}, RF_{t-10}, RF_{t-9}, RF_{t-8}, RF_{t-7}, Q_{t-2}, Q_{t-1}, Q_t$	$Q_{t+4}$
5	$RF_{t-11}, RF_{t-10}, RF_{t-9}, RF_{t-8}, RF_{t-7}, Q_{t-2}, Q_{t-1}, Q_t$	$Q_{t+5}$
6	$RF_{t-10}, RF_{t-9}, RF_{t-8}, RF_{t-7}, RF_{t-6}, Q_{t-2}, Q_{t-1}, Q_t$	$Q_{t+6}$
7	$RF_{t-6}, RF_{t-5}, RF_{t-4}, RF_{t-3}, RF_{t-2}, Q_{t-2}, Q_{t-1}, Q_t$	$Q_{t+9}$
8	$RF_{t-3}, RF_{t-2}, RF_{t-1}, RF_t, Q_{t-2}, Q_{t-1}, Q_t$	$Q_{t+12}$

Tabla 1: Variables de entrada para cada modelo de pronóstico hidrológico

## PREPARACIÓN DE LOS DATOS

Los datos horarios de precipitación y caudal corresponden a los años 2013, 2014, 2015 y 2016. Estos datos fueron separados en tres tipos de datos, los cuales se detallan a continuación

- **Datos de entrenamiento**, este tipo de datos es el que mayor porcentaje de dato posee y con el cual se encuentra el número de neuronas (parámetros) del modelo, mediante un proceso de optimización.
- **Datos de validación cruzada**, el cual es una serie de datos que permite evaluar los parámetros del modelo y evitar un sobreajuste o sobreentrenamiento de la Red Neuronal Artificial.
- **Datos de verificación**, el cual permite evaluar el desempeño y eficiencia del modelo con una serie de dato que estuvo involucrada en la calibración del modelo y en la validación cruzada.

Las propiedades estadísticas de caudales en cada serie de datos son mostradas en la Tabla 2, cabe mencionar que los caudales máximo y mínimo fueron ubicados dentro de la serie de entrenamiento.

Propiedad Estadística	Entrenamiento	Validación Cruzada	Verificación
Máximo	167.17	163.87	157.75
Mínimo	0.84	1.78	1.36
Promedio	14.45	14.12	12.64
Desviación Estándar	21.05	23.11	19.15

Tabla 2: Propiedad Estadística de cada serie de dato

## RESULTADOS Y DISCUSION

Se realizaron 8 tipos de modelos hidrológicos para 1, 2, 3, 4, 5, 6, 9 y 12 horas de tiempo de ventaja y usando como arquitectura, Redes Neuronales Artificiales con función de activación sigmoidea en la capa escondida y función lineal en la capa de salida del modelo. Los resultados de estos modelos fueron comparados con el modelo de referencia "Naive" mediante inspección visual y cálculo de funciones de error, detallados anteriormente.

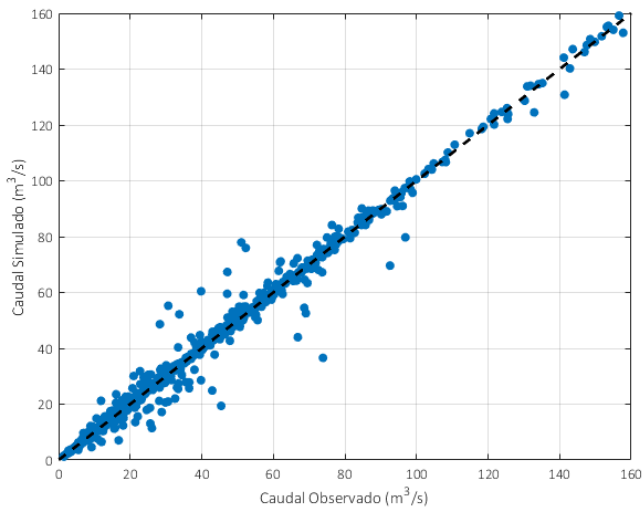


Figura 6: gráfico de dispersión entre el caudal observado y el simulado para un modelo de pronóstico de 1 hora de tiempo de ventaja

Los resultados para el cálculo de caudal con 1 hora de tiempo de anticipación (Ver Figura 6) muestran una excelente representación de los caudales altos y bajo por el modelo de Redes Neuronales Artificiales. Asimismo, los valores mostrados en la Tabla 3, nos muestra que tiene un mejor resultado que el modelo de referencia Naive, por lo que la técnica de Redes Neuronales Artificiales está brindando una mejoría que un modelo sin parámetros y autoregresivo.

