



UNIVERSIDAD
CATÓLICA
DE CUENCA

UNIVERSIDAD CATÓLICA DE CUENCA

Comunidad Educativa al Servicio del Pueblo

UNIDAD ACADÉMICA DE INGENIERÍA,

INDUSTRIA Y CONSTRUCCIÓN.

CARRERA DE INGENIERÍA CIVIL.

**EVALUACIÓN DE MODELOS DE PRONÓSTICO PARA ESTIMAR EL
INGRESO DE CAUDALES AL EMBALSE MAZAR: APLICACIÓN CON REDES
NEURONALES.**

**TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO
CIVIL**

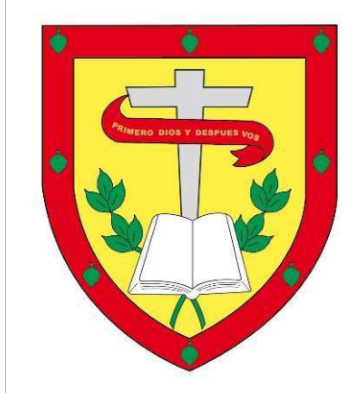
AUTOR: JUDER ANTONIO CASTILLO GONZALEZ.

DIRECTOR: ING. SANTIAGO AURELIO OCHOA PhD.

CUENCA-ECUADOR

2023

DIOS, PATRIA, CULTURA Y DESARROLLO



UNIVERSIDAD CATÓLICA DE CUENCA

Comunidad Educativa al Servicio del Pueblo

UNIDAD ACADÉMICA DE INGENIERÍA, INDUSTRIA Y CONSTRUCCIÓN.

CARRERA DE INGENIERÍA CIVIL.

EVALUACIÓN DE MODELOS DE PRONÓSTICO PARA ESTIMAR EL

INGRESO DE CAUDALES AL EMBALSE MAZAR: APLICACIÓN CON

REDES NEURONALES.

TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE

INGENIERO CIVIL.

AUTOR: JUDER ANTONIO CASTILLO GONZALEZ

DIRECTOR: ING. SANTIAGO AURELIO OCHOA PhD.

CUENCA - ECUADOR

2023

DIOS, PATRIA, CULTURA Y DESARROLLO

DECLARATORIA DE AUTORÍA Y RESPONSABILIDAD

Juder Antonio Castillo González portador de la cédula de ciudadanía N.º **1105377004**. Declaro ser el autor de la obra: **“EVALUACION DE MODELOS DE PRONOSTICO PARA ESTIMAR EL INGRESO DE CAUDALES AL EMBALSE MAZAR: APLICACIÓN CON REDES NEURONALES”** sobre la cual me hago responsable sobre las opiniones, versiones e ideas expresadas. Declaro que la misma ha sido elaborada respetando los derechos de propiedad intelectual de terceros y eximimos a la Universidad Católica de Cuenca sobre cualquier reclamación que pudiera existir al respecto. Declaro finalmente que esta obra ha sido realizada cumpliendo con todos los requisitos legales, éticos y bioéticos de investigación, que la misma no incumple con la normativa nacional e internacional en el área específica de investigación, sobre la que también me responsabilizo y eximimos a la Universidad Católica de Cuenca de toda reclamación al respecto.

Cuenca, **28 de febrero de 2023**.

F: 

Juder Antonio Castillo González.

1105377004

CERTIFICACIÓN

Yo, Ing. Santiago Aurelio Ochoa PhD con CI: 0104158142 en calidad de director de tesis certifico que el trabajo **“EVALUACIÓN DE MODELOS DE PRONOSTICO PARA ESTIMAR EL INGRESO DE CAUDALES AL EMBALSE MAZAR: APLICACIÓN CON REDES NEURONALES”**, fue realizado por el estudiante Juder Antonio Castillo González, bajo mi supervisión.

Cuenca, **28 de febrero de 2023.**

A handwritten signature in blue ink, reading "S. Aurelio Ochoa S.", is written over a horizontal dotted line. The signature is stylized and somewhat cursive.

Ing. Santiago Aurelio Ochoa PhD.

DIRECTOR.

Dedicatoria

Dedico esta tesis al creador de todo, al nuestro Padre celestial, por regalarme la oportunidad de seguir en el camino de mi formación académica, así mismo, agradezco a mi padre “Víctor Castillo” por su apoyo y constancia en este arduo camino, gracias por todos los consejos que me brindó y sigue brindándome.

Gracias a mi madrecita, porque sin usted nada fuera posible, mi pilar de vida y mi primera amiga, me enseñaste que las cosas con esfuerzo se logran y que a pesar de todas las adversidades siempre estás apoyando y ayudándome a saltar cada obstáculo que me presenta la vida.

A mis hermanos Alex Y Darwin, que ya no me acompañan en este mundo, pero desde el paraíso están cuidando y protegiendo de mí.

A mis hermanas Janeth y Magaly, que son mi ejemplo para seguir y quienes me impulsan cada día a ir por más y ser mejor persona.

Gracias a todos y todas, Juder A. Castillo.

Agradecimientos

Sin duda la formación académica va de la mano con la docencia, por lo que, mis más sinceros agradecimientos para el Ing. Santiago Aurelio Ochoa PhD por brindarme su consejo y conocimiento, por ser mi guía y mi apoyo en esta tesis.

A la Ing. Sandra Lucia Cobos Mora, quien mediante su gran calidad humana y profesional me brindó los mejores consejos para llevar a cabo esta tesis.

RESUMEN

La cuenca que conforma el río Paute tiene gran importancia estratégica nacional, aquí se encuentran las hidroeléctricas Molino, Mazar, Sopladora y se proyecta Cadernillo. La superficie que aporta caudal al embalse Mazar es de 4228.43 km².

Las redes neuronales artificiales (RNA) son un método de pronóstico que está a la vanguardia que maneja y analiza bases de datos muy extensas con el fin de innovar y desarrollar las diferentes áreas del conocimiento; se pueden aplicar al reconocimiento de imágenes, búsqueda de patrones, clasificaciones, pronósticos, etc.

Se desarrolló un modelo de red neuronal que permite pronosticar los niveles del río y mediante la curva de descarga de caudales se transformó los niveles de río a caudal, la base de datos se dividió en un 80% de los registros para el entrenamiento y un 20% para su validación, la muestra fue aleatoria y los errores entre los niveles de caudal observados y pronosticados en todas las iteraciones siempre se mantuvieron por debajo del 10% y un mínimo de 7.57%, en términos de error porcentual medio, lo que refleja la estabilidad del modelo de pronóstico desarrollado.

La arquitectura de la red neuronal consta de 7 entradas que son los valores de nivel de agua de las estaciones hidrológicas YanuncayAjTarquiL, MataderoSyausiPL, TomebambaQtaBalzayL, MachangaraLLanteraL, TarquiAjYanuncayL, TomebambaPTARL, GualaceoDjPamarPL; 1 capa oculta con 50 neuronas; y 1 capa de salida con 1 neurona, que a su vez es el pronóstico que se comparará con los datos de la estación hidrológica Paute, punto de ingreso al embalse Mazar.

Palabras clave: redes neuronales, hidroeléctricas, pronóstico, base de datos.

ABSTRACT

The basin that forms the Paute River has great national strategic importance, as it houses Molino, Mazar, and Sopladora hydroelectric power plants, and the Cadernillo project is currently underway. The area contributing to the Mazar reservoir is 4228.43 km². Artificial neural networks (ANN) are a cutting-edge forecasting method that manages and analyzes very extensive databases to innovate and develop different areas of knowledge; they can be applied to image and pattern recognition, classification, and forecasting, among other applications. A neural network model was developed to forecast river levels, and the river levels were transformed into river flow through the discharge curve. The database was split into 80% for training and 20% for validation, and the sample was randomized. The errors between the observed and predicted flow levels in all iterations were always kept below 10% and a minimum of 7.57% in terms of mean percentage error, which reflects the stability of the developed forecast model. The neural network architecture consists of seven inputs, which are the water level values of the YanuncayAjTarquiL, MataderoSayausiPL, TomebambaQtaBalzayL, MachangaraLLanteraL, TarquiAjYanuncayL, TomebambaPTARL, and GualaceoDjPamarPL hydrological stations; one hidden layer with 50 neurons, and one output layer with one neuron, which is the forecast that will be compared with the data from the Paute hydrological station —the point of entry into the Mazar reservoir—.

Keywords: neural networks, hydroelectric power plants, forecast, database

ÍNDICE GENERAL

CAPÍTULO I: INTRODUCCIÓN	1
1.1 Formulación del Problema.	1
1.2 Objetivos	2
1.2.1 Objetivo General.....	2
1.2.2 Objetivo Específico.....	2
1.3 Justificación.....	2
CAPÍTULO II: MARCO TEÓRICO	4
2.1 Tipo de Investigación	4
2.2 Nivel de Investigación.....	4
2.3 Población, Muestra, Muestreo.....	4
2.3.1 Población.....	4
2.3.2 Muestra	4
2.3.3 Muestreo	5
2.4 Técnicas e Instrumentos de Recolección de Datos	5
2.4.1 Técnicas	5
2.4.2 Observación	5
2.4.3 La medición	5
2.4.4 Recolección de información	5
2.4.5 La descripción.....	6
2.4.6 Instrumentos.....	6
2.4.7 Recolección de datos en instituciones gubernamentales y no gubernamentales.	7
2.5 Técnicas de Procesamiento y Análisis de Datos.	7
2.5.1 Técnicas de procesamiento	7
2.5.2 Análisis de datos	8
2.6 Esquema metodológico.	9
2.7 Antecedentes generales	11
2.7.1 Historia de las redes neuronales.....	11
2.7.2 Estado de arte de las RNA: Investigaciones hoy en día.....	12
2.8 Cuenca Hidrográfica	14

2.8.1	Características morfométricas de la cuenca.....	15
2.8.2	Precipitación	16
2.8.3	Escurrimiento.....	17
2.8.4	Caudal	19
2.8.5	Hidrograma	20
2.8.6	Curva de descarga.....	21
2.9	Pronóstico hidrológico	22
2.10	Redes neuronales artificiales.....	23
2.10.1	Modelo biológico.....	23
2.10.2	Analogía de una neurona biológica y Artificial.....	24
2.10.3	Definición de una red neuronal artificial	24
2.10.4	Elementos de una red neuronal artificial	25
2.10.5	Arquitectura de las RNA.....	27
2.10.6	Mecanismo de aprendizaje.....	28
2.10.7	Perceptrón multicapa (MLP).....	29
2.11	Variables de estudio.	30
2.11.1	Variable independiente.	30
2.11.2	Variable dependiente.	31
CAPÍTULO III: METODOLOGÍA DE INVESTIGACIÓN.....		32
3.1	Escala Temporal y Espacial	32
3.1.1	Escala temporal.....	32
3.1.2	Escala espacial	33
3.2	Análisis de información.	34
3.2.1	Área de estudio.	34
3.2.2	Información hidrológica.....	40
3.2.3	Estimación de la curva de descarga	49
3.2.4	Predicción de niveles mediante redes neuronales.....	52
CAPÍTULO IV: PRESENTACIÓN DE RESULTADOS		55
4.1	Resultados del modelo en la estación Paute.....	55
4.2	Estadísticos de los datos observados y pronosticado.	57
4.2.1	Varianza y desviación estándar de los datos empleados y pronosticados.	57
4.3	Estimación de errores.....	60

4.3.1	Predicción de caudales	63
4.4	Discusión de resultados	64
CAPITULO V: CONCLUSIONES.		66
RECOMENDACIONES.....		67
REFERENCIAS.....		68
ANEXOS		72
	Base de datos.....	72
	Código en R.....	83
	Información geográfica de la cuenca hidrográfica.....	97

ÍNDICE DE TABLAS

Tabla 1. Técnicas e instrumentos para la investigación.....	6
Tabla 2. Insumos para el análisis de datos.....	8
Tabla 3. Caracterización general de la cuenca hidrográfica.	36
Tabla 4. Ubicación de estaciones hidrológicas.	36
Tabla 5. Parámetros morfométricos de la cuenca.	37
Tabla 6. Resumen estadístico de las estaciones hidrometeorológicas.	41
Tabla 7. Ubicación de la estación YanuncayAjTarquiL.....	43
Tabla 8. Ubicación de la estación MataderoSyausiPL.....	43
Tabla 9. Ubicación de la estación TomebambaQtaBalzayL.....	44
Tabla 10. Ubicación de la estación MachangaraLLanteraL.	45
Tabla 11. Ubicación de la estación TarquiAjYanuncayL.....	46
Tabla 12. Ubicación de la estación TomebambaPTARL.....	47
Tabla 13. Ubicación de la estación GualaceoDjPamarPL.	48
Tabla 14. Ubicación de la estación Paute.	49
Tabla 15. Estructura de los datos de ingreso para determinar la curva de descarga.....	50
Tabla 16. Parámetros para definir la ecuación de curva de descarga.	50
Tabla 17. Estructura de los datos de ingreso a la red neuronal.....	52
Tabla 18. Resultados de los pronósticos con el modelo de red neuronal planteado.....	55
Tabla 19: Estadísticos de datos observados.....	57
Tabla 20: Estadísticos de datos pronosticados.....	59
Tabla 21: Errores de pronóstico.....	61
Tabla 22: Nivel estimado y observado con transformación a caudal en diferentes fechas.....	63
Tabla 23: Base de datos.	72

ÍNDICE DE FIGURAS

Figura 1. Diagrama metodológico de la investigación.	10
Figura 2. La cuenca hidrográfica.	14
Figura 3. Hietograma de lluvia total y exceso de precipitación.....	17
Figura 4. Escurrimiento.	18
Figura 5. Caudal.....	19
Figura 6. Hidrograma anual.	20
Figura 7. Escurrimiento directo.	21
Figura 8. Componentes de una red neuronal.	23
Figura 9. Similitud entre una neurona bilógica y una artificial.	24
Figura 10. Arquitectura de una red neuronal.	25
Figura 11. Elementos de una red neuronal artificial.	26
Figura 12. Arquitectura de una red neuronal artificial.....	27
Figura 13. Estructura del perceptrón multicapa.....	30
Figura 14: Serie de tiempo de la estación hidrológica Paute.....	32
Figura 15. Cuenca hidrográfica que aporta caudal al Embalse Mazar.	33
Figura 16. Área y subcuencas de estudio.....	35
Figura 17. Red de drenaje.	39
Figura 18. Orden de corrientes.....	40
Figura 19. Distribución de variables continuas.....	42
Figura 20. Serie de tiempo de la estación YanuncayAjTarquiL.....	43
Figura 21. Serie de tiempo de la estación MataderoSayausiPL.....	44
Figura 22. Serie de tiempo de la estación TomebambaQtaBalzayL.....	45
Figura 23. Serie de tiempo de la estación MachangaraLLanteraL.	46

Figura 24. Serie de tiempo de la estación TarquiAjYanuncayL.....	47
Figura 25. Serie de tiempo de la estación TomebambaPTARL.....	48
Figura 26. Serie de tiempo de la estación Gualaceo GualaceoDjPamarPL.	49
Figura 27. Curva de descarga.....	51
Figura 28. Porcentaje de datos ausentes de cada variable.	54
Figura 29: Resultados del modelo.....	56
Figura 30: Comparación gráfica de resultados del modelo.	65

CAPÍTULO I: INTRODUCCIÓN

La generación de energía eléctrica es un servicio que beneficia al país y, por lo tanto, a sus habitantes. Ecuador cuenta con recursos para generar una gran potencia energética; y, por lo tanto, contar con un pronóstico del potencial hídrico para generar energía eléctrica limpia es una herramienta fundamental de gran importancia, ésta una de las razones por las que se aborda el tema. Existen modelos de pronóstico de caudales muy sencillos y otros muy complejos. La complejidad del modelo se puede medir en función de la cantidad de variables o información procesada. Para desarrollar el modelo se cuenta con datos del monitoreo de información hidrometeorológica de instituciones como CELEC, ETAPA EP e INAMHI.

Los modelos para pronosticar pueden ser cualitativos o cuantitativos; en referencia a los cualitativos, son modelos que no tienen una base firme, sin embargo, están estrechamente ligados a la opinión y experiencia de los expertos. Los métodos cuantitativos se basan en la estadística y procesos bien marcados, se dividen en tres grupos: series de tiempo, métodos casuales y los algoritmos (Perdigón & González, 2020). Dentro de los algoritmos se encuentran las redes neuronales (RN). El término se lo asocia a la biología del cerebro, ya que, de forma general, el cerebro funciona con neuronas interconectadas que hacen sinapsis entre millones de ellas. Entonces, el término apropiado es red neuronal artificial (RNA), ya que mediante programación y estadística se busca emular el funcionamiento del cerebro para pronosticar (Serna, 2017).

1.1 Formulación del Problema.

Los modelos de pronóstico de caudales generalmente se ven limitados a usar regresiones lineales, estos modelos basan su análisis en medir la relación entre variables dependientes e independientes y usan como datos de entrada los valores de caudales medidos a lo largo de muchos años. Pero los datos no son lineales, por lo tanto, es indispensable incluir otro tipo de variables de tipo hidrológico y morfológico para que los pronósticos tengan un nivel de certeza de mayor calidad (Béjan, Valeriano, Ilachoque, & Sulla, 2016), entonces, estos modelos no son apropiados y aquí nace la necesidad de implementar algoritmos para la estimación de las variables de interés.

Actualmente las redes neuronales artificiales (RNA) están ganando espacio en una infinidad de aplicaciones en diferentes ramas de la ciencia, se han realizado diversos estudios que integran las

RNA cuando se busca obtener un pronóstico con un nivel de confianza de mayor calidad que el que proporcionan las regresiones lineales (Nastos, Paliatsos, Koukouletsos, Larissi, & Moustris, 2014). Una de las técnicas más usadas en este campo y es la se desarrolla en este trabajo investigativo, son las redes neuronales artificiales para obtener pronósticos con un nivel de certeza de mejor calidad, básicamente simulan el funcionamiento del cerebro para dar solución a problemas de diversa índole aprendiendo de sus aciertos y errores.

1.2 Objetivos

1.2.1 Objetivo General

- Desarrollar un modelo de pronóstico de caudales de ingreso al embalse Mazar, aplicando metodologías estadísticas con redes neuronales artificiales.

1.2.2 Objetivo Específico

- Analizar y procesar las variables hidrometeorológicas y morfológicas de la cuenca hidrográfica, para construir el modelo de pronóstico.
- Calibrar y validar un modelo de redes neuronales artificiales para obtener los caudales medios.

1.3 Justificación.

Los modelos de pronóstico de caudales usando redes neuronales artificiales son métodos que está a la vanguardia dentro de investigaciones teórico-prácticas dedicadas a realizar pronósticos de eventos o fenómenos naturales.

Esta investigación es completamente viable, no se requiere de grandes recursos materiales o humanos, sin embargo, los registros históricos de caudales y niveles de agua de las estaciones hidrológicas dentro de la cuenca son indispensables. La investigación no afecta a ningún componente social o material, sin embargo, el nivel de importancia es alto, dado que conocer el caudal es un dato indispensable para optimizar la producción de energía eléctrica limpia, pero al

ser pronósticos, siempre habrá un nivel de incertidumbre que se minimiza usando modelos de redes neuronales artificiales entrenados correctamente.

Este trabajo se puede llevar a una publicación y posteriormente a trabajar en un programa o aplicación sencilla que pueda ser usada por cualquier profesional y replicarse en otros lugares, por ende, se procura que las variables a considerar sean de fácil obtención, sin afectar la efectividad del método.

CAPÍTULO II: MARCO TEÓRICO

2.1 Tipo de Investigación

En este trabajo el tipo de investigación que se realiza es aplicada no experimental, puesto que se lleva a la práctica el componente teórico de las redes neuronales y no es experimental porque no se manipula indiscriminadamente las variables. Aquí se busca que el pronóstico del caudal sea de mayor confianza que el proporcionado por otros métodos, para ello, se parte de información básica de estaciones hidrológicas y meteorológicas. La investigación que se realiza se conoce también como Ex Post Facto, puesto que el investigador no tiene acción alguna sobre los hechos (datos hidrológicos), dado que ya ocurrieron (Sanca, 2011).

2.2 Nivel de Investigación

Esta investigación está enmarcada desde la exploración de los datos, la descripción de los mismos, las relaciones entre sí, explicación de su comportamiento, hasta la predicción de los caudales, se busca explicar la conducta de una variable en función de otra (Sanca, 2011).

2.3 Población, Muestra, Muestreo.

2.3.1 Población

Se considera población a todos los datos recopilados de la serie de tiempo, de las estaciones hidrometeorológicas dentro de la zona de estudio. Estos datos parten desde el mes de junio del año 2014 hasta el mes de junio del año 2019.

2.3.2 Muestra

La muestra es considerada como una parte de la población, esta parte de la población reúne una característica similar entre cada uno de sus datos, dicha característica es la que se desea estudiar. La muestra se considera desde el mes de agosto del 2017 a junio del 2018.

2.3.3 Muestreo

Dado que dentro de los datos que se deben proporcionar a la red neuronal se debe contar con registros en todas las estaciones en el mismo día de aforo, por lo que se seleccionó las siete estaciones que contenían más registros y que a su vez estaban mejor distribuidas en la cuenca.

2.4 Técnicas e Instrumentos de Recolección de Datos

2.4.1 Técnicas

A continuación, se van a describir las técnicas que se utilizaron para la recopilación, manejo y procesamiento de la información que se pudo obtener para este trabajo investigativo:

2.4.2 Observación

En cualquier trabajo de investigación, la observación es fundamental, ya que mediante ella podemos identificar patrones, comportamientos y hechos que pueden ser relevantes para la investigación. En cualquier etapa en la que se encuentra el trabajo investigativo, desde la recolección de los datos hasta la presentación de los resultados, se usará esta técnica.

