



**UNIVERSIDAD
DEL AZUAY**

**ESTIMACIÓN DE CAUDALES EN BASE A TÉCNICAS DE
APRENDIZAJE AUTOMÁTICO: CASO PARA EL SITIO DE
CONSTRUCCIÓN DE LA PRESA QUINGOYACU EN LA
CUENCA DEL RÍO YANUNCAY**
Tesis de Maestría Hidrosanitaria

Director:
Dr. Aurelio Ochoa

Autor:
Sebastián Montesdeoca C.

2026

DEDICATORIA

A mis seres queridos.

AGRADECIMIENTOS

Mi agradecimiento a Dios por su infinito amor.

A mi familia y amigos por su apoyo incondicional.

A la EMPRESA ELECTRO GENERADORA DEL AUSTRO – ELECAUSTRO S.A. por su ayuda y contribución en mi realización personal y profesional.

A la Universidad del Azuay por su aporte a la colectividad y al Dr. Santiago Aurelio Ochoa García por sus enseñanzas.

Sebastián Montesdeoca Cabrera

RESUMEN

Muchas obras de infraestructura demandan conocer y modelar los caudales de ríos, los pronósticos basados en metodologías estadísticas comúnmente ofrecen probabilidades de excedencia o déficit de caudal, convirtiéndose en una limitación al momento de realizar la gestión del uso del agua. Lo que se busca es contar con una metodología alternativa para la estimación de caudales basados en técnicas de aprendizaje autónomo.

Las limitaciones en la disponibilidad de registros y datos para realizar el modelado de ríos son comunes, por esta razón, en este trabajo se emplean únicamente como fuente de entrada el caudal medio diario en el punto de interés, consistente en una serie de tiempo entre 1965 y 2025, correspondiente una cuenca andina de pequeño tamaño.

Con la finalidad de elegir el modelo de aprendizaje autónomo más adecuado para para la estimación de caudales se plantearon cuatro algoritmos diferentes. Los modelos empleados fueron árboles de decisión potenciado por gradiente (XGB), Random Forest o Bosque aleatorio, Regresión de Soporte Vectorial (SVR) y Redes Neuronales (LSTM). Los resultados obtenidos muestran que el modelo Random Forest presenta el mejor desempeño general de los cuatro métodos evaluados.

Se realizó la optimización del modelo Random Forest por medio de la búsqueda aleatoria de hiperparámetros, con el cual se realizó la estimación de caudales para 90, 180 y 365 días, obteniendo resultados en todos los casos un factor de correlación superior a 0.9 y un índice NSE superior a 0.8.

Se realizó un análisis de sensibilidad del modelo obteniendo que las variables que mayor importancia tienen son la ventana de días previos utilizados para el pronóstico de caudales (10 días) y el valor del caudal en los días inmediatamente anteriores al pronóstico.

Palabras clave: Aprendizaje autónomo, machine learnig, Random Forest, optimización, análisis de sensibilidad, estimación de caudales.

ABSTRACT

Many infrastructure projects require knowledge and modeling of river flows. Forecasts based on statistical methodologies commonly provide probabilities of flow exceedance or deficit, which becomes a limitation when managing water use. Therefore, there is a need for an alternative methodology for flow estimation based on autonomous learning techniques.

Limitations in the availability of records and data for river modeling are common. For this reason, this study uses only the daily mean streamflow at the point of interest as the input source, consisting of a time series from 1965 to 2025, corresponding to a small Andean watershed.

In order to select the most appropriate machine learning model for streamflow estimation, four different algorithms were tested. The models evaluated were Extreme Gradient Boosting decision trees (XGB), Random Forest, Support Vector Regression (SVR), and Long Short-Term Memory neural networks (LSTM). The results show that the Random Forest model presents the best overall performance among the four evaluated methods.

The Random Forest model was optimized through random search hyperparameter tuning, which was then used to estimate streamflow for 90, 180, and 365 days. The results showed, in all cases, a correlation coefficient greater than 0.9 and a Nash–Sutcliffe Efficiency (NSE) index greater than 0.8.

A sensitivity analysis of the model was also performed, showing that the most influential variables are the window of previous days used for streamflow forecasting (10 days) and the streamflow values immediately preceding the forecast.

Keywords: Machine learning, Random Forest, hyperparameter optimization, sensitivity analysis, streamflow estimation.