Similar análisis se realizó para los modelos de pronóstico de 3, 6, 9 y 12 horas; en donde se puede observar en el diagrama de dispersión entre observados y simulado (Figura 7) que la eficiencia de los modelo disminuye conforme se aumenta el tiempo de ventaja. Este resultado está conforme a la gráfica presentada en Figura 2 en donde se describe que un pronóstico de 12 horas tiene un gran valor, pero es la precisión y eficiencia del modelo es mucho menor comparado a un modelo de 1 hora de tiempo de ventaja

Los valores de cálculo de error para el pronóstico de 12 horas detallan un satisfactorio resultado entre el caudal simulado y el observado; no obstante, existen algunos eventos de caudal máximo que no son bien reproducido por el modelo. Esta pobre representación de caudales altos podría deberse a la distribución espacial de la precipitación, ya que este tipo de modelo agrupado en una sola cuenca (Lumped) no considera dicha distribución. Se podría explorar un modelo de Redes Neuronales semi-distribuido, en donde cada variable de entrada del modelo es un dato horario de una estación pluviométrica.

Otra de las fuentes de incertidumbre que podrían estar jugando un papel importante en el cálculo del pronóstico de 12 horas, podría estar relacionado a la curva de descarga. Debido a que no se ha desarrollado algún método eficiente de extrapolación de caudales altos, así como también a incertidumbre aleatoria en la captura de datos, la cual es de forma irreversible.

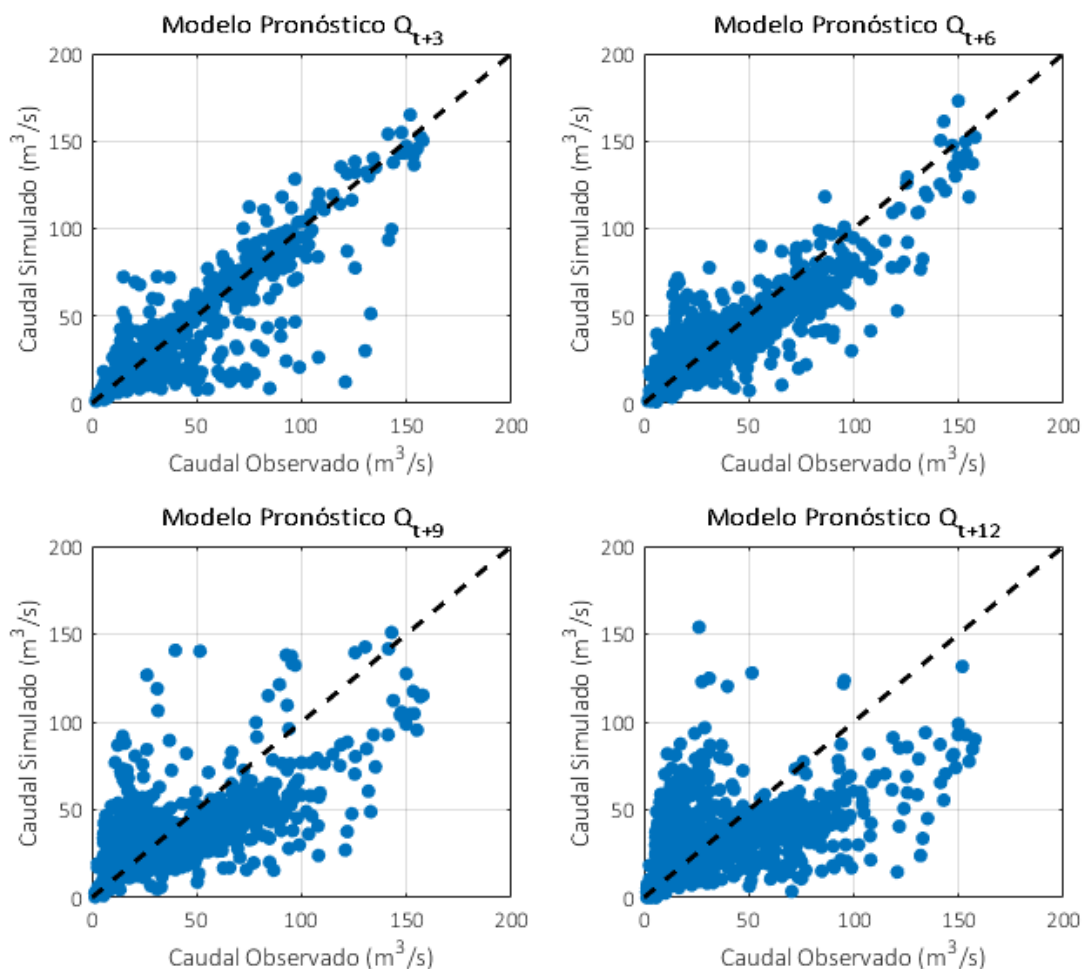


Figura 7: diagrama de dispersión entre caudal observado y simulado para pronóstico hidrológico de 3, 6, 9 y 12 horas