2.4.3 La medición

Considero que esta técnica dentro de cualquier investigación hidrológica es indispensable, ya que, generalmente, siempre se trabaja con variables que son cuantificables como la precipitación, el caudal, el nivel del río, entre otras variables. También existen variables que no se pueden cuantificar, sin embargo, se pueden categorizar, como el tipo de suelo de la cuenca hidrográfica.

2.4.4 Recolección de información

Esta parte hace referencia a toda la información disponible que se puede recolectar ya sea de entidades gubernamentales y no gubernamentales. Para este trabajo investigativo es indispensable los registros de las estaciones hidrometeorológicas, asimismo, como cierta información geográfica que nos ayudará a delimitar la zona de estudio.

2.4.5 La descripción

La investigación se enmarca en el campo aplicado no experimental, como se explicó en el apartado 2.1, parte con la recopilación de datos hidrometeorológicos cuyas fuentes de información fueron el INAMHI, CELEC EP y ETAPA EP. Dentro del análisis de la información se discretiza la variable independiente que ofrezca mejores resultados, se probó la precipitación como variable independiente, sin embargo, el modelo no ofrecía un nivel de certeza alto, dado que la precipitación a pesar de estar relacionada al caudal pasa por diversos procesos para formar parte de él, lo cual agrega incertidumbre al pronóstico. Es por ello, que se pasa a analizar la variable “nivel de agua” de las estaciones que contaban con registros. Luego de analizar esta información se ingresan los datos en el modelo, se obtienen los pronósticos y se verifica que ofrezcan un mejor nivel de certeza.

2.4.6 Instrumentos

Las técnicas anteriores necesitan ciertos instrumentos, además, el talento humano, para su correcto funcionamiento:

Tabla 1. Técnicas e instrumentos para la investigación.

Fuente: Elaboración propia.

Técnicas	Instrumentos
Observación	<ul style="list-style-type: none">• Computador
La medición	<ul style="list-style-type: none">• Computador• Estaciones hidrometeorológicas
Recolección de información	<ul style="list-style-type: none">• Computador• Instrucciones para obtención de datos
La descripción	<ul style="list-style-type: none">• Computador• Sistemas de información geográfica (ArcGIS, Qgis)

	<ul style="list-style-type: none"> • Software para procesamiento de información (Editor de texto, Microsoft Excel, R)
--	--

2.4.7 Recolección de datos en instituciones gubernamentales y no gubernamentales.

La información hidrometeorológica se obtuvo de la Empresa Municipal de Telecomunicaciones, Agua Potable, Alcantarillado y Saneamiento de Cuenca (ETAPA EP), que tiene un constante monitoreo de la cuenca del Río Paute. Estos datos se complementan con los obtenidos de la Corporación Eléctrica del Ecuador (CELEC) y el Instituto Nacional de Meteorología e Hidrología (INAMHI) para formar una data lo suficientemente extensa y válida para ejecutar dentro del algoritmo de la red neuronal artificial, ya que mientras mayor sea el número de registros diarios, más ejemplos tendrá la red neuronal y podrá brindar un pronóstico con un mayor nivel de certeza. La población de datos parte en junio de 2014 hasta junio de 2019, pero la muestra se reduce al periodo de agosto de 2017 hasta junio de 2018 debido a que no todas las estaciones seleccionadas cuentan con los registros del mismo día.

2.5 Técnicas de Procesamiento y Análisis de Datos.

2.5.1 Técnicas de procesamiento

Las técnicas de procesamiento que se aplicaron en esta investigación fueron dos: estadística descriptiva y estadística inferencial.

Estadística descriptiva: Todos los registros hidrológicos y meteorológicos obtenidos en la fase de recolección de datos (tablas, gráficos) son de carácter cuantitativo, dado esto, se puede aplicar ciertas técnicas de exploración y evaluar las variables de entrada para el modelo de RNA.

Estadística inferencial: Se examinan los datos que provienen de las estaciones hidrológicas y meteorológicas, para ello se hace uso de estadísticos como: raíz cuadrada del error cuadrático medio (RMSE) (Béjan et al., 2016), error porcentual medio absoluto (MAPE) (Béjan et al., 2016), suma cuadrada de residuos (SSE) y coeficiente de determinación (R²) (Oré, 2019).

2.5.2 Análisis de datos

Los datos que se recopilaron deben ser analizados y procesados en softwares y lenguajes de programación, por ello, se describen de manera breve el software o lenguaje de programación y en que va a ser usado:

Tabla 2. Insumos para el análisis de datos.

Fuente: Elaboración Propia.

Global Mapper, ArcGIS o QGIS	Estos softwares sirven para la exploración, análisis y creación de nueva información geográfica, en este punto, se analiza las propiedades físicas de la cuenca y se concreta las estaciones que se encuentran dentro de la zona de interés y, posteriormente, si hay registros formaran parte de la data.
R	Este es un lenguaje de programación muy potente y de libre acceso, dentro se realiza el procesamiento de los datos y la construcción del algoritmo de red neuronal, así mismo, se analiza los resultados y diversos estadísticos que evaluarán la precisión del modelo
Microsoft Excel	Este software es imprescindible para la exploración y organización de la información recopilada.

2.6 Esquema metodológico.

El diagrama metodológico busca representar las etapas a seguir en esta investigación, con el fin de llegar al producto final que es el pronóstico de caudal. En la figura 1 se representa el diagrama metodológico que sigue esta investigación.

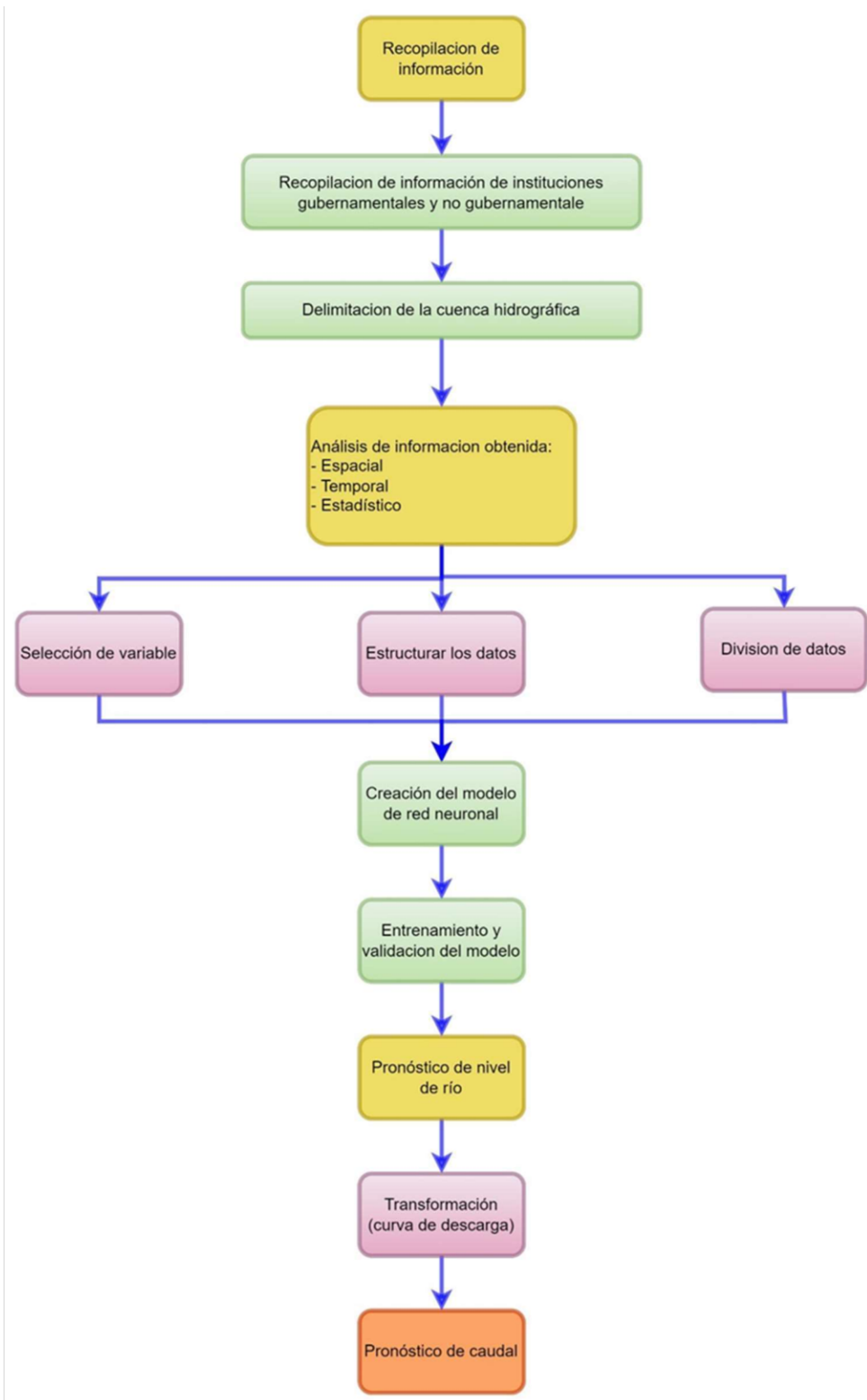


Figura 1. Diagrama metodológico de la investigación.

Fuente: Elaboración propia.

2.7 Antecedentes generales

2.7.1 Historia de las redes neuronales.

A continuación, se describe en orden cronológico el avance y desarrollo teórico de las Redes Neuronales:

En 1936, Alan Turing, miró desde la computación el funcionamiento del cerebro, sin embargo, en 1943, Warren McCulloch y Walter Pitts, neurofisiólogo y matemático, respectivamente, propusieron el primer método teórico de cómo trabajan las neuronas en su publicación titulada “Un Cálculo Lógico de la Inminente Idea de la Actividad Nerviosa - Boletín de Matemática Biofísica 5: 115-133” (Matich, 2001).

En 1949, Donal Hebb, explico desde la perspectiva psicológica los procesos de aprendizaje y como estos ocurrían. Se puede decir que los trabajos de Hebb fueron los principios teóricos de las RNA (Andrade Tepán, 2013).

En los años posteriores, hasta 1956, fue que Karl Lashley descubrió que la información guardada en el cerebro no estaba en el centro de éste, si no que era repartida sobre él. En este mismo año se celebra el Congreso de Dartmouth, que es concebido como el origen de la inteligencia artificial como campo de actividad (Matich, 2001).

En 1957, Frank Rosenblatt, empezó el desarrollo del perceptrón, la más simple y antigua red neuronal. Este perceptrón es capaz de generalizar y reconoces patrones similares sin que antes se hayan plasmado en el entrenamiento, pero no era capaz de clasificar datos no lineales (Matich, 2001).

En 1959, Frank Rosenblatt propone y demuestra el Teorema de Convergencia del Perceptrón, es decir, el perceptrón converge, bajo ciertas condiciones, hacia un estado finito (Matich, 2001).

Bernard Widrow y Marcial Hoff, en 1960, presentaron la red neuronal “Adaptative Linear Elements” o “ADALINE”, este modelo resolvió el problema del eco en las líneas telefónicas. En 1967, Stephen Grossberg, presentó la red neuronal “Avalancha”, esta tenía su aplicación en reconocimiento continuo del habla (Andrade Tepán, 2013).

Marvin Minsky y Seymour Papert en 1969, criticaron fuertemente al perceptrón, ya que no podía resolver problemas fáciles como el aprendizaje de una función no lineal, tomando en cuenta que las funciones no lineales son extensamente empleadas en el mundo real. Sin embargo, se continuaron las investigaciones sobre RNA y en ese mismo año, James Anderson, presentó un modelo lineal denominado “Asociador Lineal” (Andrade Tepán, 2013).

Paul Werbos, en 1974, introduce la primera idea del algoritmo *Backpropagation* o *aprendizaje hacia atrás*, pero su estudio se aclaró hasta 1985 (Matich, 2001).

En 1977, Sephen Grossberg, logra simular destrezas del cerebro como memoria a corto y largo plazo y lo denomina como “Teoría de resonancia adaptada”. En ese mismo año, Teuvo Kohonen, concibió un modelo de red neuronal similar al de Anderson en 1969. Años después, Kunihiko Fukushima, desarrollo una RNA que podía reconocer patrones visuales (Andrade Tepán, 2013).

A partir de 1985, el desarrollo de redes neuronales e investigación fue más amplia, además, John Hopfield hizo un gran aporte con su libro “Computación neuronal de decisiones en problemas de optimización”, y un año más tarde se redescubrió el algoritmo de *backpropagation* por David Rumelhart y Geoffrey Hinton (Matich, 2001).

2.7.2 Estado de arte de las RNA: Investigaciones hoy en día.

Diversos estudios dedicados a la predicción y proponen usar modelos de redes neuronales, por su eficacia y simplicidad del método. A continuación, se presentan algunos trabajos recientes:

En el trabajo investigativo denominado “Modelado del caudal natural en la cuenca hidrográfica Chambo con redes neuronales” se aplica diversas estructuras de de redes neuronales con una data de 14 años, sin embargo, el modelo reconstruye de manera correcta los caudales con una data de 6 años, así mismo, se miden diversas métricas con el fin de validar la eficiencia del modelo. Finalmente, se desarrolla una aplicación para simular el estudio en otras cuncas hidrográficas (Turleone, 2022).

En el estudio realizado por Angarita Jhon y Pérez Walter en 2021, sobre “Comportamiento tarifario de energías renovables tipo eólica y solar en la generación de energía eléctrica en Colombia”, se aplica las RN al modelamiento de datos, con ello se busca pronosticar los costos de la electricidad en la bolsa nacional en Colombia y en las fuentes no convencionales de energía renovables, obteniendo excelentes resultados (Angarita & Perez, 2021).

En el artículo titulado “Redes neuronales para pronóstico de series de tiempo hidrológicas del Caribe colombiano” se trabaja una red neuronal tipo perceptrón multicapa y se establecen 4 fases bien marcadas: diseño estructural, diseño funcional, selección del mejor modelo y validación. Analizan seis bases de datos y dos algoritmos de entrenamiento que son ADAM y SGD. Como punto de comparación se usó la Raíz del Error Cuadrático Medio (RMSE), luego de aplicar el algoritmo experimental propuesto, se obtuvieron resultados satisfactorios, además, los niveles de error alcanzados fueron cercanos a cero (de la Hoz, Canchano, Cordero, & Sánchez, 2019).

Amador (2021), en su investigación titulada “Red Neuronal Artificial (RNA) para pronóstico de intensidad de lluvia a 10 minutos en el área Metropolitana de Guadalajara”, usa RNA tipo Multilayer FeedForward, para pronosticar precipitación nula, fuerte y tempestuoso. El código en MATLAB de retro-propagación del error Levenberg-Marquardt se usó en el entrenamiento de la red. Se obtuvo un error medio cuadrático de 1.13×10^{-14} , en el periodo de ensayo y admisión el 100 % de los datos fue clasificado sin problemas. Se ejecutaron 7 ensayos con datos diferentes a los de entrenamiento y validación de la red, obteniendo un pronóstico correcto en el 70% de los casos (Amador, 2018).

Para trabajar con redes neuronales artificiales, ningún autor establece un número mínimo o máximo de registros, porque el propósito de la investigación en todos los casos es diferente, por ejemplo, si se trabaja con redes neuronales para clasificación binaria el resultado puede ser categórico como un “sí” o “no”, o puede ser numérico como “1” o “0”, en estos casos la red neuronales tiene que decidir entre dos posibles resultados, sin embargo, cuando las redes neuronales se utilizan para pronosticar ya no se tiene solo dos posibles resultados, si no un universo de números que pueden ser posibles soluciones. Es así que, cuando se busca pronosticar valores es necesario el mayor número de registros

posibles, para que la red neuronal en la fase de entrenamiento aprenda que con cierta combinación de registros se obtiene un resultado.

2.8 Cuenca Hidrográfica

En la noción de cuenca hidrográfica se basa todo diseño hidrológico. Dado que una gran cuenca hidrológica consta de muchas cuencas de menor tamaño, la cuenca está restringida por los puntos de mayor elevación que forman el límite entre la cuenca y las subcuencas, y el agua de lluvia que escurre desde estos puntos proporcionará el caudal de salida.

Se representa mejor gráficamente, en la Figura 2 representa la cuenca de salida en el punto A. El agua se deposita en un punto de salida de muchas áreas más pequeñas, generalmente es el punto con menor elevación, que también son cuencas hidrográficas más pequeñas, y pueden dar origen a ríos, arroyos o simplemente corrientes de corta duración (Oré, 2019).

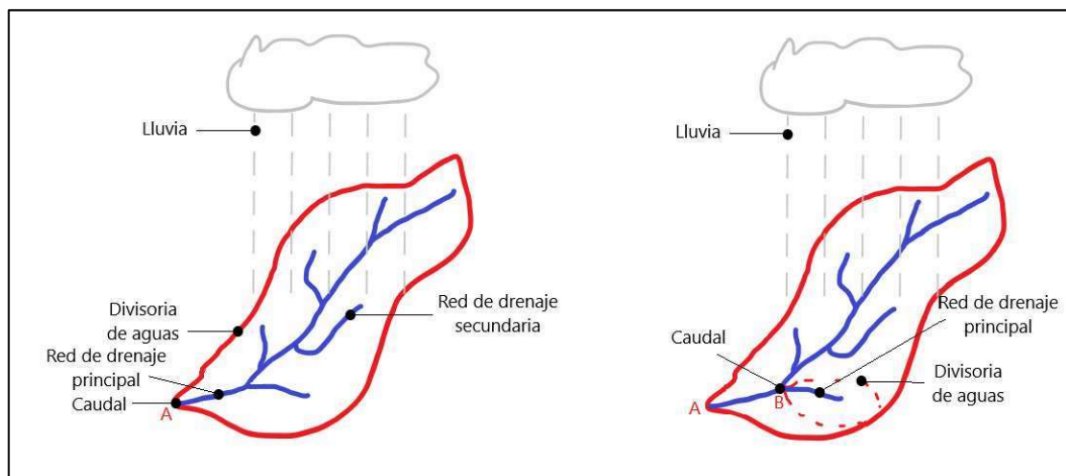


Figura 2. La cuenca hidrográfica.

Fuente: Elaboración propia.

Por ejemplo, si diseña en el punto B en lugar de A (Figura 2), la cuenca será el área pequeña enmarcada dentro de la línea punteada. La cuenca en el punto B también constará de cuencas o subcuencas más pequeñas (Oré, 2019).

2.8.1 Características morfométricas de la cuenca

Los caudales guardan una relación muy cercana con las características físicas de la cuenca hidrográfica. Los parámetros morfométricos componen un grupo de valores calculados, en general, al principio de un estudio hidrológico, con propósitos de explotación o control (Gaspari et al., 2012).

El análisis morfométrico está compuesto por variables lineales, de superficie, de relieve y drenaje; que ofrecen una idea intuitiva de las propiedades físicas de una cuenca (Lux, 2020), esto permite hacer comparaciones entre cuencas hidrográficas y definir la mejor estrategia para su manejo.

El área de la cuenca es la proyección horizontal de la superficie de escorrentía que está delimitada por la divisoria de aguas, este parámetro se expresa en la mayoría de los casos en kilómetros cuadrados (km^2) (Oré, 2019).

La longitud de la divisoria de aguas proyectada en un plano horizontal, se conoce como perímetro de la cuenca y se expresa en kilómetros (km) o en metros (m) (Oré, 2019).

Longitud del cauce principal es la longitud del curso de agua más largo de la cuenca hidrográfica, desde la parte alta de cuenca hasta la desembocadura, se mide sobre un plano horizontal y se expresa en kilómetros (km) o metros (m) (Oré, 2019).

El factor de forma es el cociente entre el ancho medio y la longitud del cauce principal de la cuenca hidrográfica.

La densidad de drenaje se puede definir como la suma de los cauces de agua entre el área de drenaje de la cuenca, si mayor es el área de drenaje, mayor el volumen de escurrimiento (Oré, 2019).

La pendiente de la cuenca depende netamente de la orografía, es importante para determinar el tiempo de concentración y de respuesta de la cuenca hidrográfica (Oré, 2019).

La pendiente media del cauce principal representa la diferencia de altura máxima y mínima del cauce principal, entre la longitud del cauce principal, también se puede realizar promediando cada segmento y alturas de la longitud principal (Oré, 2019).

El orden de corriente u orden de los ríos muestra la amplitud de ramificación de la corriente, estas pueden ser de orden 1, 2, 3 o superior. Cuando dos corrientes que se juntan tienen el mismo orden,

forman una de orden superior, pero si no son del mismo orden, conservan el orden mayor (Oré, 2019).

La curva hipsométrica es un gráfico que representa el área de la cuenca que hay por encima de una cota determinada (Oré, 2019).

El uso y ocupación del suelo afecta principalmente a la escorrentía ya que todos los usos y ocupación del suelo son diferentes. Por ejemplo, un suelo de ganadería es diferente de un suelo de agricultura por la vegetación que lo cubre, por ende, la escorrentía es diferente entre esos suelos (Oré, 2019).

Los suelos son diversos en toda la cuenca hidrográfica, más si la cuenca es extensa, estas variaciones en las características de los suelos son el tamaño de las partículas, estructura y porosidad; por ello, la infiltración también es diversa (Oré, 2019).

2.8.2 Precipitación

Según Chereque en 1985, la precipitación puede ser definida como toda forma de humedad que tiene su origen en las nubes, por lo tanto, las lluvias, las granizadas, las garuas y nevadas, son formas del fenómeno de la precipitación (Chereque, 2010).

El agua, producto de la precipitación marca el inicio o punto de para cualquier investigación, control o uso del agua, desde el punto de vista hidrológico. Además, es una variable que se presenta de forma aleatoria, varía notablemente en el tiempo y en el espacio (Oré, 2019).

La unidad de medida de la lluvia es el mm, esta medida representa la precipitación de un litro de agua en un m². En la actualidad existen métodos directos para medir la precipitación como el pluviógrafo y el pluviómetro, que son instrumentos que se utilizan a nivel terrestre. Sin embargo, se puede medir la precipitación de forma indirecta, para ello se utiliza radares y satélites de observación de la tierra (World Bank Group, 2022).

Precipitación efectiva

Cuando el agua, producto de la precipitación no se puede retener, pero tampoco se infiltra en el suelo, se le conoce como exceso de precipitación o precipitación efectiva. Por lo que, el exceso de precipitación fluye sobre la superficie de la cuenca hidrográfica dando origen a la escorrentía

directa que se dirige aguas abajo hacia la salida de la cuenca, bajo la suposición de flujo superficial (Chow, R. Maidment, & W. Mays, 1994).

Las abstracciones o pérdidas resultan de dos hietogramas, primero el de lluvia total y segundo el exceso de precipitación, sin embargo, las pérdidas también toman en cuenta el agua que es absorbida producto de la infiltración, la intercepción y almacenamiento superficial (Chow et al., 1994).

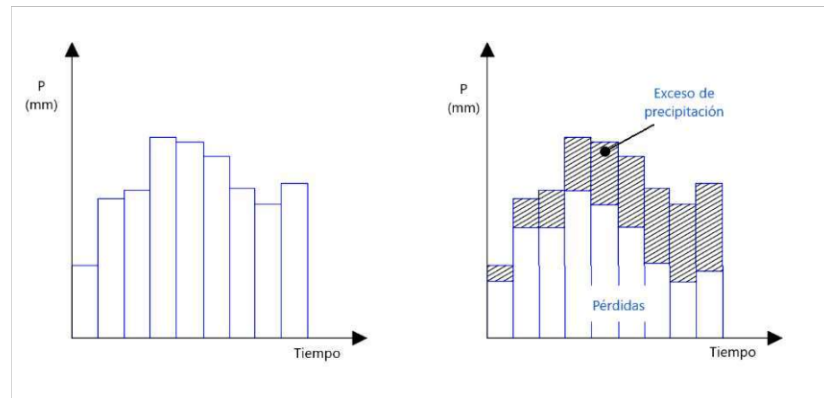


Figura 3. Hietograma de lluvia total y exceso de precipitación.

Fuente: Elaboración propia.

2.8.3 Escurrimiento

Cuando ocurre la precipitación, parte del agua circula sobre la superficie de la cuenca y otra parte debajo de la superficie, sin embargo, tanto la circulación superficial y la que se moviliza por debajo de la tierra convergen en un punto de drenaje, hasta la salida de la cuenca (Aparicio Mijares, 1989).

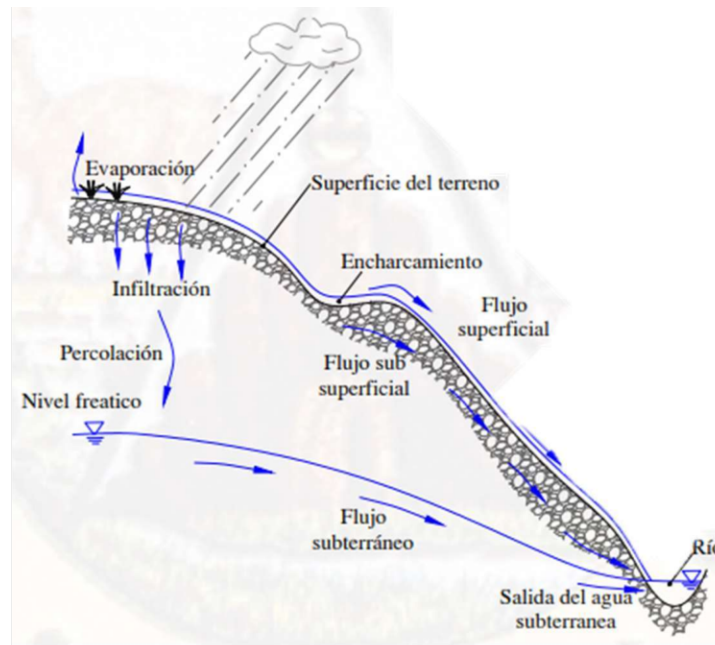


Figura 4. Escurrimiento.

Fuente: (Oré, 2019).

Para ilustrar mejor el fenómeno del escurrimiento se emplea la figura 4. La precipitación cae en forma de lluvia, parte de esta es retenida por la intersección de las hojas de la vegetación, ocurre almacenamiento superficial en las zonas donde hay depresiones en la superficie de la cuenca. Mientras la precipitación continúa, se forma una fina capa de agua la cual es conocida como detención superficial y empieza a fluir aguas abajo hacia los cursos de agua. Seguido aparece la escorrentía subsuperficial, estos dos tipos de escorrentía conforman la escorrentía directa. La percolación es el fenómeno en el cual el agua se mueve entre las partículas dentro del suelo y posteriormente ocurre la infiltración, el agua que se infiltra en exceso puede llegar a formar parte del agua subterránea y posteriormente se dirige a los cursos de agua (Chereque, 2010).

Tipos de escurrimiento

Luego de que haya caído la precipitación, que la evaporación se haya producido y que la interceptación por parte de la vegetación se haya dado, el agua restante continúa hacia el punto de drenaje. Existen tres formas en las que se da este fenómeno (Aparicio Mijares, 1989):

Escorrimento superficial: es el flujo de agua que se puede ver sobre la superficie del suelo, además, es el primero en llegar a los cauces de agua, tienes su origen en la precipitación efectiva, también forma parte del escurrimento directo.

Escorrimento subsuperficial: este flujo de agua se encuentra infiltrado en el suelo y su recorrido es más o menos paralelo a la superficie.

Escorrimento subterráneo: Este flujo de agua es el que más lento llega a la salida de la cuenca, debido a esto, también es el que más se infiltra, llegando incluso por debajo del nivel freático, asimismo, cuando no hay precipitación, este escurrimento forma el escurrimento base.

También existe la precipitación directa sobre el río, aportando volumen al cauce durante el ciclo inicial de la lluvia.

2.8.4 Caudal

Es la cantidad de agua, generalmente en m^3 , medida sobre el tiempo, generalmente en segundos, que circula en un cauce de agua en un tiempo específico.

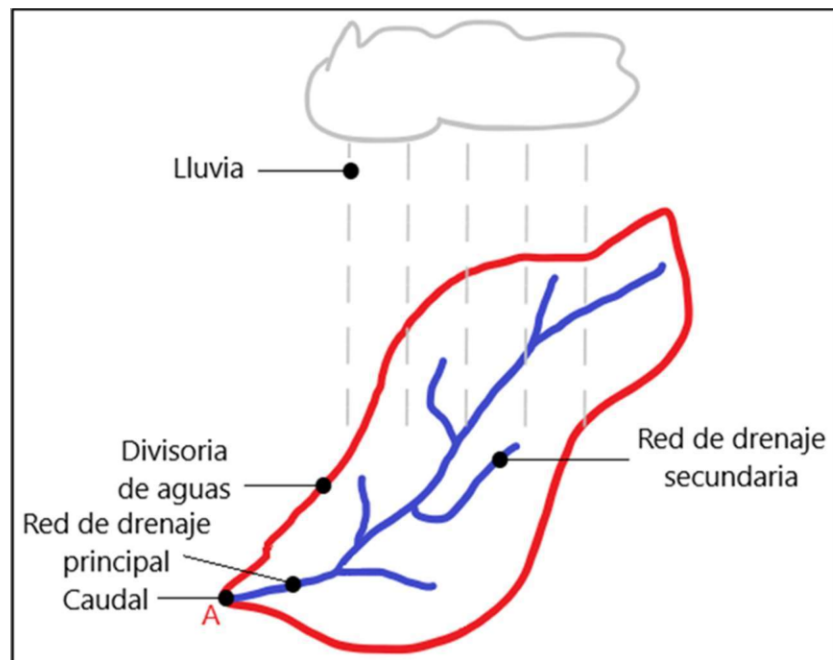


Figura 5. Caudal.

Elaboración propia.

Es imprescindible conocer los recursos hidráulicos que presenta una cuenca hidrográfica, ya que para cualquier estudio con fines de explotación se debe conocer el potencial hídrico, para esto se utiliza estaciones de aforo de caudal a nivel diario y durante el mayor tiempo posible. Todos los países velan por brindar esta información, en Ecuador no es la excepción, esta tarea la realiza el INAMHI en todo el territorio y adicionalmente en la parte de la provincia de Azuay lo realiza ETAPA EP.

2.8.5 Hidrograma

El hidrograma representa el caudal o volumen de agua que circula por una sección del río en un tiempo específico. Por lo tanto, el hidrograma representa las propiedades fisiográficas y climáticas que determinan las relaciones entre lluvia y escorrentía de una cuenca hidrográfica (Chow et al., 1994).

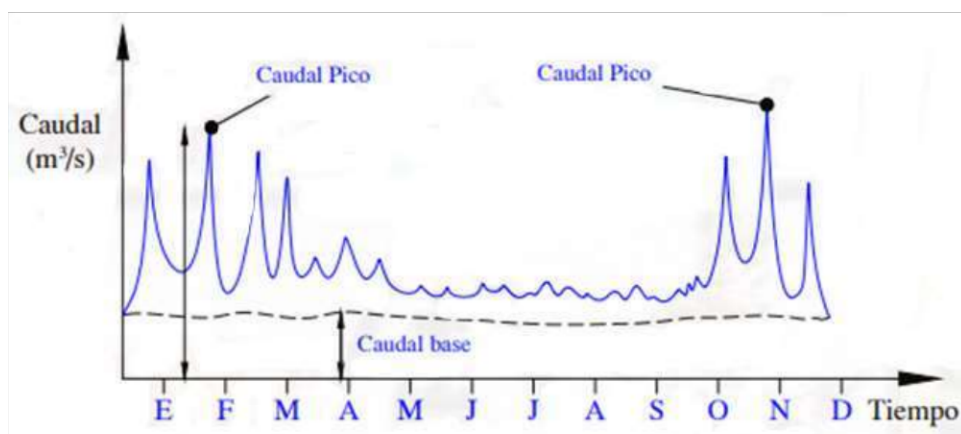


Figura 6. Hidrograma anual.

Fuente: (Oré, 2019).

En la figura 6 se ilustra el hidrograma anual, sin embargo, si la escala de tiempo se maximiza con la finalidad de observar solamente el caudal o escurrimiento que produce una tormenta, la Ilustración sería como la figura 7.

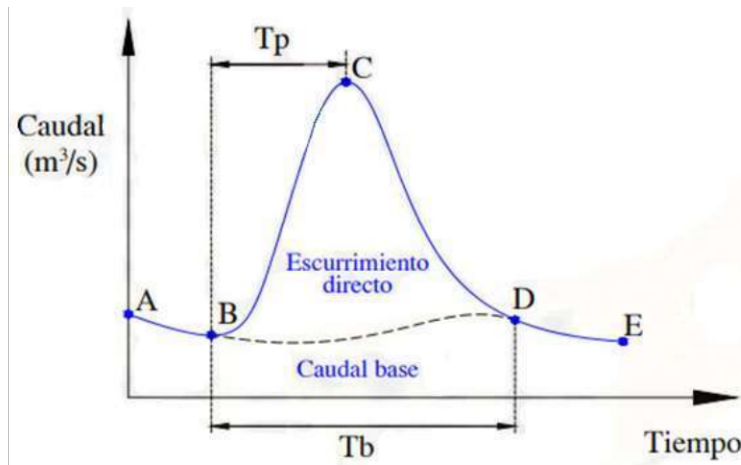


Figura 7. Esguimiento directo.

Fuente: (Oré, 2019).

Entonces, el volumen total de esguimiento directo generado por una tormenta se puede obtener del área total de la gráfica menos el caudal base (Aparicio Mijares, 1989). A continuación, la explicación de la terminología de la figura 7 (Oré, 2019):

- **B:** Punto inicial. Representa el momento en el que el agua empieza a observarse en el punto de la cuenca, producto de la tormenta.
- **C:** Pico. Es el caudal máximo que produjo la precipitación.
- **D:** Punto de inflexión. Llegados a este punto, aquí se marca el final del flujo superficial.
- **E:** Fin del esguimiento directo. Este punto representa que el esguimiento es solo subterráneo, iniciando la gráfica de agotamiento.
- **Tp:** Tiempo de pico. Es el tiempo desde el punto del levantamiento hasta cuando se produce el caudal pico.
- **Tb:** tiempo base. Es el tiempo que transcurre entre el punto de levantamiento y el punto de inflexión, conocido también como tiempo de esguimiento directo.

2.8.6 Curva de descarga.

La curva de calibración de caudales, consiste en un número bastante extenso de aforos de caudal durante un largo tiempo, con el objetivo de que dentro de los aforos haya tanto máximos como mínimos y se ajusten a la realidad del cauce, con esto se forma una relación entre el caudal y el nivel de agua en un tramo de río específico (Aparicio Mijares, 1989).

Ecuación 1: Ecuación general de la curva de calibración de caudales.

Fuente: (Arboleda, Moreno, & Santos, 2012).

$$Q = k(H - h_0)^n$$

Donde en la Ec. 1 se tiene:

Q : Caudal m^3/s

k, n : parámetros de ajuste (constante)

h_0 : Nivel cuando el caudal es cero (m).

H : Nivel actual del cauce (m)

2.9 Pronóstico hidrológico

Se puede definir como pronóstico hidrológico aquello que busca estimar el estado futuro de fenómenos hidrológicos; la necesidad de conocer con antelación estos fenómenos aumenta con el crecimiento económico de un país. Se clasifican según la lejanía del fenómeno que se busca pronosticar. Los pronósticos instantáneos que se usan los aforos en tiempo real, a corto plazo no van más allá de los 2 días, a mediano plazo de 2 a 10 días y a largo plazo mayor a 10 días (Oré, 2019).

La necesidad de datos para las predicciones hidrológicas dependerá de diversos factores como el propósito y el tipo de pronóstico, características hidrológicas de la cuenca, la disponibilidad de datos y la fiabilidad de ellos, por ende, la precisión del pronóstico siempre será sensible.

Además, el entorno económico siempre es una de las más grandes limitaciones (Organización Meteorológica Mundial (OMM), 2011).

2.9.1.1 Métodos de pronóstico de caudales.

Cuando se trata de pronósticos de caudales, históricamente se ha trabajado con modelos lineales estocásticos y determinísticos, ya sea univariados o multivariados. Con la llegada de la computación y sus evidentes avances, los modelos predictivos fundamentados en análisis de datos también han evolucionado. Actualmente existen modelos de pronósticos hidrológicos que permiten mejorar el pronóstico y que incluyen la no linealidad de los datos; es aquí que entran las

redes neuronales e incluyen inteligencia y aprendizaje computacional (Carvajal, Correa, & Muños, 2007).

2.10 Redes neuronales artificiales

Se han creado diversos grupos dedicados al desarrollo de las RNA y se ofrecen varios puntos de vista, donde participan neurólogos, psicólogos, matemáticos, físicos y programadores. Además, algunos investigadores siguen trabajando en redes, como Grossberg que continúa ejerciendo en la Universidad de Boston, mientras Teuvo Kohonen se encuentra en la Universidad de Helsinki, también, Rumelhart, McClelland y Hinton formaron el grupo Parallel Distributed Processing o PDP, que es uno de los grupos de investigación más grande de los últimos años (Basogain, 2007).

2.10.1 Modelo biológico

Las Redes Neuronales Artificiales (RNA) o ANN por su definición en inglés (Artificial Neural Networks) funcionan de forma similar a las neuronas biológicas del cerebro de las personas.

Una neurona está compuesta básicamente por tres elementos: dendritas, cuerpo de la célula o soma y axón. Las dendritas vienen a ser receptoras de la red, cargan de señales eléctricas al cuerpo de la célula, luego, estas señales se suman o procesan en el cuerpo de la célula para después ser transportadas por el axón que es una fibra larga hasta otras neuronas (ver Figura 8) (Serna, 2017).

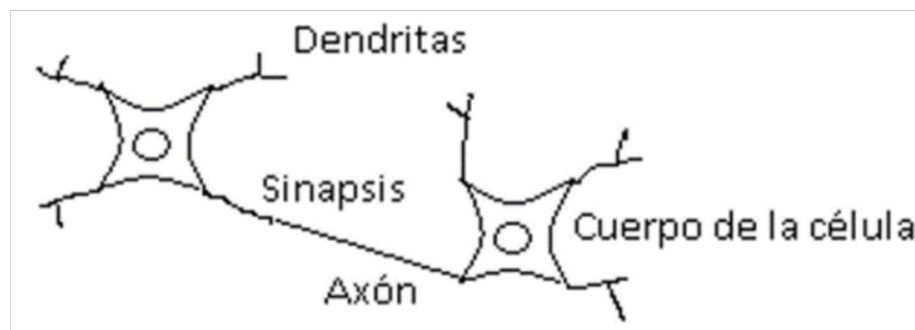


Figura 8. Componentes de una red neuronal.

Fuente: (Serna, 2017).

2.10.2 Analogía de una neurona biológica y Artificial

De cierta manera, las neuronas artificiales son muy similares a las biológicas, en la figura 9 se puede apreciar de forma general una neurona artificial (fig. 9A) y una neurona biológica (fig. 9B).

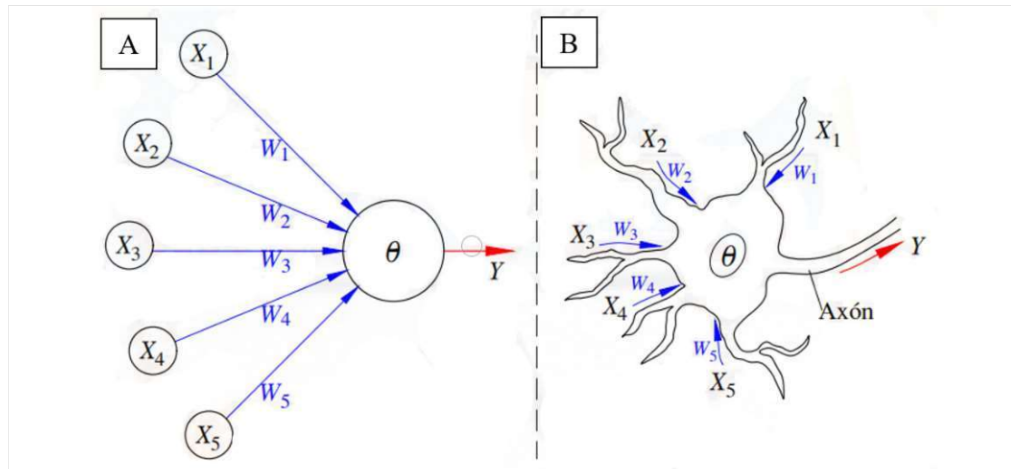


Figura 9. Similitud entre una neurona biológica y una artificial.

Fuente: (Oré, 2019).

Las similitudes halladas entre neuronas artificiales (fig. 9A) y biológicas (fig. 9B) son en las entradas, los pesos sinápticos y la función de activación:

- Las señales que emiten las neuronas son receptoras por las dendritas, estas se representan como entradas X_i .
- La intensidad de conexión entre neuronas se representa como pesos W_i .
- La función de activación theta (Θ) consiste en un umbral que activará la neurona siempre que dicho umbral sea sobrepasado (Oré, 2019).

2.10.3 Definición de una red neuronal artificial

Las redes neuronales artificiales son estructuras flexibles que pueden reconocer las complejas interacciones no lineales entre los datos de las capas de entrada y salida, aun sin tratar de entender la naturaleza de los fenómenos que se están suscitando.

Algunos modelos de pronósticos de caudales trabajan con herramientas basadas en Inteligencia Artificial como Redes Neuronales Artificiales (RNA). Este conjunto de métodos y técnicas que

buscan solucionar un problema que a su vez es difícil y ninguna solución es óptima o satisfactoria se conoce como un modelo heurístico, por lo tanto, las redes neuronales artificiales son modelos heurísticos que buscan emular el funcionamiento del cerebro biológico con el fin de obtener una funcionabilidad similar. En general, una red neuronal consta de capa de entrada, ocultas o intermedias y capa de salida, que a su vez forman un paralelismo entre sí, cada neurona de una capa conecta por medio de pesos sinápticos a la neurona de la siguiente capa (ver figura 10) (Arenas Cárdenas, 2010).

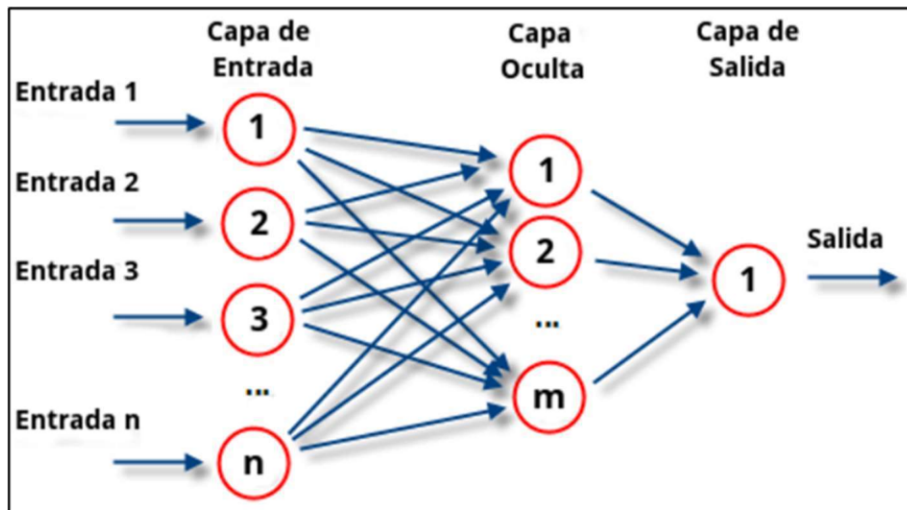


Figura 10. Arquitectura de una red neuronal.

Fuente: Elaboración propia.