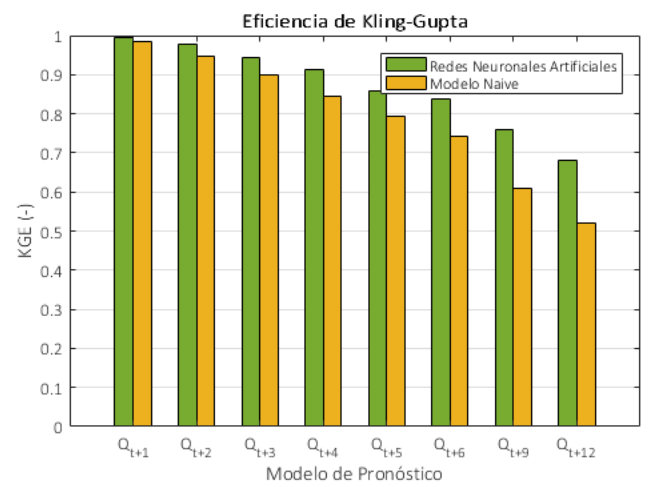
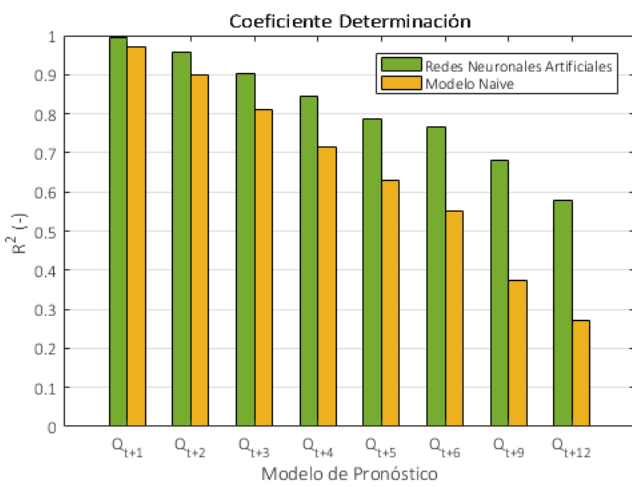
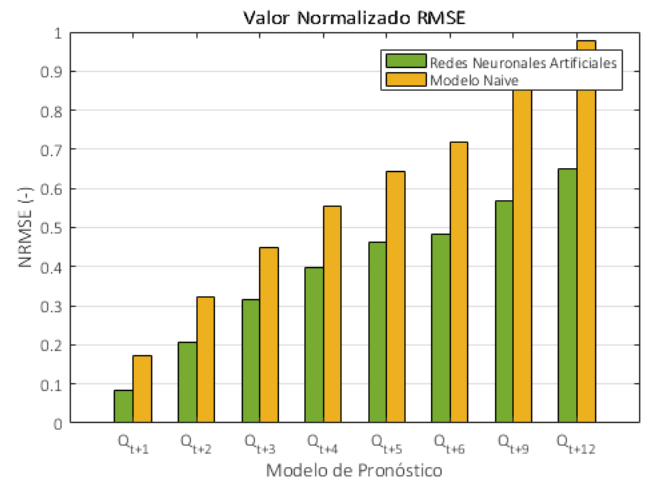
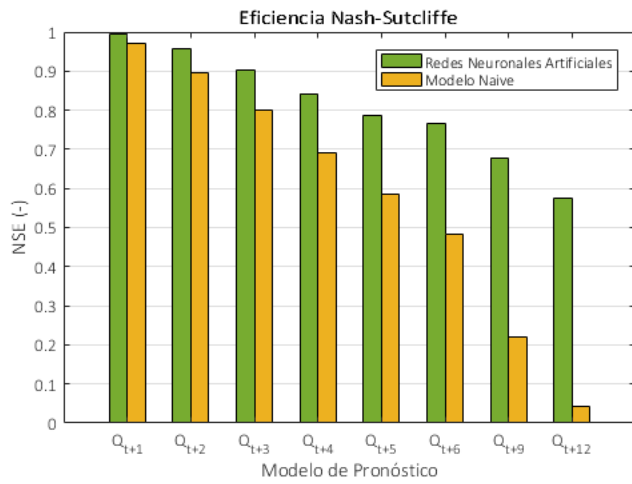


Figura 8: Desempeño y eficiencias de modelos hidrológicos de pronóstico para 1, 2, 3, 4, 5, 6, 9 y 12 horas de tiempo de anticipo