2.10.4 Elementos de una red neuronal artificial

De forma general, las redes neuronales constan de los siguientes elementos indispensables: entradas, pesos sinápticos, regla de entrada o propagación, función de activación y función de salida (Castro, 2006).

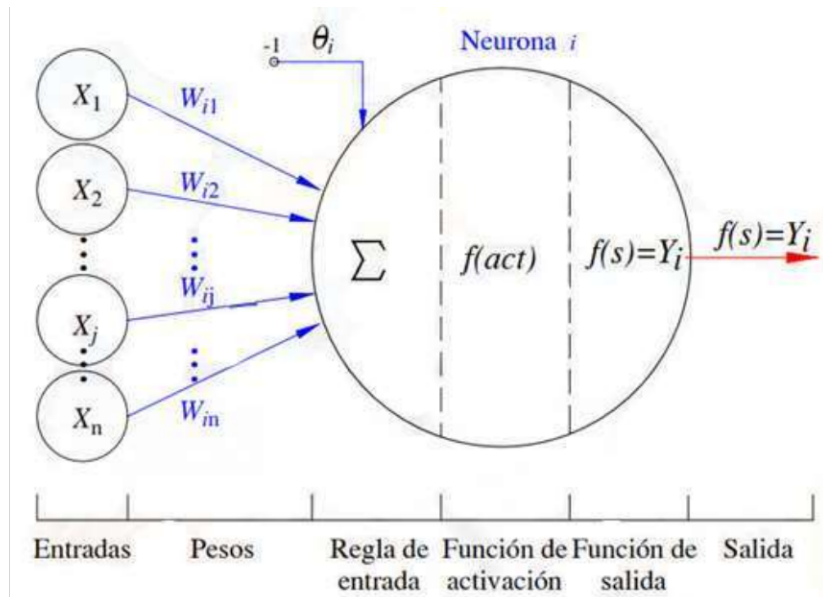


Figura 11. Elementos de una red neuronal artificial.

Fuente: (Oré, 2019).

Entradas

Las entradas X_j son pequeñas señales que salen de una neurona y son recibidas por otra, mediante las dendritas.

Pesos sinápticos

Los pesos sinápticos W_{ij} se pueden considerar como la fuerza de interacción entre la neurona presináptica y la postsináptica.

Regla de entrada o propagación

Debido a su sencillez, la regla más usada es la suma de cada una de las entradas, ponderadas con su respectivo peso sináptico.

Ecuación 2: Regla de entrada o retropropagación.

Fuente: (Oré, 2019).

$$h_i(t) = \sum W_{ij}X_j$$

Función de activación

Restringe la extensión de salida de la neurona. La función de activación busca establecer un valor (generalmente de 0 a 1 o -1 a 1) con el cual la neurona se activará o se mantendrá apagada.

Es importante indicar que (Shamseldin, Nasr, & O'Connor, 2002), demostraron que en la mayoría de los casos las mejores funciones de transferencias eran la funciones sigmoidea y logística, sin embargo, la función arco tangente era la peor, esto basado en los resultados obtenidos.

Función de salida

Esta es la última parte de lo que compone a la neurona, el valor que resulte de esta función será el valor de salida de la neurona y a su vez la entrada de la neurona de la capa siguiente (Oré, 2019).

2.10.5 Arquitectura de las RNA.

Este apartado hace referencia a la colocación y distribución de las capas, las neuronas, y a su vez las conexiones entre capas y neuronas que dependen de cuatro parámetros primordiales: (1) número de capas de la red, (2) el número de neuronas en cada capa, (3) el nivel de conectividad y (4) tipo de conexión.

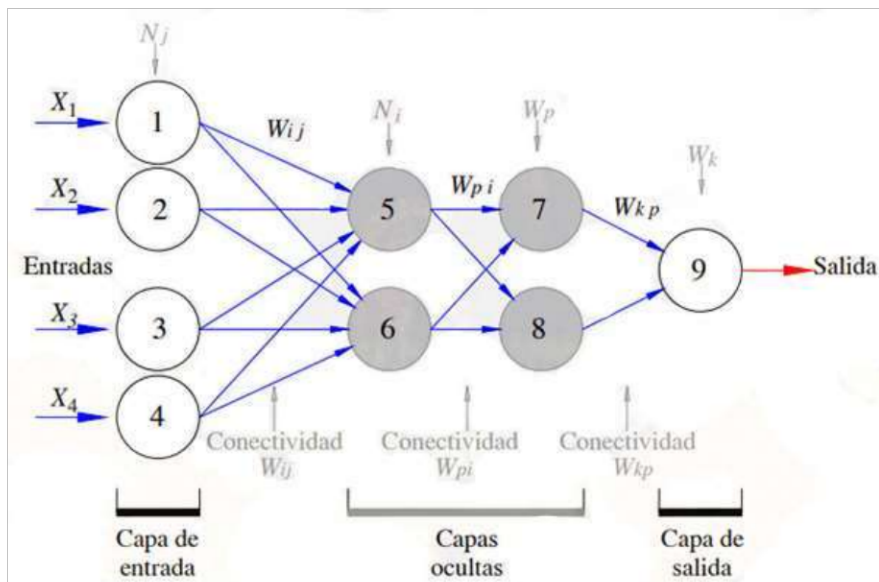


Figura 12. Arquitectura de una red neuronal artificial.

Fuente: (Oré, 2019).

En la figura 12 se presenta una capa de entrada (j) que consta de 4 neuronas, dos capas ocultas (i , p) de dos neuronas cada una y finalmente la salida (k) tiene una neurona. Todas están conectadas por el nivel de conectividad (W_{ij} ; W_{pi} ; W_{kp}) y el tipo de conexión intercapa.

La arquitectura de las RNA puede definirse en función de diferentes razonamientos, pero generalmente se dividen en dos: según su estructura y según el flujo de los datos.

Según su estructura se puede dividir en redes monocapa y redes multicapas. Las redes monocapa solo tienen una capa de neuronas, por lo tanto, solo un nivel de procesamiento hasta la salida de la red. Las redes multicapa cuentan con dos o más niveles de procesamiento y, por ende, dos o más capas ocultas antes de la capa de salida, habiendo conexión integral entre neuronas.

Según el flujo de los datos, las RNA pueden ser feedforward y feedback. Las redes feedforward son redes de propagación hacia delante, por lo tanto, la salida de la neurona no es entrada de un nivel anterior de un mismo nivel, es decir la información fluye en un solo sentido, desde las entradas hasta la salida de la red, pero jamás en sentido contrario (Haykin, 2018). Por otro lado, las RNA feedback son redes que pueden crear conexiones laterales, es decir, conectar la salida a la entrada de otra neurona del mismo nivel y también conexiones con niveles anteriores. Vale aclarar que los sistemas recurrentes son redes feedback que presentan lazos cerrados.

2.10.6 Mecanismo de aprendizaje

Ya establecida la arquitectura de la red, se la debe entrenar con el fin de que “aprenda” el comportamiento que debe presentar ante las entradas que se le va a proporcionar. En función de la información disponible, el mecanismo de aprendizaje se divide en tres grupos:

- Aprendizaje supervisado
- Aprendizaje no supervisado
- Aprendizaje por refuerzo

Aprendizaje supervisado.

En este tipo de aprendizaje de debe contar con la información de los datos de entrada y también los datos a pronosticar, este tipo de aprendizaje se divide en dos categorías:

- Regresión: los valores a predecir generalmente son continuos, un ejemplo es pronosticar el valor de una vivienda en función de su tamaño, número de habitación, baños, ubicación, etc. (Antona, 2017).
- Clasificación: las entradas corresponden a dos o más clases, con este tipo de aprendizaje se busca discernir a que clase corresponden las nuevas entradas, en base a las entradas y salidas que ya conocemos.

Aprendizaje no supervisado.

En este tipo de aprendizaje, el conjunto de entradas no cuenta con información previa para el grupo de salidas. Por lo tanto, se exploran técnicas auto-organizativas y automáticas que buscan revelar la organización de los datos (Antona, 2017).

Aprendizaje por refuerzo.

Este tipo de aprendizaje se enmarca entre los dos anteriores. Se proporciona un conjunto de datos de entrada, la red proporciona una salida y se le indica si esa salida es o no correcta, por ende, no se enseña durante el entrenamiento la salida que se busca conseguir para una respectiva entrada (Haykin, 2018).

2.10.7 Perceptrón multicapa (MLP)

El perceptrón multicapa consta de capa de entrada, muchas capas ocultas y capa de salida, es un tipo de red neuronal y se simboliza con las letras MLP que significan Multi-Layer Perceptron. La mayoría de los problemas cotidianos son del tipo no lineal, sin embargo, mediante el uso del MLP pudieron ser explicados. Es por esta razón que varios autores consideran al perceptrón multicapa como un aproximador universal de funciones (Luna, 2013).

En la figura 12 se muestra la estructura de un perceptrón multicapa:

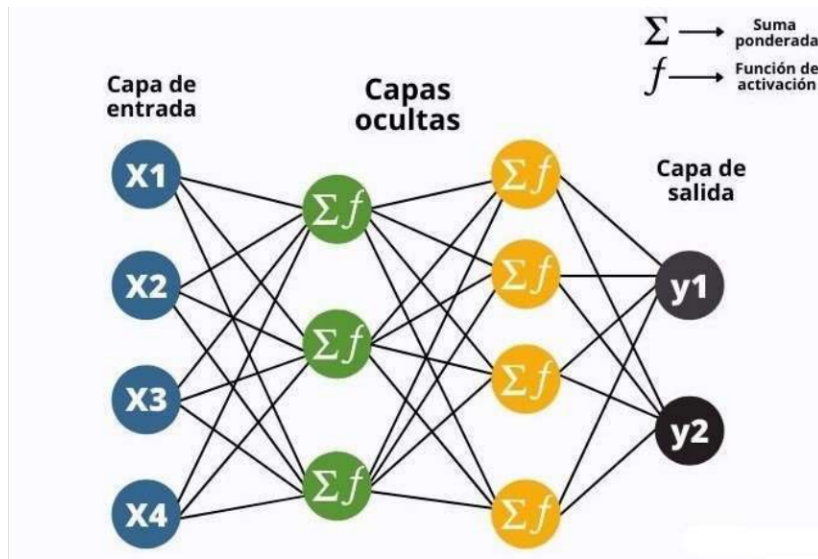


Figura 13. Estructura del perceptrón multicapa.

Fuente: (Oré, 2019).

Se puede observar tres zonas bien marcadas, la capa de entrada, las capas ocultas y la capa de salida. La capa de entrada recibe todas las entradas y se encarga de propagarlas hacia la siguiente capa de procesamiento. Las capas ocultas son las que se encargan de hacer el procesamiento (generalmente no lineal) de los datos que son enviados desde la capa anterior y la capa de salida es la encargada de suministrar los resultados de la red. Este tipo de red neuronal se entrena con el algoritmo de retropropagación de errores o backpropagation.

2.11 Variables de estudio.

2.11.1 Variable independiente.

La variable que buscamos pronosticar y a su vez la más importante es el caudal, producto de la precipitación, escorrentía y demás procesos, sin embargo, producto de la falta de información conlleva a no contar con el registro antes mencionado. Por lo que, la variable independiente o predictora son los niveles de agua medidos en las estaciones hidrológicas aguas arriba del embalse Mazar y que más adelante se mencionan.

2.11.2 Variable dependiente.

La variable dependiente es el pronóstico del caudal de la estación hidrológica Paute, dado que esta estación es muy cercana a la entrada al embalse Mazar. Desde luego, no se puede obtener directamente el caudal, porque la variable independiente es nivel de agua, por lo que se hace uso de la curva de descarga de caudal de la estación Paute.

CAPÍTULO III: METODOLOGÍA DE INVESTIGACIÓN

3.1 Escala Temporal y Espacial

3.1.1 Escala temporal

El presente trabajo de investigación parte de la recopilación de datos hidrometeorológicos y posterior procesamiento de la información recolectada, dicha información fue obtenida del Instituto Nacional de Meteorología en Hidrología (INAMHI), del periodo de junio del año 2014 hasta junio de 2019.

La serie de tiempo que se presenta en la figura 14 corresponde a la estación Paute, ubicada en las coordenadas 2° 45' 31" N y 78° 44' 24" E, cabe aclarar que se cuenta con más datos de esta estación, sin embargo, al construir la base de datos de la red neuronal, no existen registros del mismo día de las estaciones vecinas dentro de la cuenca, y que son vitales para generar el pronóstico, es por ello, que se reduce la base de datos al periodo de agosto de 2017 a junio de 2018.

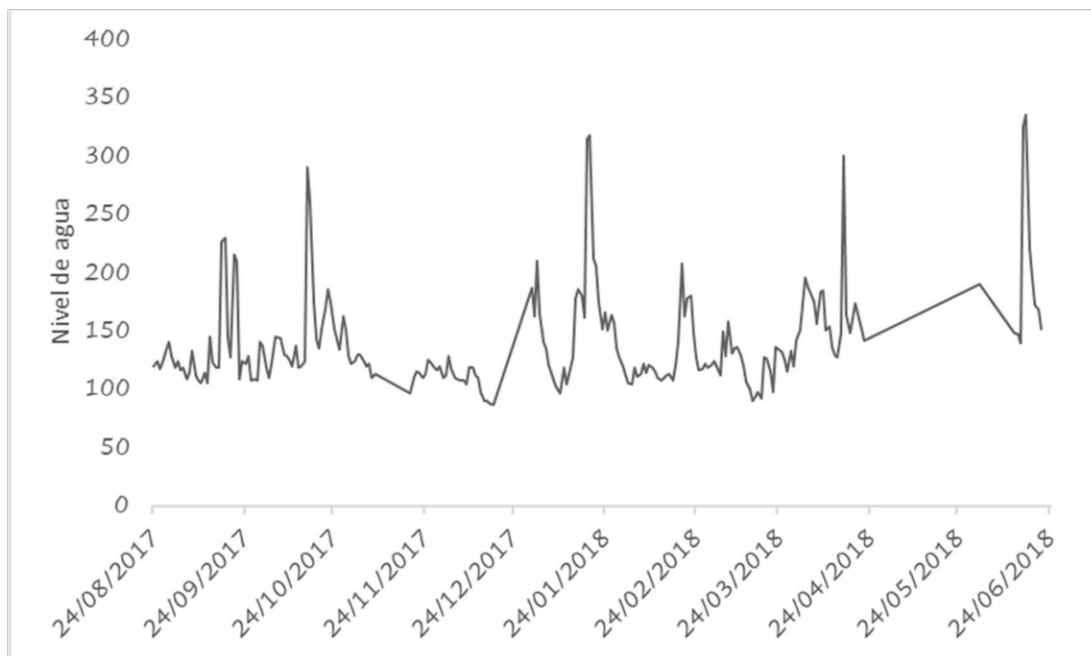


Figura 14: Serie de tiempo de la estación hidrológica Paute.

Fuente: Elaboración propia.

De la serie de tiempo antes presentada se evidencia que la mínima altura del nivel de agua registrada se da el 17 de diciembre de 2017 con 0.87 m y el mayor registro se da el 16 de junio de

2018 con 3.35 m. La media de la altura del nivel del agua de la serie de tiempo de la figura 14 se evidencia en 139.67 m.

3.1.2 Escala espacial

La cuenca que conforma el Río Paute tiene gran importancia estratégica nacional (Ministerio del Ambiente, 2009), ya que aquí se encuentra una de las centrales hidroeléctricas más grandes del Ecuador (Molino, 1100 MW), además, la hidroeléctrica Mazar con capacidad de 170 MW y Sopladora con 487 MW.

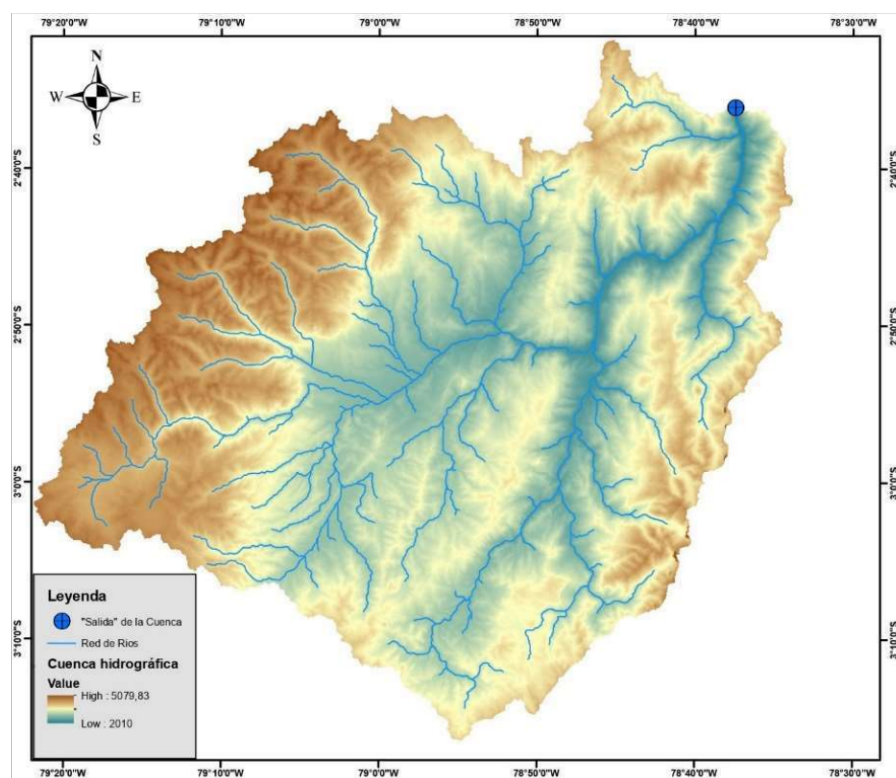


Figura 15. Cuenca hidrográfica que aporta caudal al Embalse Mazar.

Fuente: Elaboración propia.

La cuenca del Río Paute tiene una superficie aproximada de 6 500 km², sin embargo, en este trabajo teórico-práctico solo se toma en cuenta la superficie de aporte de caudal al Embalse Mazar, cuya superficie calculada es de 4 228.43 km² (Figura 15).

En la figura 15 se muestra de forma gráfica que mientras más se acerca al punto de interés, el río principal va creciendo y haciéndose más evidente debido al aporte de los afluentes.

3.2 Análisis de información.

Los procedimientos y resultados se presentan en los siguientes puntos de este capítulo, a fin de llegar a cumplir con el objetivo planteado en el capítulo 1.3. Para una mejor comprensión se hace uso de tablas y gráficos construidos en software y lenguaje de programación informático como Excel, R, ArcGIS, entre otros.

3.2.1 Área de estudio.

El área de estudio está dividida en 13 microcuencas (Fig. 16), presentan relieve montañoso en las zonas más altas y la cuenca media y baja presenta un relieve escarpado, se encuentra comprendida entre las coordenadas $2^{\circ} 31' 34''$ sur, $79^{\circ} 22' 16''$ oeste y $3^{\circ} 17' 23''$ sur, $78^{\circ} 32' 28''$ oeste, y contempla parte de las provincias de Azuay y Cañar al sur del Ecuador.



Figura 16. Área y subcuencas de estudio.

Fuente: Elaboración propia.

La cuenca del Río Paute se encuentra ubicado sobre la Cordillera de los Andes, debido a esto, sus características se ven sesgadas ya que la morfología y ubicación geográfica tienen influencia sobre los factores meteorológicos:

Durante todo el año, especialmente en la cuenca baja, desde la región amazónica llegan masas de aire húmedo que causan lluvias especialmente entre junio y agosto lo cual produce máximas de precipitación y mínimas entre noviembre y febrero (Cordero, 2013).

Al occidente de la cuenca alta llegan masas de aire provenientes del océano pacífico, lo cual provoca dos estaciones lluviosas, la primera entre febrero y mayo, y la segunda entre octubre y noviembre (Cordero, 2013).

El Anticiclón del Pacífico Sur, el fenómeno de Niño y la Niña también afectan significativamente la cuenca, en especial a la cuenca media (Cordero, 2013).

En función de la altitud y la topografía, la cuenca se puede dividir en tres sectores:

Tabla 3. Caracterización general de la cuenca hidrográfica.

Fuente: (Cordero, 2013).

	Clima	Altitud	Precipitación anual	Temperatura media anual	Humedad relativa media anual	Uso de suelo
Cuenca Alta	Frío semi-húmedo	2600 – 5080 msnm	1000 – 2000 mm/año	8 °C	88%	Vegetación de páramo, cereales y hortalizas
Cuenca Media	Templado semi-húmedo	2200 – 2600 msnm	500 - 1000 mm/año	15 °C	84%	Pastos, caña de azúcar, frutales y flores
Cuenca baja.	Mesotérmico húmedo	500 – 2200 msnm	> 2000 mm/año	20 °C	91%	Bosque natural y agricultura

A continuación, se presenta en la tabla 4 el resumen de las estaciones hidrológicas que fueron tomadas en cuenta para esta investigación, debido a su importancia regional y disponibilidad de los datos.

Tabla 4. Ubicación de estaciones hidrológicas.

Fuente: Etapa EP.

Nombre	Latitud	Longitud
---------------	----------------	-----------------

YanuncayAjTarquiL	722104	9677886
MataderoSayausiPL	714889	9682022
TomebambaQtaBalzayL	718377	9680325
MachangaraLLanteraL	725473	9681510
TarquiAjYanuncayL	721493	9677789
TomebambaPTARL	728627	9682487
GualaceoDjPamarPL	743850	9668469

3.2.1.1 Parámetros Morfométricos

En Ecuador, el Instituto Geofísico Militar está en la obligación de proveer información cartográfica de libre acceso para temas prácticos y/o investigativos, a partir de esta información primaria se generó el modelo digital de terreno o modelo digital de elevaciones que es imprescindible para obtener los parámetros morfométricos de la cuenca. Alguno de los parámetros morfométricos de la cuenca que se van a calcular son área, perímetro, longitud, razón de circularidad factor de forma, entre otros (Dolores & Agüero, 2019).

En la Tabla 5 se resumen los resultados de cada parámetro morfométrico de la cuenca calculado:

Tabla 5. Parámetros morfométricos de la cuenca.

Fuente: Elaboración propia.

Parámetros relativos a la forma de la cuenca:		
Parámetro	Modo de medición	Resultado
Área (A) km ²	Digital	4228.434
Perímetro (P) km	Digital	386.914
Longitud (L) km	Digital	97.93
Ancho (W) km	W = A/L	43.178
Razón de Circularidad de Miller (M) km ⁻¹	M = P/A	0.092

Factor de Forma de Horton	$K = A/L^2$	0.441
---------------------------	-------------	-------

Parámetros relativos al relieve de la cuenca:

Parámetro	Modo de medición	Resultado
Desnivel altitudinal de cuenca (G) m	$G = H_{max_c} - H_{min_c}$	3069.834
Pendiente media de cuenca (S) m	$S = G/L$	31.347

Parámetros relativos al perfil de la cuenca:

Parámetro	Modo de medición	Resultado
Longitud del cauce principal (Y) m	Digital	123176.70
Desnivel del cauce principal (V) m	$V = H_{max_r} - H_{min_r}$	1580
Pendiente media del cauce principal (J) %	$J = (V/Y) * 100\%$	1.3
Centro de gravedad	X	730432.469 m E
Centro de gravedad	Y	9679695.878 m S

Parámetros relativos al sistema de drenaje:

Parámetro	Modo de medición	Resultado
Red de drenaje (RD)	Digital	Ver figura 18
Orden de corrientes (O)	Digital	Ver figura 19
Longitud de los cauces de agua (C) km	Digital	990.236
Densidad de drenaje (Z) km^{-1}	$Z = C/A$	0.234
Frecuencia de drenaje (F) km^{-2}	$F = N/A$	0.055

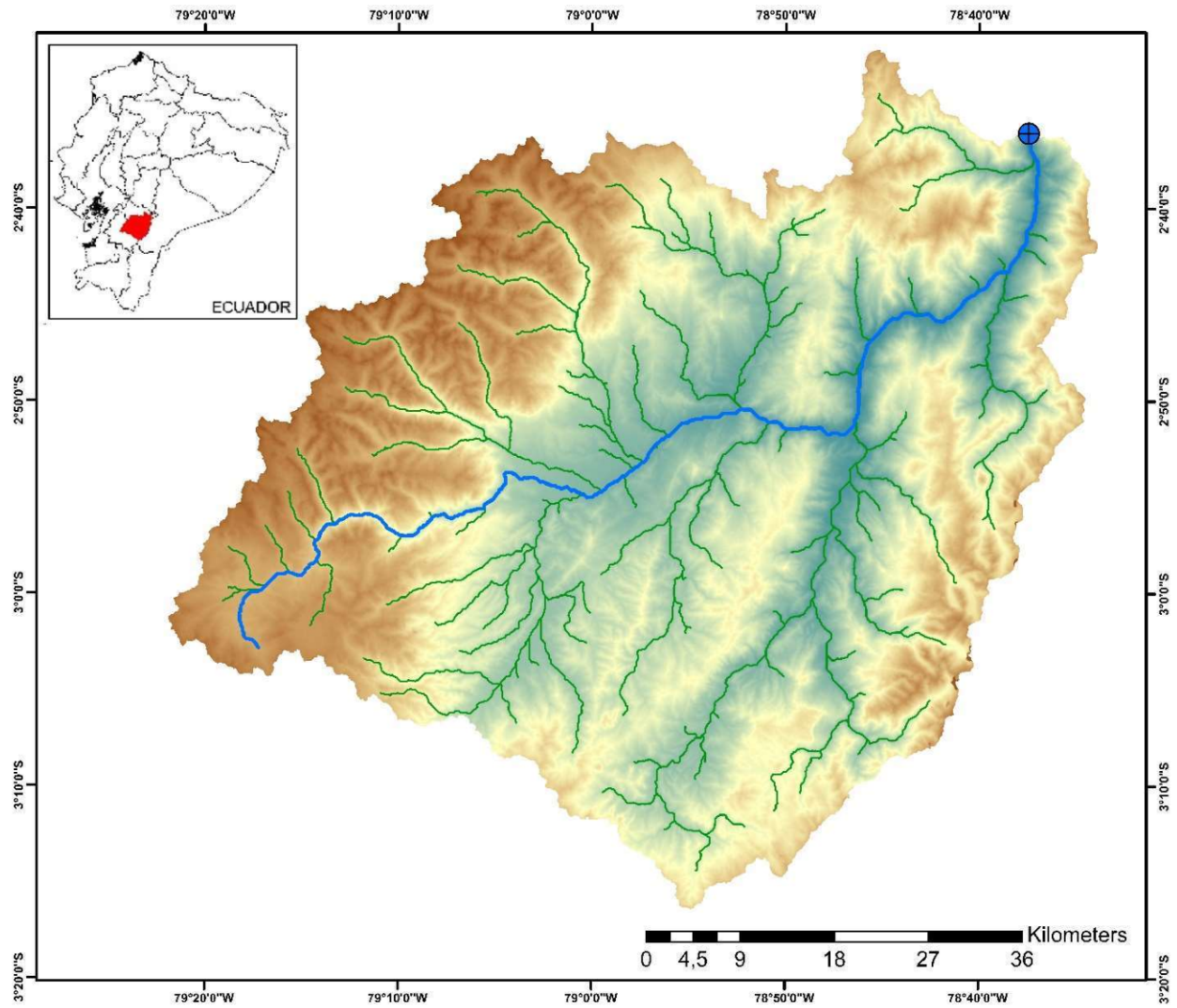


Figura 17. Red de drenaje.

Fuente: Elaboración propia.

La red de drenaje (ver fig. 17) es bastante extensa y ramificada, se puede observar claramente el cauce principal y los afluentes que lo alimentan aguas abajo. En la figura 18 se puede apreciar el orden de los ríos, siendo el 5 el mayor de todo, por lo tanto, lo catalogaremos como una cuenca hidrográfica con orden de corriente 5.

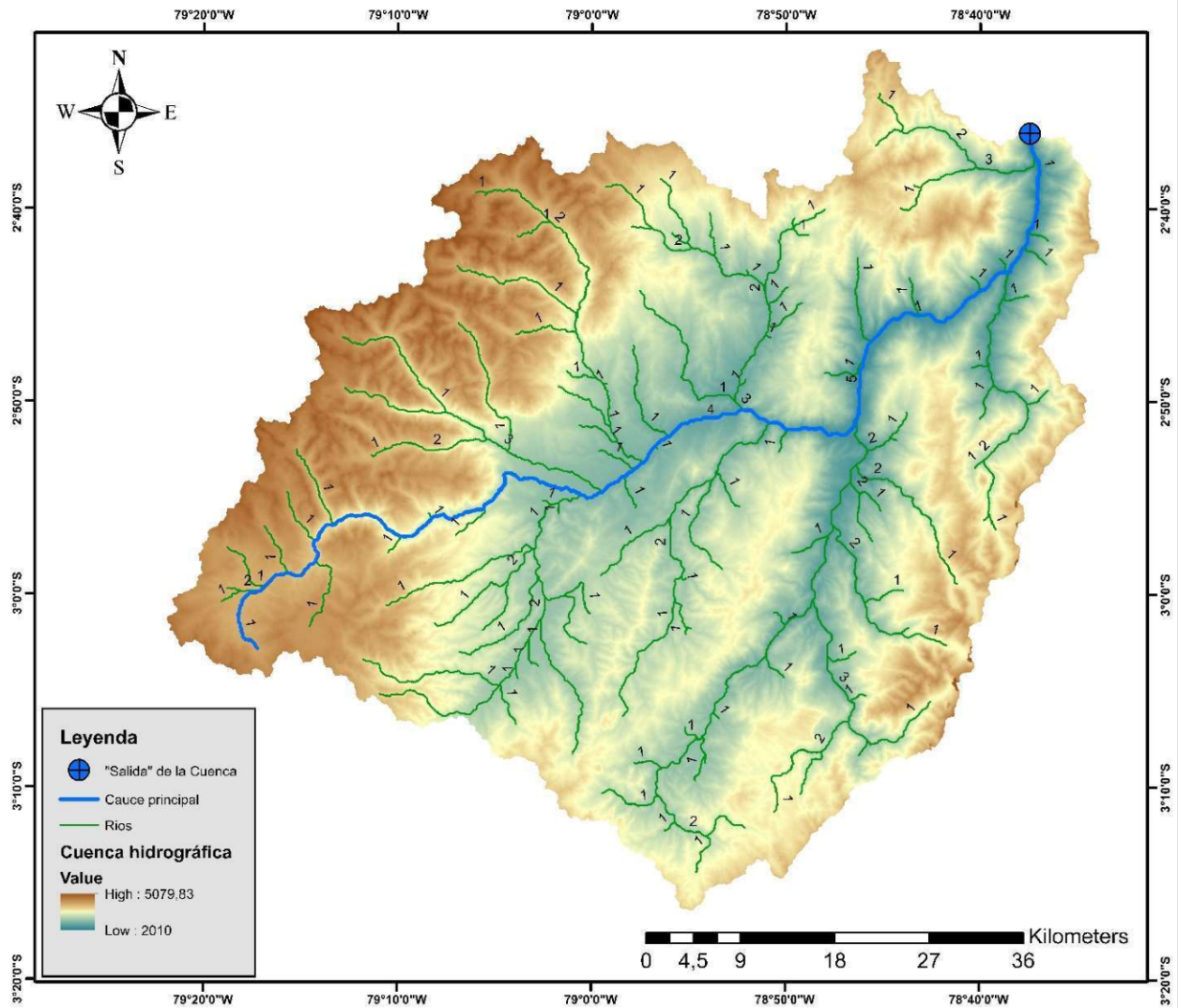


Figura 18. Orden de corrientes.

Fuente: Elaboración propia.

3.2.2 Información hidrológica.

Las estaciones hidrológicas proveen registros diarios de nivel de río o altura del agua, especialmente de los ríos Tarqui, Yanuncay, Tomebamba, Machángara y Cuchipamba, cuyas estimaciones del nivel de río nos ayudaran como variable principal dentro del modelo de red neuronal.

Tabla 6. Resumen estadístico de las estaciones hidrometeorológicas.

Fuente: Etapa EP.

	Yanuncay Aj TarquiL	Matadero Sayausi PL	Tomebamba Qta Balzay L	Machangara Llantera L	Tarqui Aj YanuncayL	Tomebamba PTARL	Gualaceo Dj PamarPL
Latitud	722104	714889	718377	725473	721493	728627	743850
Longitud	9677886	9682022	9680325	9681510	9677789	9682487	9668469
Promedio de nivel (cm)	100.81	78.68	62.9	63.27	90.97	135.05	89.26
Mínimo (cm)	79.78	35.93	35.93	9.91	2.66	99.3	59.31
Máximo (cm)	167.9	168.13	138.32	111.36	579.65	220	256.1

En el siguiente grafico se evalúa la distribución de las variables continuas:

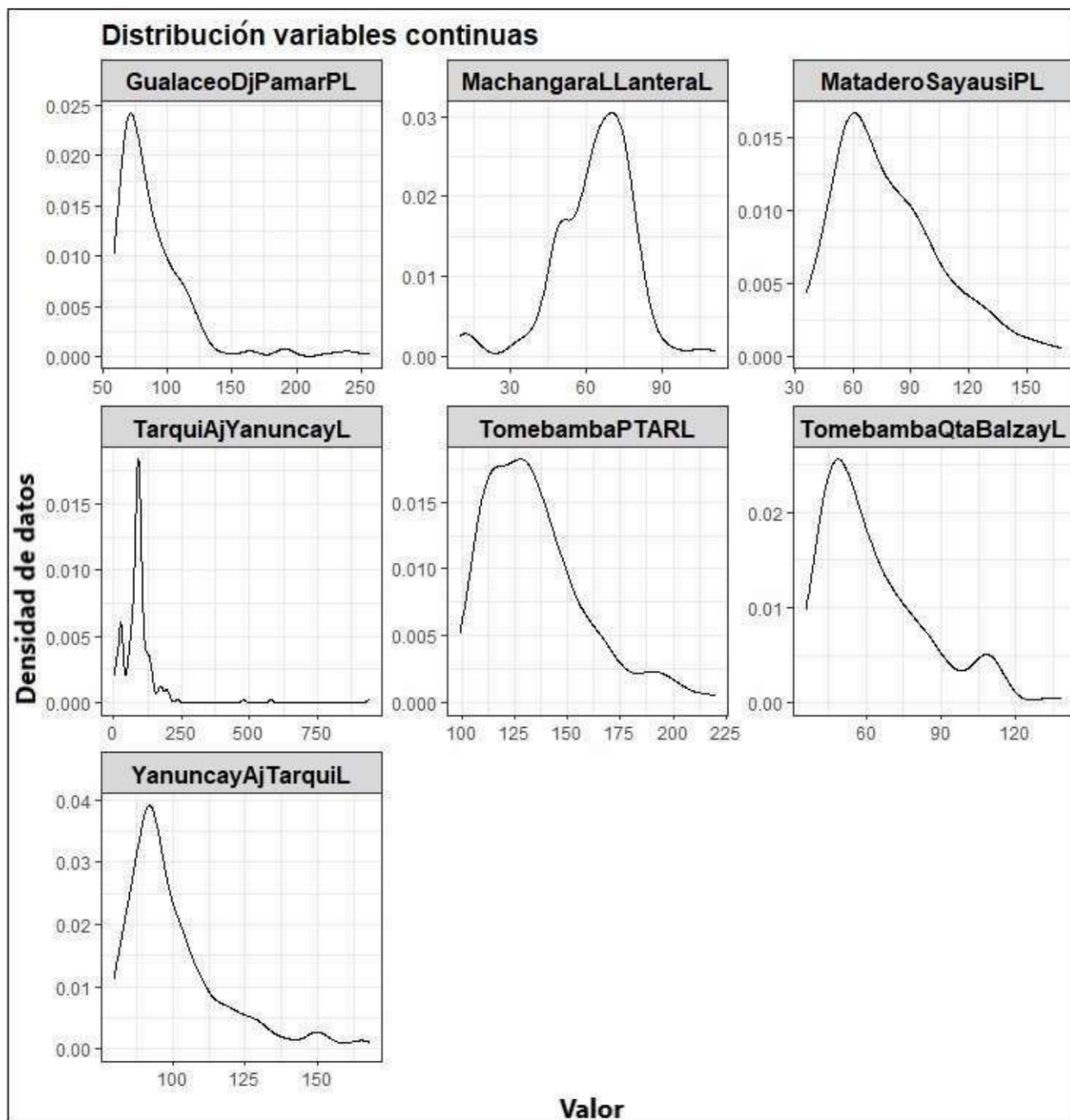


Figura 19. Distribución de variables continuas.

Fuente: Elaboración propia.

En la figura 20, se muestra la serie de tiempo que provienen de los registros de la estación Yanuncay Tarqui 1:

Tabla 7. Ubicación de la estación YanuncayAjTarquiL.

Fuente: Etapa EP.

Estación	Latitud	Longitud
YanuncayAjTarquiL	722104	9677886

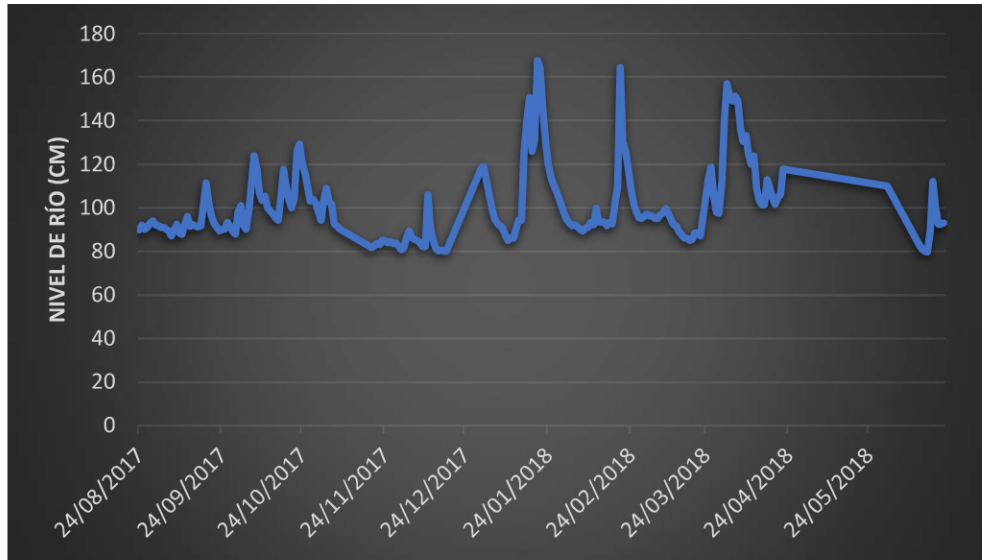


Figura 20. Serie de tiempo de la estación YanuncayAjTarquiL.

Fuente: Etapa EP.

En la figura 21, se muestra la serie de tiempo que provienen de los registros de la estación MataderoSayausiPL:

Tabla 8. Ubicación de la estación MataderoSayausiPL.

Fuente: Etapa EP.

Estación	Latitud	Longitud
MataderoSayausiPL	714889	9682022

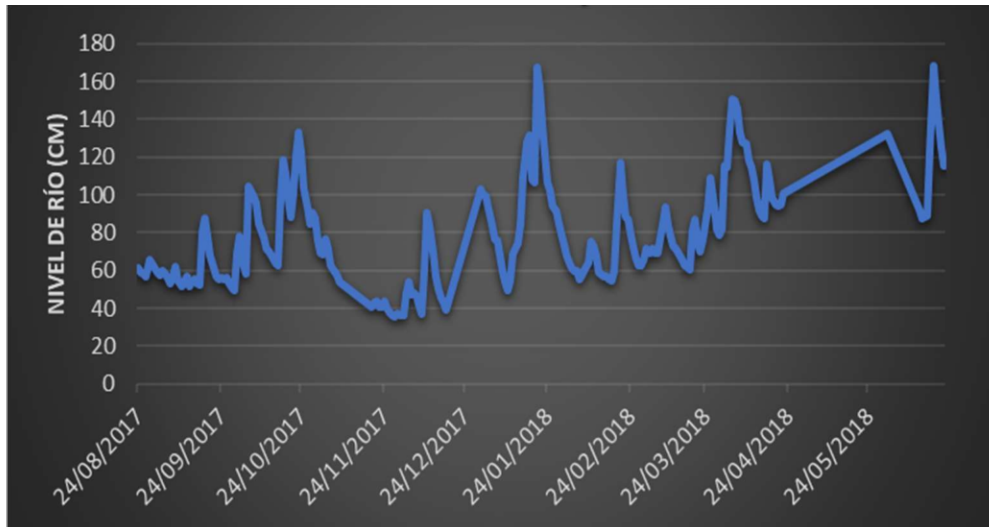


Figura 21. Serie de tiempo de la estación MataderoSayausiPL.

Fuente: Etapa EP.

En la figura 22, se muestra la serie de tiempo que provienen de los registros de la estación TomebambaQtaBalzayL:

Tabla 9. Ubicación de la estación TomebambaQtaBalzayL.

Fuente: Etapa EP.

Estación	Latitud	Longitud
TomebambaQtaBalzayL	718377	9680325

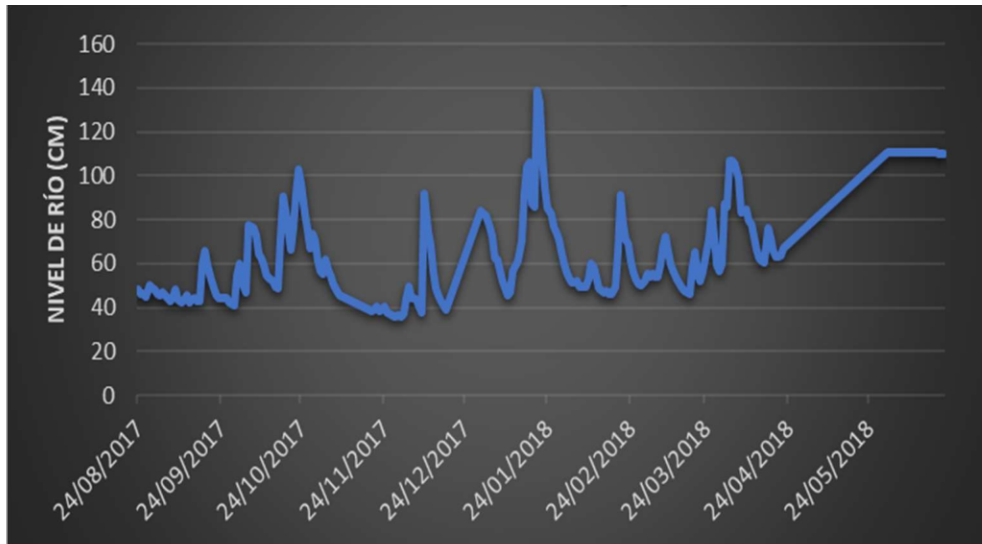


Figura 22. Serie de tiempo de la estación TomebambaQtaBalzayL.

Fuente: Etapa EP.

En la figura 23, se muestra la serie de tiempo que provienen de los registros de la estación MachangaraLLanteraL:

Tabla 10. Ubicación de la estación MachangaraLLanteraL.

Fuente: Etapa EP.