Modelos	ME (m <sup>3</sup> /s)	NSE (-)	RMSE (m <sup>3</sup> /s)	NRMSE (-)	R <sup>2</sup> (-)	KGE (-)	
Q <sub>t+1</sub>	Naive	0.001	0.971	3.282	0.171	0.971	0.985
	ANN	0.033	0.993	1.578	0.082	0.993	0.996
Q <sub>t+2</sub>	Naive	0.001	0.896	6.164	0.322	0.899	0.948
	ANN	0.071	0.957	3.957	0.207	0.958	0.977
Q <sub>t+3</sub>	Naive	0.002	0.799	8.588	0.448	0.809	0.899
	ANN	-0.015	0.901	6.021	0.314	0.902	0.943
Q <sub>t+4</sub>	Naive	0.003	0.693	10.616	0.554	0.716	0.846
	ANN	0.056	0.843	7.596	0.397	0.845	0.914
Q <sub>t+5</sub>	Naive	0.003	0.586	12.318	0.643	0.629	0.793
	ANN	-0.154	0.787	8.382	0.461	0.788	0.860
Q <sub>t+6</sub>	Naive	0.004	0.484	13.757	0.718	0.551	0.742
	ANN	-0.137	0.766	9.263	0.484	0.766	0.836
Q <sub>t+9</sub>	Naive	0.006	0.220	16.910	0.883	0.372	0.610
	ANN	-0.482	0.679	10.855	0.567	0.680	0.761
Q <sub>t+12</sub>	Naive	0.008	0.042	18.749	0.979	0.271	0.521
	ANN	-0.487	0.576	12.467	0.651	0.578	0.682

Tabla 3: Resumen de valores de error para cada modelo de pronóstico hidrológico

## CONCLUSIONES Y RECOMENDACIONES

Los objetivos de este estudio fueron construir modelos de pronóstico a corto plazo y con variables hidrometeorológicas observadas para diferente tiempo de ventaja. La bondad de esta cuenca hidrográfica de tener un tiempo de respuesta cercano a 15 horas, nos permitió elaborar un pronóstico hidrológico de 12 horas de tiempo de anticipación. Los resultados de todos los modelos fueron evaluados mediante fórmula de funciones de error tales como Coeficiente de Nash-Sutcliffe, Kling-Gupta, Raíz del error cuadrado medio (RMSE), coeficiente de determinación, entre otros.

Los resultados muestran un excelente resultado para pronósticos hidrológicos no mayores a 4 horas debido a que funciones de eficiencia tales como el Coeficiente Nash-Sutcliffe brinda valores mayores a 80%. Asimismo, valores de pronóstico de hasta 9 horas de tiempo de anticipo brinda resultados buenos debido a que el valor de NRMSE es de 0.5. Por último, el pronóstico de 12 horas de tiempo de ventaja brinda resultados satisfactorios, ya que su desempeño en reproducir el caudal observado es cercano al 60%.

Además, los modelos de predicción hidrológica a corto plazo muestran mejor reproducción comparada con un modelo de referencia Naive, lo cual indica que la técnica sofisticada de Redes Neuronales Artificiales se está beneficiando tanto de la precipitación en la cuenca, como en el caudal pasado para la predicción de caudales.

Las incertidumbres de estos modelos provienen principalmente de la distribución espacial de la precipitación, la cual no está contemplada en la arquitectura de Redes Neuronales Artificiales. Esta cuenca hidrográfica muestran gran variabilidad en sus regiones climática, la cuales responde de forma diferente conforme a la localización de tormentas de gran magnitud. Por lo anterior, sería interesante explorar un modelo de Redes Neuronales Artificiales de forma semi-distribuida, en donde cada variable de entrada es un dato de estación pluviométrica. Dicho resultado serían comparados con los brindados en este estudio, con el fin de verificar que si aumentando la

complejidad del modelo, se reduce la incertidumbre del mismo producto de la distribución espacial de la lluvia.

## REFERENCIAS

- Abebe, A., & Price, R. (2003). Managing uncertainty in hydrological models using complementary models.
- Frank, E., Chui, C., & Witten, I. (2000). Text categorization using compression models. .
- Gupta, H., Sorooshian, S., & Yapo, P. (1998). Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources*.
- Moriasi, D., Arnold, J., & Van Liew, M. (2007). Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations.
- Nash, J., & Sutcliffe, J. (1970). River flow forecasting through conceptual models part I — A discussion of principles.
- Quinlan, J. (1992). Learning with continuous classes.
- Shannon, C. (1948). A mathematical theory of communication, Part I, Part II. .
- Singh, J., Knapp, H., & Demissie, M. (s.f.). Hydrologic modeling of the Iroquois River watershed using HSPF and SWAT.
- SNET. (2005). Balance hídrico integrado y dinámico en El Salvador.
- Solomatine, D., & Xue, Y. (2004). M5 model trees and neural networks: application to flood forecasting in the upper reach of the Huai River in China.
- Werner, M., Reggiani, P., De Roo, A., Bates, P., & Sprokkereef, E. (2005). Flood forecasting and warning at the river basin and at the European scale.