Estación	Latitud	Longitud
MachangaraLLanteraL	725473	9681510

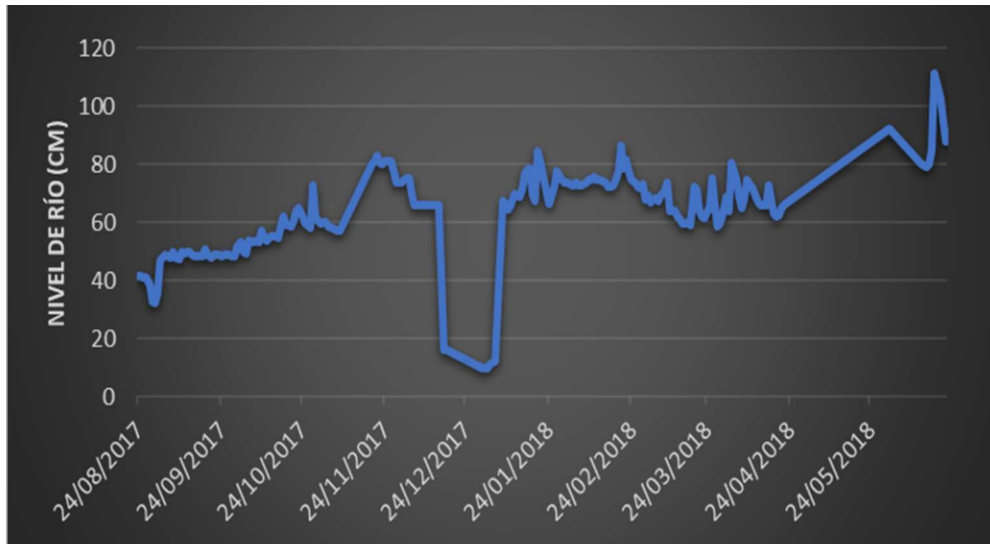


Figura 23. Serie de tiempo de la estación MachangaraLLanteraL.

Fuente: Etapa EP.

En la figura 24, se muestra la serie de tiempo que provienen de los registros de la estación TarquiAjYanuncayL:

Tabla 11. Ubicación de la estación TarquiAjYanuncayL.

Fuente: Etapa EP.

Estación	Latitud	Longitud
TarquiAjYanuncayL	721493	9677789

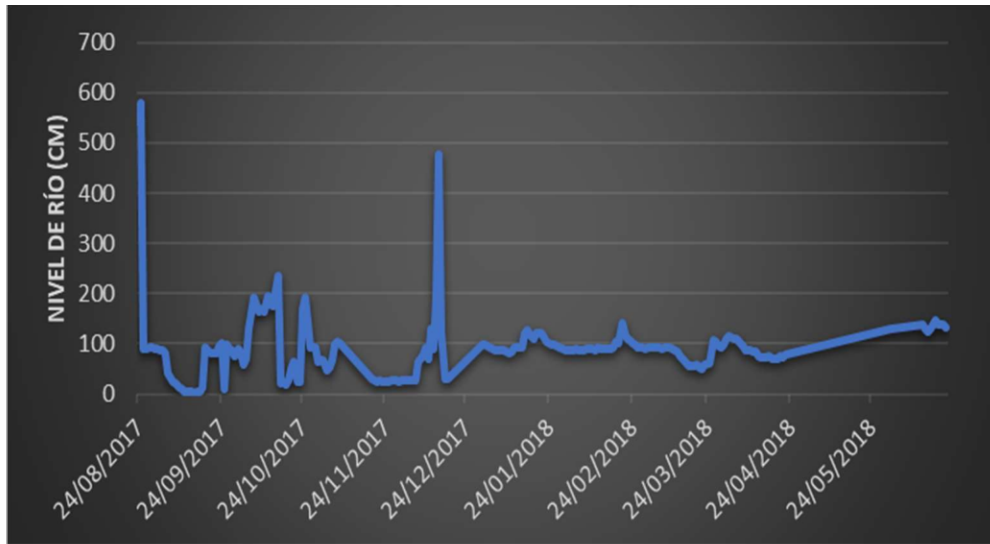


Figura 24. Serie de tiempo de la estación TarquiAjYanuncayL.

Fuente: Etapa EP.

En la figura 2, se muestra la serie de tiempo que provienen de los registros de la estación TomebambaPTARL:

Tabla 12. Ubicación de la estación TomebambaPTARL.

Fuente: Etapa EP.

Estación	Latitud	Longitud
TomebambaPTARL	728627	9682487

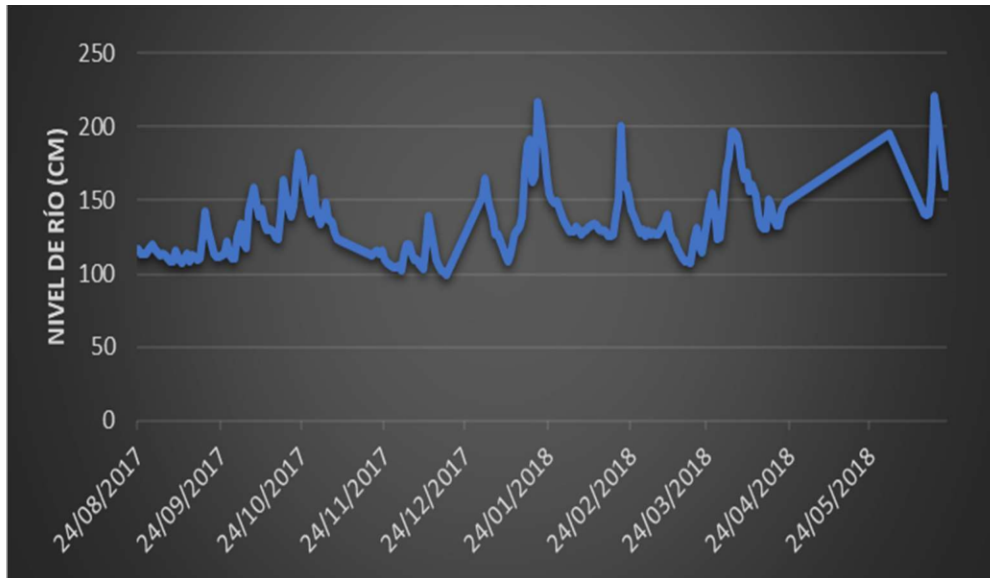


Figura 25. Serie de tiempo de la estación TomebambaPTARL.

Fuente: Etapa EP.

En la figura 26, se muestra la serie de tiempo que provienen de los registros de la estación GualaceoDjPamarPL:

Tabla 13. Ubicación de la estación GualaceoDjPamarPL.

Fuente: Etapa EP.

Estación	Latitud	Longitud
GualaceoDjPamarPL	743850	9668469

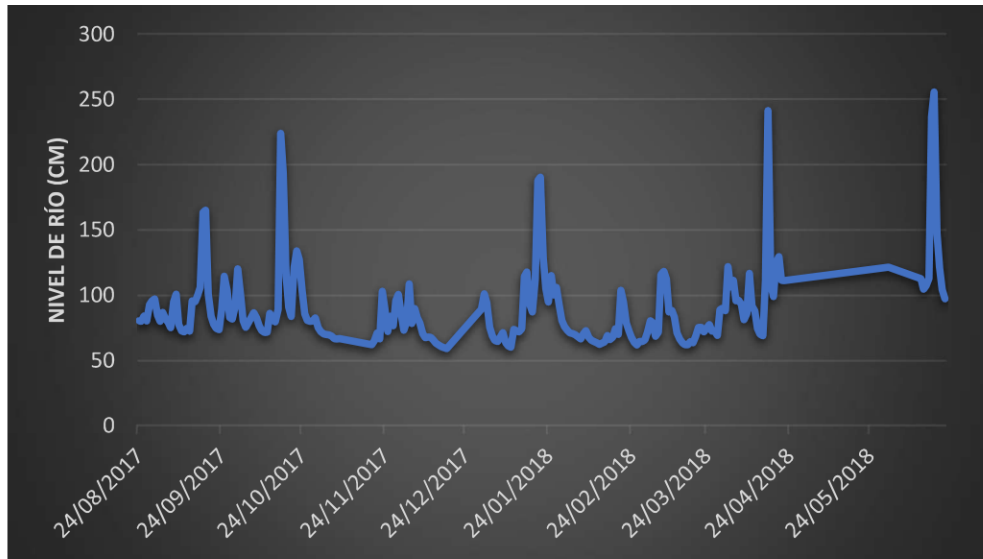


Figura 26. Serie de tiempo de la estación Gualaceo GualaceoDjPamarPL.

Fuente: Etapa EP.

3.2.3 Estimación de la curva de descarga

Para estimar la curva de descarga existen varios métodos, pero en esta investigación se hace uso del software R y la librería “minpack.lm” que permite obtener ecuaciones no lineales y obtener la curva de caudal simple (Ramírez, Carvajal, Bocanegra, & Sandoval, 2018). Para determinar la curva de descarga se hace uso de una base de datos que contine el nivel de rio y aforo de caudal de la estación hidrológica Paute desde junio de 2000 hasta diciembre de 2015 (Ver tabla 14).

Tabla 14. Ubicación de la estación Paute.

Fuente: Etapa EP.

Estación	Latitud	Longitud	Elevación (msnm)
Paute en Paute (DJ Gualaceo)	751257,2	9694849,6	2104

Tabla 15. Estructura de los datos de ingreso para determinar la curva de descarga.

Fuente: INAMHI.

Fecha	Nivel (m)	Q (m ³ /s)
16/06/2000	1.72	44.015
17/06/2000	2.49	125.526
18/06/2000	2.67	142.58
19/06/2000	2.16	83.48
20/06/2000	1.99	66.597
21/06/2000	2.09	77.216
22/06/2000	2.72	150.839
23/06/2000	2.54	126.621
24/06/2000	2.17	85.052
25/06/2000	2	68.171
26/06/2000	1.99	67.4
27/06/2000	2.38	111.664
28/06/2000	2.44	116.915
29/06/2000	2.16	84.103
30/06/2000	2.13	81.181

Para calcular la curva de descarga se hace uso de la ecuación 1, que se describió en el apartado 2.8.6 y del Anexo “Código para obtener los parámetros de la curva de calibración de caudales”.

Luego de aplicar el código anterior se obtienen los parámetros H_0 , k y n que se muestran en la tabla 16:

Tabla 16. Parámetros para definir la ecuación de curva de descarga.

Fuente: Elaboración propia.

Parámetro	Valor
H_0	0.763864
k	46.606474

n

1.717117

La ecuación de la curva de descarga nos permite transformar los niveles de río a caudal, la ecuación generalizada queda de la siguiente manera:

Ecuación 3: Ecuación de curva de descarga del modelo.

Fuente: Elaboración propia.

$$Q = 46.606474 * (H - 0.763864)^{1.717117}$$

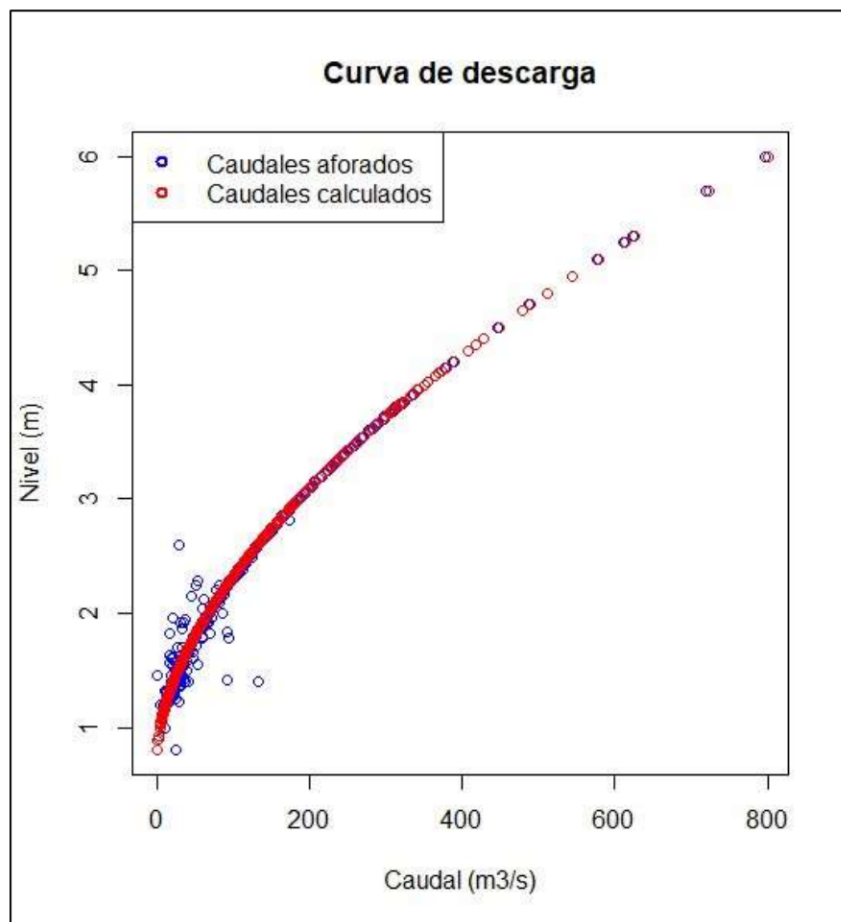


Figura 27. Curva de descarga.

Fuente: Elaboración propia.

En la figura 27 se muestra el caudal aforado y caudal calculado donde se puede observar el ajuste de los datos después de usar la ecuación de la curva de descarga, además, también se calculó el coeficiente de correlación R que nos da una medida intuitiva de que tan bien se ajusta el caudal calculado al caudal aforado.

$$R = 0.998502$$

3.2.4 Predicción de niveles mediante redes neuronales.

El objetivo de esta investigación es realizar el pronóstico de caudales de ingreso al embalse Mazar, debido a que no se cuenta con datos de caudal en las estaciones aguas arriba, se optó por pronosticar niveles y luego transformarlos a caudales mediante la curva de descarga. Por lo tanto, la primera parada de la investigación es pronosticar los niveles de río que se presentan al ingreso del embalse Mazar.

3.2.4.1 Selección de datos de ingreso

En el apartado 4.1.3 se especificaron las estaciones hidrológicas que forman parte de la data que servirán como variables de ingreso a la red neuronal, además, también consta la variable que se busca pronosticar (niveles en la estación Paute en Paute). En la tabla 17 se presenta la estructura de los datos de ingreso del modelo y no se coloca la fecha para no generar una confusión dentro de la red neuronal.

Tabla 17. Estructura de los datos de ingreso a la red neuronal.

Fuente: Etapa EP.

Yanuncay Aj Tarqui L (cm)	Matadero Sayausi PL (cm)	Tomebamba Qta Balzay L (cm)	Machangara Llantera L (cm)	Tarqui Aj Yanuncay L (cm)	Tomebamba PTARL (cm)	Gualaceo Dj PamarPL (cm)	Nivel Estación Paute (cm)
89.81136	61.09452	47.89163	41.44691	941.2202	116.6	80.5	120

92.03187	58.70042	46.2682	41.50793	579.6457	113.8	80	124
90.20307	58.2067	45.91819	41.31524	89.92062	113.7	84.2	117
91.10789	56.97771	44.76317	41.18425	90.50838	113.4	80.4	123
92.91611	65.58868	50.11893	38.78531	92.73712	117.4	92.8	134
94.05383	63.08563	48.57261	32.67458	92.94611	120.2	95.9	140
92.23344	61.18069	47.94634	32.21975	90.84936	117.2	97.2	128
91.89617	58.67913	46.45367	36.0056	89.06114	114.6	84.8	118
90.98981	57.70163	45.57159	47.18933	87.97279	113.2	80.1	124
91.00433	59.5979	46.68775	48.00071	88.09338	113.8	86.9	116
90.31583	58.14233	45.37106	49.01287	80.96861	112.5	81.8	119
89.53402	56.28582	44.17864	48.56001	41.40256	111.2	79.42	109
87.10898	53.39045	42.85604	47.82656	32.94394	108.6	75.11	114
89.52873	56.53816	44.54349	49.91663	23.234	108.8	95	133
92.50565	61.70224	48.10027	47.84366	22.4369	115.5	100.9	112
88.52354	54.29119	43.5929	47.46	12.19616	109.6	77.59	108
87.76101	51.85405	42.16859	49.72882	10.60069	108	72.7	105
92.45802	54.20839	43.47572	49.57457	5.092768	110.3	72.02	114
95.98975	56.50436	45.85008	49.76445	5.85595	114.2	74.77	105

Los datos en la tabla 17 están contemplados en centímetros.

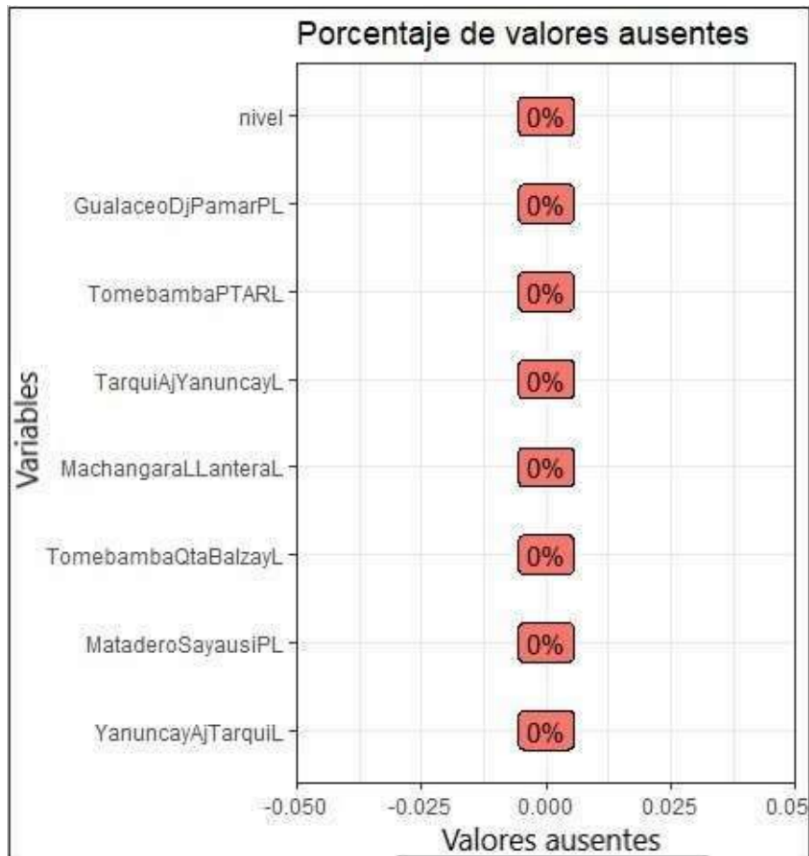


Figura 28. Porcentaje de datos ausentes de cada variable.

Fuente: Elaboración propia.

No existen datos ausentes ya que previamente fueron trabajados por el investigador a fin de evitar cualquier error de procesamiento en el modelo.

No hay ningún fundamento teórico que pueda sustentar el porcentaje de datos necesarios para el entrenamiento y prueba del modelo, sin embargo, algunos autores coinciden que es recomendable trabajar con el 80% de los datos para entrenamiento y un 20% para test, otra forma interesantes es trabajar con el 75% de datos para entrenamiento, 20% para validación y un 5% para prueba (Oré, 2019). En esta investigación abordaremos solo las etapas entrenamiento y test, con 80% y 20% de registros respectivamente, debido a la disponibilidad de los datos.

CAPÍTULO IV: PRESENTACIÓN DE RESULTADOS

4.1 Resultados del modelo en la estación Paute.

Las variables de la capa de entrada corresponden a las siete estaciones hidrológicas que se tomaron en cuenta, 1 capa oculta con 50 neuronas y 1 capa de salida con 1 neurona, por lo tanto, el modelo final queda definido como 7 – 50 – 1. Dado que el código se encarga de tomar los datos para entrenamiento y test de forma autónoma, los resultados de error varían, pero todos los modelos están por debajo del 10%, en términos de error porcentual medio.

Tabla 18. Resultados de los pronósticos con el modelo de red neuronal planteado.

Fuente: Etapa EP y elaboración propia.

Nro de Obs.	Estimado	Observado	Nro de Obs.	Estimado	Observado
1	143.0824654	120	24	111.8816361	140
2	123.5562973	134	25	115.196627	116
3	111.6123498	108	26	111.1537366	121
4	123.9901994	123	27	112.9127118	124
5	125.5382414	118	28	113.8843864	120
6	112.4986935	107	29	121.5552414	136
7	122.8765282	120	30	118.5899991	128
8	188.0510499	175	31	111.2209399	106
9	155.43745	152	32	104.8396534	90
10	108.3637401	110	33	103.181	93
11	108.3883157	114	34	103.1348724	98
12	108.6208052	112	35	120.0296559	127
13	116.4665339	119	36	157.7184029	150
14	107.5265878	87	37	161.3334351	172
15	155.1890629	165	38	155.8026213	175
16	105.4745244	119	39	154.518789	156
17	115.2585429	127	40	151.6836105	154

18	164.0892899	161	41	150.7287324	136
19	173.5398118	205	42	120.5030755	148
20	153.9513545	152	43	130.9873727	160
21	145.0027823	157	44	161.358012	148
22	110.4537333	105	45	157.7848228	139
23	111.0666359	110	46	167.1666485	168

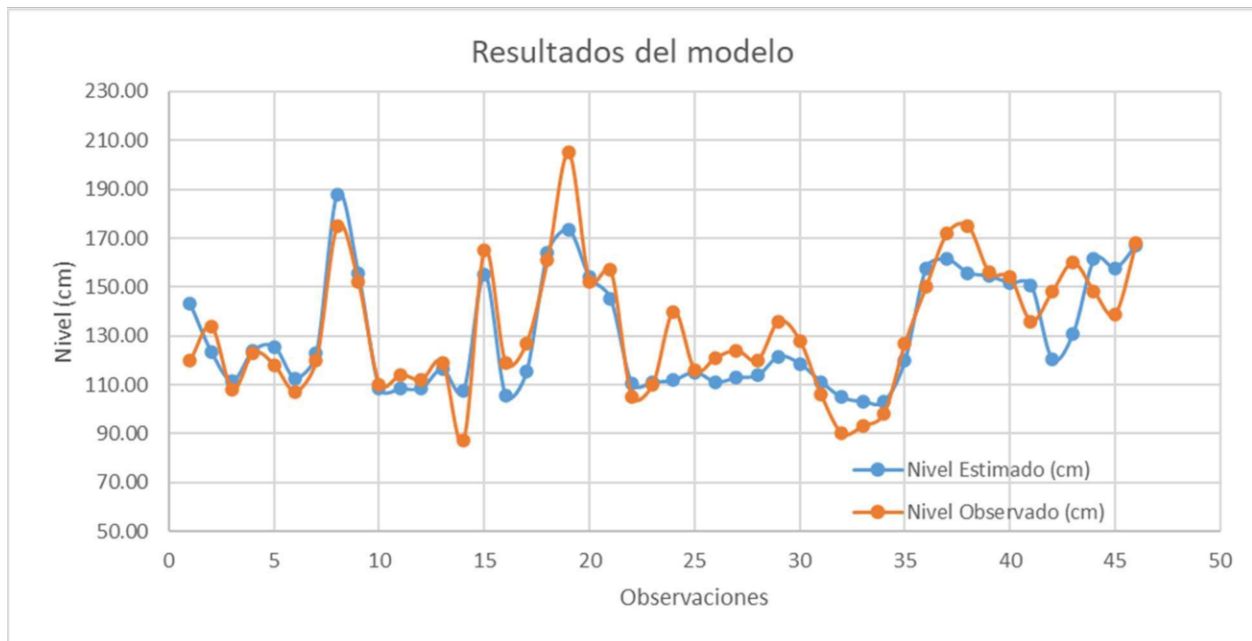


Figura 29: Resultados del modelo.

Fuente: Elaboración propia.

Con esta configuración se obtiene un PEMA de 7.57% y RMSE de 12.91, lo cual representa un muy buen ajuste para la red neuronal. Dado que no existe ninguna metodología que sustente el porqué de usar cierta configuración de capas, neuronas, funciones de activación y parámetros a introducir, queda evidenciado que toda red neuronal es diferente y la precisión de esta dependerá de la experiencia del investigador.

4.2 Estadísticos de los datos observados y pronosticado.

4.2.1 Varianza y desviación estándar de los datos empleados y pronosticados.

La incertidumbre siempre aparece cuando se tratan magnitudes, comúnmente se la conoce como la medida cuantitativa del resultado, es muy importante ya que nos permite realizar comparaciones entre las variables de estudio. A continuación, se presenta los datos estadísticos que nos facilitaran la comparación entre los datos observados y pronosticados de la estación hidrológica Paute:

Media de los datos observados: $\chi_m = 132.72 \text{ cm}$

Tabla 19: Estadísticos de datos observados.

Fuente: Elaboración propia.

Nivel Observado (cm)	$\chi_m - X$	$(\chi_m - X)^2$
120	-12.72	161.8
134	1.28	1.64
108	-24.72	611.08
123	-9.72	94.48
118	-14.72	216.68
107	-25.72	661.52
120	-12.72	161.8
175	42.28	1787.6
152	19.28	371.72
110	-22.72	516.2
114	-18.72	350.44
112	-20.72	429.32
119	-13.72	188.24
87	-45.72	2090.32
165	32.28	1042
119	-13.72	188.24
127	-5.72	32.72
161	28.28	799.76
205	72.28	5224.4
152	19.28	371.72
157	24.28	589.52
105	-27.72	768.4
110	-22.72	516.2
140	7.28	53
116	-16.72	279.56
121	-11.72	137.36

124	-8.72	76.04
120	-12.72	161.8
136	3.28	10.76
128	-4.72	22.28
106	-26.72	713.96
90	-42.72	1825
93	-39.72	1577.68
98	-34.72	1205.48
127	-5.72	32.72
150	17.28	298.6
172	39.28	1542.92
175	42.28	1787.6
156	23.28	541.96
154	21.28	452.84
136	3.28	10.76
148	15.28	233.48
160	27.28	744.2
148	15.28	233.48
139	6.28	39.44
168	35.28	1244.68

Sumatoria del cuadrado de la diferencia entre los datos observados y la media:

$$\sum_{i=1}^n (x - x_m)^2 = 30401.4 \text{ cm}^2$$

La varianza y desviación estándar se calcula de la siguiente manera:

$$\sigma_{obs}^2 = \frac{\sum_{i=1}^n (x - x_m)^2}{n - 1} = 675.48 \text{ cm}^2$$

$$\sigma_{obs} = \sqrt{\frac{\sum_{i=1}^n (x - x_m)^2}{n - 1}}$$

$$\sigma_{obs} = 25.99 \text{ cm}$$

A continuación, se presenta el cálculo de incertidumbre de los datos pronosticados:

Media de los datos observados: $\chi_m = 130.59 \text{ cm}$

Tabla 20: Estadísticos de datos pronosticados.

Fuente: Elaboración propia.

Nivel Estimado (cm)	$x_m - X$	$(x_m - X)^2$
143.08	10.36	107.33
123.56	-9.16	83.91
111.61	-21.11	445.63
123.99	-8.73	76.21
125.54	-7.18	51.55
112.5	-20.22	408.85
122.88	-9.84	96.83
188.05	55.33	3061.41
155.44	22.72	516.2
108.36	-24.36	593.41
108.39	-24.33	591.95
108.62	-24.1	580.81
116.47	-16.25	264.06
107.53	-25.19	634.54
155.19	22.47	504.9
105.47	-27.25	742.56
115.26	-17.46	304.85
164.09	31.37	984.08
173.54	40.82	1666.27
153.95	21.23	450.71
145.00	12.28	150.8
110.45	-22.27	495.95
111.07	-21.65	468.72
111.88	-20.84	434.31
115.2	-17.52	306.95
111.15	-21.57	465.26
112.91	-19.81	392.44
113.88	-18.84	354.95
121.56	-11.16	124.55
118.59	-14.13	199.66
111.22	-21.5	462.25
104.84	-27.88	777.29
103.18	-29.54	872.61
103.13	-29.59	875.57
120.03	-12.69	161.04
157.72	25	625
161.33	28.61	818.53
155.8	23.08	532.69

154.52	21.8	475.24
151.68	18.96	359.48
150.73	18.01	324.36
120.5	-12.22	149.33
130.99	-1.73	2.99
161.36	28.64	820.25
157.78	25.06	628
167.17	34.45	1186.8

Sumatoria del cuadrado de la diferencia entre los datos observados y la media:

$$\sum_{i=1}^n (x - x_m)^2 = 24631.08 \text{ cm}^2$$

La varianza y desviación estándar se calcula de la siguiente manera:

$$\sigma_{pr}^2 = \frac{\sum_{i=1}^n (x - x_m)^2}{n - 1} = 547.36 \text{ cm}^2$$

$$S_{pr} = \sqrt{\frac{\sum_{i=1}^n (x - x_m)^2}{n - 1}}$$

$$S_{pr} = 23.40 \text{ cm}$$

Los valores de desviación estándar para la muestra de registros observados fueron de 25.99 cm y para la muestra de datos pronosticados fue de 23.40 cm, además, las medias fueron muy similares, siendo 132.72 cm para los datos observados y de 130.59 cm para los datos pronosticados. Esto afirma que los resultados del pronóstico son muy similares a los observados, por lo tanto, se afirma la funcionalidad del modelo.

4.3 Estimación de errores.

La determinación de los valores de error del pronóstico se realizó mediante cinco índices de estimación: error medio del pronóstico, error medio al cuadrado, desviación estándar de los errores, desviación absoluta de la media y error porcentual medio. A continuación, se presenta la formulación para calcular el error:

Error medio del pronóstico:

$$EPM = \frac{\sum_{t=1}^n (Y_t - Y'_t)}{n}$$

Error medio al cuadrado:

$$EMC = \frac{\sum_{t=1}^n (Y_t - Y'_t)^2}{n}$$

Desviación estándar de los errores

$$DDE = \sqrt{\frac{\sum_{t=1}^n (e_t - EMP)^2}{n - 1}}$$

Desviación absoluta de la media

$$DAM = \frac{\sum_{t=1}^n |Y_t - Y'_t|}{n}$$

Error porcentual medio

$$PEMA = \frac{\sum_{t=1}^n |Y_t - Y'_t|^2 * 100}{n}$$

Resumen del cálculo del error según los índices anteriores:

Tabla 21: Errores de pronóstico.

Fuente: Elaboración propia.

Nro.	Nivel Estimado (cm)	Nivel Observado (cm)	EPM	EMC	DEE	DAM	PEMA
1	143.08	120.00	23.08	532.80	635.47	23.08	19.24%
2	123.56	134.00	-10.44	109.07	69.18	10.44	7.79%
3	111.61	108.00	3.61	13.05	32.93	3.61	3.34%
4	123.99	123.00	0.99	0.98	9.71	0.99	0.81%
5	125.54	118.00	7.54	56.83	93.40	7.54	6.39%
6	112.50	107.00	5.50	30.24	58.14	5.50	5.14%
7	122.88	120.00	2.88	8.27	25.03	2.88	2.40%
8	188.05	175.00	13.05	170.33	230.34	13.05	7.46%
9	155.44	152.00	3.44	11.82	30.95	3.44	2.26%
10	108.36	110.00	-1.64	2.68	0.24	1.64	1.49%
11	108.39	114.00	-5.61	31.49	12.15	5.61	4.92%
12	108.62	112.00	-3.38	11.42	1.57	3.38	3.02%

13	116.47	119.00	-2.53	6.42	0.17	2.53	2.13%
14	107.53	87.00	20.53	421.34	513.14	20.53	23.59%
15	155.19	165.00	-9.81	96.25	59.06	9.81	5.95%
16	105.47	119.00	-13.53	182.94	129.95	13.53	11.37%
17	115.26	127.00	-11.74	137.86	92.46	11.74	9.25%
18	164.09	161.00	3.09	9.54	27.20	3.09	1.92%
19	173.54	205.00	-31.46	989.74	860.49	31.46	15.35%
20	153.95	152.00	1.95	3.81	16.63	1.95	1.28%
21	145.00	157.00	-12.00	143.93	97.44	12.00	7.64%
22	110.45	105.00	5.45	29.74	57.45	5.45	5.19%
23	111.07	110.00	1.07	1.14	10.19	1.07	0.97%
24	111.88	140.00	-28.12	790.64	675.60	28.12	20.08%
25	115.20	116.00	-0.80	0.65	1.75	0.80	0.69%
26	111.15	121.00	-9.85	96.95	59.60	9.85	8.14%
27	112.91	124.00	-11.09	122.93	80.30	11.09	8.94%
28	113.88	120.00	-6.12	37.40	15.92	6.12	5.10%
29	121.56	136.00	-14.44	208.65	151.75	14.44	10.62%
30	118.59	128.00	-9.41	88.55	53.06	9.41	7.35%
31	111.22	106.00	5.22	27.26	53.98	5.22	4.93%
32	104.84	90.00	14.84	220.22	287.84	14.84	16.49%
33	103.18	93.00	10.18	103.65	151.46	10.18	10.95%
34	103.13	98.00	5.13	26.37	52.72	5.13	5.24%
35	120.03	127.00	-6.97	48.59	23.47	6.97	5.49%
36	157.72	150.00	7.72	59.57	96.91	7.72	5.15%
37	161.33	172.00	-10.67	113.78	72.94	10.67	6.20%
38	155.80	175.00	-19.20	368.54	291.43	19.20	10.97%
39	154.52	156.00	-1.48	2.19	0.42	1.48	0.95%
40	151.68	154.00	-2.32	5.37	0.04	2.32	1.50%
41	150.73	136.00	14.73	216.94	284.08	14.73	10.83%
42	120.50	148.00	-27.50	756.08	643.68	27.50	18.58%
43	130.99	160.00	-29.01	841.73	722.89	29.01	18.13%
44	161.36	148.00	13.36	178.44	239.76	13.36	9.03%
45	157.78	139.00	18.78	352.87	437.27	18.78	13.51%
46	167.17	168.00	-0.83	0.69	1.67	0.83	0.50%
Valor de error			-2.13	4.52	12.88	10.05	7.57%

En base a los errores del modelo calculados, se puede percibir que el pronóstico está un poco subestimado ya que el valor del EPM es negativo, pero poco alejado del cero. El MAPE presenta un valor más perceptible sobre el valor error, que es 7.57%, esto representa un muy buen ajuste de la red neuronal, esto considerando que los registros fueron de un periodo corto de tiempo (ver apartado 2.3.2).

4.3.1 Predicción de caudales

La información hidrológica que se recolectó consistía en nivel de río o altura del agua, por lo tanto, esto debe transformarse a valores de caudal. La ecuación de la curva de descarga nos permite hacer esta transformación, para ello, nos apoyaremos de la ecuación 3.

La variable H de la curva de descarga será quien tome los valores de nivel de río que se pronosticaron mediante la red neuronal

A continuación, se presenta el nivel de río medido en la estación hidrológica Paute en Paute (Latitud: 751257.20 y Longitud: 9694849.60) con su respectiva transformación a caudal utilizando la curva de descarga, y análogamente también se presenta el pronóstico del nivel del río para la estación Paute en Paute y el caudal que representa el pronóstico.

Tabla 22: Nivel estimado y observado con transformación a caudal en diferentes fechas.

Fuente: Etapa EP y elaboración propia.

Fecha	Nivel Estimado (cm)	Nivel Observado (cm)	Caudal Estimado (m3/s)	Caudal Observado (m3/s)
25/08/2017	143.08	120	23.25	11.21
29/08/2017	123.56	134	12.83	18.08
09/09/2017	111.61	108	7.77	6.45
28/08/2017	123.99	123	13.03	12.57
01/09/2017	125.54	118	13.77	10.34
29/09/2017	112.5	107	8.11	6.11
25/08/2017	122.88	120	12.51	11.21
18/10/2017	188.05	175	56.33	45.5
21/10/2017	155.44	152	31.13	28.84
03/10/2017	108.36	110	6.58	7.17
08/02/2018	108.39	114	6.59	8.69
01/02/2018	108.62	112	6.67	7.92
04/09/2017	116.47	119	9.7	10.77
18/12/2017	107.53	87	6.29	0.99
03/01/2018	155.19	165	30.96	37.87
04/09/2017	105.47	119	5.59	10.77
20/09/2017	115.26	127	9.2	14.48
18/01/2018	164.09	161	37.2	34.98
22/01/2018	173.54	205	44.35	71.8

21/10/2017	153.95	152	30.13	28.84
28/01/2018	145	157	24.41	32.19
10/09/2017	110.45	105	7.34	5.44
03/10/2017	111.07	110	7.56	7.17
30/08/2017	111.88	140	7.87	21.43
29/11/2017	115.2	116	9.18	9.5
06/01/2018	111.15	121	7.6	11.66
26/08/2017	112.91	124	8.27	13.03
25/08/2017	113.88	120	8.65	11.21
11/03/2018	121.56	136	11.91	19.17
31/08/2017	118.59	128	10.6	14.97
14/03/2018	111.22	106	7.62	5.77
15/12/2017	104.84	90	5.38	1.52
17/03/2018	103.18	93	4.86	2.14
19/11/2017	103.13	98	4.84	3.36
20/09/2017	120.03	127	11.22	14.48
29/10/2017	157.72	150	32.69	27.54
02/04/2018	161.33	172	35.22	43.15
18/10/2017	155.8	175	31.37	45.5
07/04/2018	154.52	156	30.51	31.51
11/04/2018	151.68	154	28.63	30.16
11/03/2018	150.73	136	28.01	19.17
24/02/2018	120.5	148	11.43	26.27
19/04/2018	130.99	160	16.49	34.28
24/02/2018	161.36	148	35.24	26.27
15/06/2018	157.78	139	32.73	20.86
21/06/2018	167.17	168	39.47	40.1

4.4 Discusión de resultados

El modelo fue ejecutado con las funciones de activación Rectificadora, Maxout, Tangente Hiperbólica y Rectifier With Dropout, pero los mejores resultados se obtuvieron con la función Tangente Hiperbólica.

Existen registros muy similares entre en nivel de agua estimado y observado que representan la eficacia del método. Pero también existen valores que son considerablemente diferentes y esto se debe a que la base de datos o registros es de un periodo corto de tiempo.

Al comparar los datos estadísticos del pronóstico y las observaciones, se puede afirmar la similitud entre medias, varianza y desviación estándar, así mismo, los valores del pronóstico y

observaciones son muy cercanos y esto se evidencia en el cálculo del error, ya que se obtiene un error de 7.57% en términos de error porcentual medio.

La ecuación de curva descarga no se ajusta en ningún caso a todos los datos, es por ello, siempre se concebirá un error producto de este ajuste. Es por esta razón que el error del modelo se calculó en base al valor de pronóstico de nivel de río.

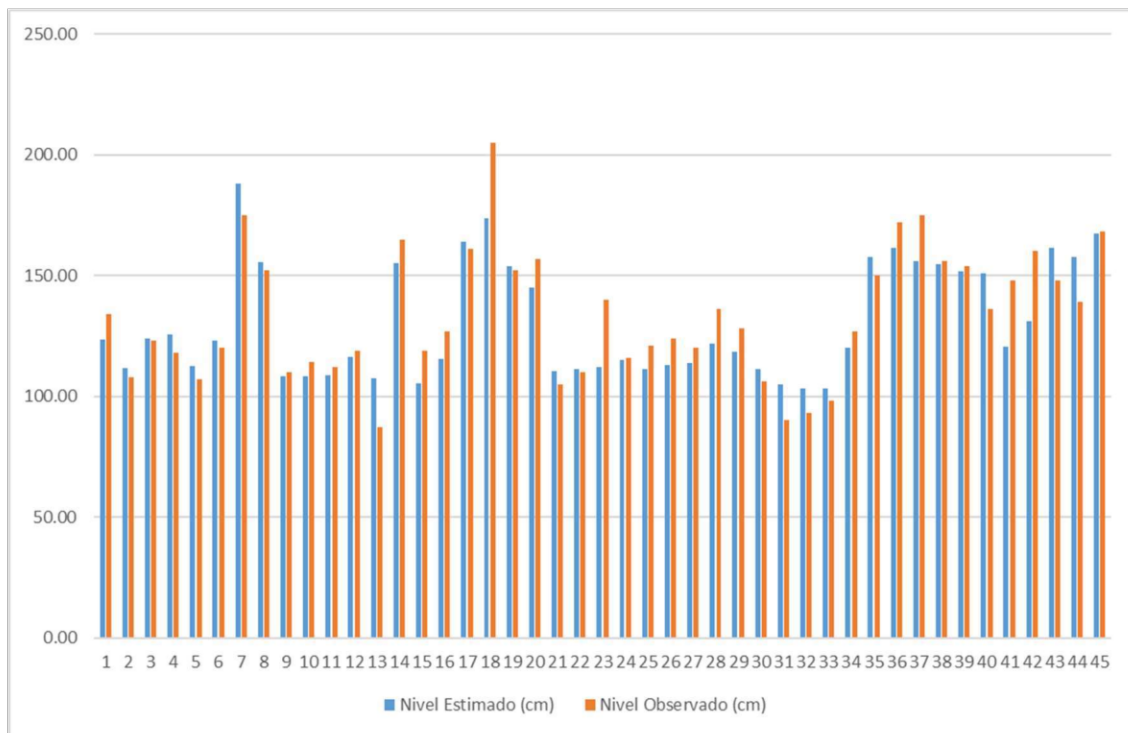


Figura 30: Comparación gráfica de resultados del modelo.

Fuente: Elaboración propia.

En la gráfica de barras (ver fig. 31) se puede evidenciar que lo observado y lo estimado tienden a tener un comportamiento similar, incluso los picos de nivel de agua tienden a comportarse de manera similar.

Tanto la media y el coeficiente de variación son valores muy similares, lo cual representa un muy buen ajuste del pronóstico. Es necesario contar con una base de datos más grande para lograr un mejor ajuste, sin embargo, el modelo presenta un error aceptable, menor al 10%, por lo que, se considera un muy buen ajuste.

CAPITULO V: CONCLUSIONES.

Las redes neuronales nacen de la necesidad de entender el comportamiento del cerebro humano o de la necesidad de imitar el comportamiento propio del ser humano, por ello, es indispensable estudiar a profundidad métodos y aplicar modelos matemáticos que nos permitan entender su funcionamiento, especialmente cuando hablamos de redes neuronales que hacen clasificación o regresiones, que buscan pronosticar variables que están estrechamente ligadas a otras variables, como sucede en esta investigación.

No existe ninguna configuración predeterminada de número de capas de entrada, capas ocultas, número de neuronas o capas de salida que puedan arrojar los mejores resultados en un modelo y en otro. Cada red neuronal y cada algoritmo se plantea de forma diferente y el investigador en base a su experiencia y conocimiento con prueba y error decide la mejor configuración. Para nuestros datos, la mejor configuración resultó ser siete capas de entrada, que son las siete estaciones hidrológicas que forman la base de la investigación, una capa oculta con 50 neuronas u una capa de salida con una neurona. Por lo tanto, el modelo se puede definir como $7 - 50 - 1$.

La base de datos que usamos en esta investigación parte de junio de 2014 a julio de 2019, sin embargo, existen muchos registros faltantes, por lo que la base de datos se ve disminuida, aun así, la robustez del modelo arrojó datos muy buenos y bastante similares a lo observado en campo.

Ya que los datos de caudal de las diversas estaciones no se pudieron obtener, se tuvo que hacer uso de los niveles de los ríos, siendo las estaciones YanuncayAjTarquiL, MataderoSayausiPL, TomebambaQtaBalzayL, MachangaraLLanteraL, TarquiAjYanuncayL, TomebambaPTARL, GualaceoDjPamarPL las que mayor número de registros tuvieron. Luego se hace uso de la curva de descarga donde se obtuvo un excelente ajuste con un coeficiente de correlación muy bueno ($R=0.998502$) y se procedió a transformar los datos de nivel pronosticados a caudal. El error entre los niveles de caudal observados y pronosticados en todas las iteraciones siempre se mantuvo en menos del 10% y un mínimo de 7.57% hablando en términos de error porcentual medio (PEMA).

Finalmente, con base en los datos estadísticos presentados en el capítulo anterior, se evidencia que el modelo planteado funciona y da muy buenos resultados, aún sin tener una base de datos extensa, además, se da un gran paso para vincular la inteligencia artificial que propone los modelos de redes

neuronales con la finalidad de pronosticar caudales y prever en parte el comportamiento de los fenómenos naturales como la lluvia y posterior transformación a caudales.

RECOMENDACIONES.

- Las instituciones gubernamentales y no gubernamentales deben ofrecer la información de calidad y de libre acceso para temas de investigación, de igual forma, deben contar con registros actualizados ya que estos son la base de cualquier exploración investigativa.
- La aplicación de lenguajes de programación de libre acceso debe hacerse una prioridad para que cualquier investigador pueda replicar el estudio y acoplarlo a su realidad.
- Contar una base de datos firme y lo más extensa posible, a fin de ofrecer miles de ejemplos a la red neuronal, así mismo, esto le ayuda a generalizar y a prender mucho más rápido, mejora la precisión del modelo y ofrece mejores resultados.
- Aumentar el número de neuronas, capas ocultas, variables de entrada, funciones de activación forma parte de la investigación, debido a que se busca mediante prueba y error la mejor configuración posible para el modelo propuesto.
- Desarrollar nuevos algoritmos en base al ya creado que puedan ser alimentados con más información, y usar código sencillo a fin de optimizar los tiempos de ejecución del modelo.
- Es importante difundir esta investigación, ya que los desastres naturales como inundaciones y sequías son una gran amenaza, producen pérdidas, tanto económicas como humanas que pueden ser evitadas.

REFERENCIAS

- Amador, N. (2018). *Red Neuronal Artificial (RNA) para pronóstico de intensidad de lluvia a 10 minutos en el Area Metropolitana de Guadalajara. 1*, 14. Retrieved from http://iam.cucei.udg.mx/sites/default/files/adjuntos/correcciones_modularhechas.pdf
- Andrade Tepán, E. C. (2013). *Estudio de los principales tipos de redes neuronales y las herramientas para su aplicación*. Retrieved from <http://dspace.ups.edu.ec/bitstream/123456789/1506/14/UPS-CT002062.pdf>
- Angarita, J., & Perez, W. (2021). *Comportamiento tarifario de energías renovables tipo eólica y solar en la generación de energía eléctrica en Colombia*. Retrieved from <https://repository.universidadean.edu.co/bitstream/handle/10882/11402/PerezWalter2021.pdf?sequence=1>
- Antona, C. (2017). *Herramientas modernas en redes neuronales: la librería Keras*. Retrieved from <https://repositorio.uam.es/handle/10486/677854>
- Aparicio Mijares, F. J. (1989). Fundamentos De Hidrología De Superficie. In *Journal of Chemical Information and Modeling* (Vol. 53). Retrieved from http://webdelprofesor.ula.ve/ingenieria/adamoreno/HIDRO/Fundamentos_de_hidrologia_de_superficie_-_Aparicio.pdf
- Arboleda, F., Moreno, C., & Santos, D. (2012). Compendio curvas altura gasto periodo 2011-2012. *Senamhi*. Retrieved from <https://repositorio.senamhi.gob.pe/handle/20.500.12542/2146>
- Arenas Cárdenas, S. (2010). *Desarrollo de un modelo de pronóstico de cuadales semanales asociado a la variabilidad climática intranual en Colombia*. Retrieved from <http://www.bdigital.unal.edu.co/2029/1/98668319.2010.pdf>
- Basogain, X. (2007). *Redes Neuronales Artificiales y sus aplicaciones. Escuela Superior de Ingeniería de Bilbao, 1*. Retrieved from http://cvb.ehu.es/open_course_ware/castellano/tecnicas/redes_neuro/contenidos/pdf/libro-del-curso.pdf
- Béjan, W., Valeriano, K., Ilachoque, J., & Sulla, J. (2016). Predicción de caudales medios diarios

en la cuenca del Amazonas aplicando redes neuronales artificiales y el modelo neurodifuso ANFIS. *Research in Computing Science*, 113(1), 23–35. <https://doi.org/10.13053/rcs-113-1-2>

Carvajal, Y., Correa, G., & Muños, F. (2007). Modelos De Predicción De Caudal Utilizando Variables Macroclimáticas Y Técnicas Estadísticas Multivariadas En El Valle Del Río Cauca. *Ingeniería de Recursos Naturales y Del Ambiente*, (6), 67–81. Retrieved from <https://www.redalyc.org/articulo.oa?id=231120826008>

Castro, J. F. (2006). Fundamentos para la implementación de red neuronal perceptrón multicapa mediante software. *Universidad San Carlos de Guatemala*, 114. Retrieved from http://biblioteca.usac.edu.gt/tesis/08/08_0155_EO.pdf

Chereque, W. (2010). *Hidrología para estudiantes de ingeniería civil*. Retrieved from <https://repositorio.pucp.edu.pe/index/handle/123456789/28689>

Chow, V. Te, R. Maidment, D., & W. Mays, L. (1994). Hidrología aplicada. *McGraw Hill Higher Education.*, 1, 390. Retrieved from <https://formandoconstructores.com/hidrologia-aplicada-ven-te-chow-1ed-1994-mcgraw-hill/>

Cordero, I. (2013). *Evaluacion de la Cuenca del río Puate, estrategias y líneas de acción para superarlas*. 151. Retrieved from <http://web.ambiente.gob.ec/sites/default/files/users/dhermida/estudiovulnerabilidad.pdf>

de la Hoz, B., Canchano, O., Cordero, L., & Sánchez, P. (2019). Redes neuronales para pronóstico de series de tiempo hidrológicas del Caribe Colombiano. *Revista I+D En TIC*, 10(2), 18–31. Retrieved from <https://revistas.unisimon.edu.co/index.php/identific/article/download/4008/4360>

Dolores, D., & Aguero, J. (2019). Delimitación hidrográfica y caracterización morfométrica de la Cuenca del Río Anzulón. *Universidad Cooperativa de Colombia Sede Villavicencio*, 1–9. Retrieved from <https://inta.gob.ar/documentos/delimitacion-hidrografica-y-caracterizacion-morfometrica-de-la-cuenca-del-rio-anzulon-1>

Gaspari, F., Rodríguez, A., Sinisterra, G., Denegri, G., Delgado, M., & Besteiro, S. (2012). Caracterización morfométrica de la cuenca alta del río Sauce Grande, Buenos Aires,

- Argentina. *AUGMDOMUS*, 4(1852–2181), 16. Retrieved from <https://revistas.unlp.edu.ar/domus/article/view/476>
- Haykin, S. (2018). Neural networks and learning machines. In *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics* (Vol. 1–3). <https://doi.org/10.1016/B978-0-12-809633-8.20339-7>
- Luna, J. (2013). Predicción y pronóstico de tormentas en regiones de Montaña - Aplicación en la cuenca del río La Paz, Bolivia. *Universidad Nacional Autónoma de México*, 277. Retrieved from www.ptolomeo.unam.mx:8080/jspui/bitstream/132.248.52.100/6012/1/tesis.pdf
- Lux, B. (2020). *Conceptos básicos de Morfometría de Cuencas Hidrográficas*. Retrieved from <https://core.ac.uk/download/pdf/35294551.pdf>
- Matich, D. J. (2001). *Redes Neuronales: Conceptos Básicos y Aplicaciones*. 55. Retrieved from <ftp://decsai.ugr.es/pub/usuarios/castro/Material-Redes-Neuronales/Libros/matich-redesneuronales.pdf>
- Ministerio del Ambiente. (2009). *Estudio de vulnerabilidad actual a los riesgos climáticos en el sector de los recursos hídricos en las cuencas de los Ríos Paute, Jubones, Catamayo, Chone, Portoviejo y Babahoyo*. Retrieved from <http://web.ambiente.gob.ec/sites/default/files/users/dhermida/estudiovulnerabilidad.pdf>
- Nastos, P. T., Paliatsos, A., Koukouletsos, K., Larissi, I., & Moustris, K. (2014). Artificial neural networks modeling for forecasting the maximum daily total precipitation at Athens, Greece. *Atmospheric Research*, 144, 141–150. <https://doi.org/10.1016/j.atmosres.2013.11.013>
- Oré, R. (2019). Predicción de caudales mediante redes neuronales artificiales empleando información de sensores remotos en la cuenca experimental del Río Ichu. *Universidad Nacional de Huancavelica*, 1. Retrieved from <https://repositorio.unh.edu.pe/items/50016a96-287c-4d9e-ba7b-5a72c822f904>
- Organización Meteorológica Mundial (OMM). (2011). *Guía de prácticas hidrológicas. Volumen II. Gestión de recursos hídricos y aplicación de prácticas hidrológicas*. Retrieved from https://library.wmo.int/index.php?lvl=notice_display&id=9404#.X10SsWgzbDc
- Perdigón, R., & González, N. (2020). Una revisión bibliográfica sobre modelos para predecir las

- producciones de leche. *Revista Ingeniería Agrícola*, 10(4), 69–77. Retrieved from <https://revistas.unah.edu.cu/index.php/IAgric/article/download/1312/2325>
- Ramírez, C. A., Carvajal, Y., Bocanegra, R. A., & Sandoval, M. C. (2018). Determinación De Relaciones Nivel – Caudal Simple O Compleja En Un Río. Caso Del Río Cauca. *Ingeniería Y Competitividad*, 20(2), 45. <https://doi.org/10.25100/iyc.v20i2.6354>
- Sanca, M. (2011). Clases y tipos de Investigación Científica. *Exposiciones y/o Temas de Explicación, Metodología de Investigación*, 9, 621–624. Retrieved from <https://investigacionestodo.wordpress.com/2012/05/19/clases-y-tipos-de-investigacion-cientifica/>
- Serna, E. (2017). *Principios y características de las redes neuronales artificiales*. Retrieved from https://www.researchgate.net/publication/331498946_Principios_y_caracteristicas_de_las_redes_neuronales_artificiales
- Shamseldin, A. Y., Nasr, A. E., & O'Connor, K. M. (2002). Comparison of different forms of the Multi-layer Feed-Forward Neural Network method used for river flow forecasting. *Hydrology and Earth System Sciences*, 6(4), 671–684. <https://doi.org/10.5194/hess-6-671-2002>
- Turleone, A. (2022). Modelado del caudal natural en la cuenca hidrográfica Chambo con redes neuronales. *Universidad de Cantabria*. Retrieved from <https://repositorio.unican.es/xmlui/handle/10902/27063>
- World Bank Group. (2022). *Measuring Precipitation: on the ground and from space*. 1–7. Retrieved from <https://olc.worldbank.org/sites/default/files/sco/E7B1C4DE-C187-5EDB-3EF2-897802DEA3BF/Nasa/chapter2.html>

ANEXOS

Base de datos.

Tabla 23: Base de datos.

Fuente: Etapa EP.

YanuncayAj TarquiL (cm)	Matadero SayausiPL (cm)	Tomebamba QtaBalzayL (cm)	Machangara LLanteraL (cm)	TarquiAj YanuncayL (cm)	Tomebamba PTARL (cm)	GualaceoDj PamarPL (cm)	Estación Paute (cm)
89.81136	61.09452	47.89163	41.44691	941.2202	116.6	80.5	120
92.03187	58.70042	46.2682	41.50793	579.6457	113.8	80	124
90.20307	58.2067	45.91819	41.31524	89.92062	113.7	84.2	117
91.10789	56.97771	44.76317	41.18425	90.50838	113.4	80.4	123
92.91611	65.58868	50.11893	38.78531	92.73712	117.4	92.8	134
94.05383	63.08563	48.57261	32.67458	92.94611	120.2	95.9	140
92.23344	61.18069	47.94634	32.21975	90.84936	117.2	97.2	128
91.89617	58.67913	46.45367	36.0056	89.06114	114.6	84.8	118
90.98981	57.70163	45.57159	47.18933	87.97279	113.2	80.1	124
91.00433	59.5979	46.68775	48.00071	88.09338	113.8	86.9	116
90.31583	58.14233	45.37106	49.01287	80.96861	112.5	81.8	119
89.53402	56.28582	44.17864	48.56001	41.40256	111.2	79.42	109
87.10898	53.39045	42.85604	47.82656	32.94394	108.6	75.11	114

89.52873	56.53816	44.54349	49.91663	23.234	108.8	95	133
92.50565	61.70224	48.10027	47.84366	22.4369	115.5	100.9	112
88.52354	54.29119	43.5929	47.46	12.19616	109.6	77.59	108
87.76101	51.85405	42.16859	49.72882	10.60069	108	72.7	105
92.45802	54.20839	43.47572	49.57457	5.092768	110.3	72.02	114
95.98975	56.50436	45.85008	49.76445	5.85595	114.2	74.77	105
91.43995	51.94363	42.46529	49.74865	5.591231	108.3	72.45	145
92.33202	53.81015	43.51229	48.82579	4.768138	113.1	95.8	123
91.69425	55.40968	44.1617	48.37132	2.655859	111.4	95.1	119
91.2593	52.84053	42.96912	48.51197	3.277589	110	100	118
91.72599	52.56404	42.82098	48.31701	3.261683	110.7	106.1	226
101.5518	80.15228	59.75597	48.59132	12.48716	125.1	163.7	230
111.4978	87.6701	65.61779	51.03392	93.52383	142.1	165.3	144
102.1962	78.18639	59.04869	49.10802	85.18187	131	101.3	127
96.34917	68.09084	52.6923	48.00777	80.09788	120.5	84.1	215
92.93202	61.24941	48.88321	48.71588	81.00189	114.1	77.84	209
91.20845	57.1612	45.30994	48.8931	80.92608	111.5	74.77	109
89.35536	55.34956	44.36209	48.80946	96.69757	112	73.64	124
90.08092	55.82418	44.27039	48.47741	101.5562	112.8	89.1	122
90.30627	55.71203	44.00023	48.82316	10.67566	113.4	114.8	128
93.39917	56.47402	44.43642	48.75695	97.56642	121.7	103.7	108
91.29892	52.84271	42.5326	48.78002	90.76963	114.1	83.1	109

89.06813	50.68724	41.43329	48.38146	86.02357	110.7	81.8	107
87.94814	49.17815	41.02518	48.49789	76.11553	110.6	91.7	140
98.1554	69.24882	54.46725	51.62248	88.79454	124.8	120.4	137
101.1228	78.05051	59.88216	53.50593	80.30081	133.7	100.2	119
91.96703	63.4194	50.1763	49.83113	59.97235	120.4	80.6	110
90.20933	57.97775	46.92655	49.47718	69.43538	117.6	75.29	122
98.9856	104.1548	77.16801	53.76468	131.215	143.6	79.23	145
124.1256	99.08323	76.42815	53.3961	193.3419	158.6	86.7	144
118.0303	95.3441	73.11923	53.90305	176.834	148.5	82.6	129
106.6896	84.37361	64.40894	53.46928	165.0957	139.2	76.71	128
103.389	81.28629	62.15408	57.44857	169.3442	144.1	73	125
105.4324	77.50806	58.89597	54.83254	163.573	135.3	71.45	120
100.0191	71.6825	54.79133	53.73786	195.9714	129.6	71.82	137
98.66246	70.25885	53.6258	55.14452	178.1035	129.6	86.2	119
97.04581	68.09048	52.35357	55.28035	175.0255	129	81.7	120
95.52261	63.80492	49.72744	55.09087	206.51	125.2	79.58	124
94.21861	62.95773	49.02284	54.8402	234.9452	123.6	89.2	290
105.4134	97.51375	72.3188	58.61839	22.24391	141	224.3	260
117.662	118.5959	90.31581	62.29921	23.87044	163.2	194	175
109.8929	110.7187	84.22203	59.39764	18.31793	153.5	115.2	143
103.5985	97.17948	75.04919	59.18937	28.67402	145	90.9	135
100.0365	88.03174	66.15105	58.64945	51.53654	139.4	83.8	152

103.9645	101.0456	75.99709	60.76404	62.95282	146.7	121.5	169
126.451	120.7202	91.80294	64.04054	24.88471	167.8	134.2	186
129.3505	132.7727	102.6211	65.08655	24.74056	182.3	127.5	175
120.7001	122.9652	96.47869	62.12593	169.9511	171.9	103.3	152
115.7445	103.4467	82.71919	59.50363	192.3555	157.5	85.9	143
109.5648	93.32617	75.69155	59.13938	143.8043	149.8	80.7	134
103.0957	84.4751	67.20149	58.23652	93.30656	141.4	80.1	163
103.971	90.69884	73.2318	72.73643	92.54214	164.4	80.3	150
102.19	87.23611	70.5198	62.30238	92.44427	147.1	82.9	128
98.39463	76.24188	60.93779	60.4289	63.50429	139.1	75.17	122
94.21732	69.15027	56.42999	59.66454	67.67535	134.6	72.12	124
103.2404	68.3304	55.13894	59.94103	67.8993	138.6	70.55	130
109.0749	76.94982	61.77087	60.30391	48.49047	148.6	69.92	130
103.0383	71.34106	57.21714	58.89878	52.16431	136.8	69.68	124
101.5127	62.66857	54.46337	58.42491	65.8519	134.7	68.68	120
92.68314	60.75951	50.59944	57.61122	97.14673	127.5	66.76	122
91.6325	57.624	47.77164	57.17266	105.0784	124.4	66.65	110
90.36961	54.28026	45.49979	57.36801	101.5634	122.9	66.89	113
81.82661	40.8207	38.52611	80.36201	29.67331	113.2	62.22	96
82.67844	42.96939	38.89003	81.24873	27.71985	114.8	65.1	110
83.66535	43.48986	40.20596	83.03384	25.41513	116.3	71.28	115
83.02148	40.38869	38.48908	80.78799	26.55114	113.4	66.63	114

85.07928	41.09356	38.88006	79.97523	25.40361	115.7	103.1	110
84.8364	44.03424	40.01284	80.98148	25.76543	110.2	88.3	113
83.98922	39.70631	37.99456	80.95561	25.69805	107.7	72.15	125
84.48855	37.46741	37.21347	81.31434	26.24739	105.2	83.9	123
83.54472	36.71643	36.49485	77.7818	26.68996	104.8	76.73	119
83.83722	35.93121	35.92807	73.61806	25.98159	105.1	95.8	116
82.69107	36.80338	36.11449	73.72036	25.52759	106.1	100.7	120
80.70028	36.44628	36.00307	73.73246	26.76531	102.3	85.4	110
81.76646	36.03894	37.09595	74.57478	27.0646	113.5	73.24	112
85.88263	47.90666	44.42326	75.12644	26.51106	120.2	79.82	128
89.36211	53.94851	49.28113	75.51109	27.04975	120.2	108.7	117
86.41416	47.25499	44.47018	70.09788	27.0504	113.4	78.34	110
85.65533	48.42468	44.51309	66.18887	26.17967	109.4	90	109
84.96317	46.43779	43.28305	66.18887	63.88467	109.4	83	108
83.94274	41.78343	40.28658	66.18887	71.4324	106.9	78.87	107
82.40916	37.02291	38.04816	66.18887	79.79985	103.2	70.74	104
82.26377	59.45556	91.6861	66.18887	93.29507	119.2	67.47	119
106.0419	90.5448	74.8721	66.18887	70.91936	139.4	68.01	119
90.82268	83.59619	68.23676	66.18887	130.2467	129.2	67.68	112
84.92194	69.37056	56.88162	66.18887	89.92142	119.5	65.79	110
81.32455	56.95895	48.70773	66.18887	197.9784	109.7	63.29	98
80.27984	50.25432	45.37708	66.18887	478.0813	106.1	61.8	90

80.723	46.17287	42.79485	41.88774	125.0873	102.6	60.66	90
80.20227	42.81927	40.79432	16.30289	30.67515	101.1	59.88	88
80.08375	38.97385	38.9254	16.30054	30.67515	99.3	59.31	87
118.4553	102.8314	84.11913	10.00031	98.28808	152.6	89.4	187
118.6976	100.0638	82.7673	10.14476	99.87251	164.5	101.1	162
112.2867	99.60268	82.05531	9.914393	94.41067	152	94.4	210
105.8439	92.57298	77.00202	11.33682	91.74523	144	75.45	165
99.59991	86.77695	72.19493	11.89978	90.32333	138.3	69.01	140
95.23593	76.783	62.57399	12.14935	86.48084	126.9	65.34	135
93.01839	75.84204	62.0154	31.99763	85.827	126.5	64.55	121
90.73182	59.79068	51.74949	67.44912	85.93939	118.2	71.53	104
87.29887	53.91531	48.48859	66.26542	83.11816	113.2	63.57	100
84.93949	49.72126	45.26223	64.64998	81.61854	108.7	61.09	96
86.37695	54.61221	46.85217	66.02425	83.36847	113.3	60.3	119
86.08204	68.6617	56.54931	69.7202	91.78358	126.5	73.94	104
94.38347	74.48701	61.46067	68.96738	93.96045	132.2	72.01	127
94.28384	84.97586	70.18519	71.39603	92.86425	139.3	74.31	178
125.9707	110.4741	92.45947	76.94122	120.5825	169.8	114.4	186
139.3204	127.9674	104.0386	78.26997	127.5017	186.8	118	180
150.9408	131.5808	106.0592	78.67331	118.4493	190.8	93.2	161
125.9349	107.826	87.16432	69.75872	109.0465	162.5	87.2	315
131.4762	106.6659	85.84456	67.5472	122.6594	166.9	112.3	318

167.9003	167.4752	138.3175	84.55797	122.5918	216.8	187.8	212
164.2323	157.1618	133.1877	81.08829	121.2468	207.4	190.6	205
146.2567	137.4841	109.5889	75.03437	112.2313	185.7	128.2	175
130.9983	121.0496	94.41985	70.5406	105.2678	167.4	105.2	152
120.4922	106.3972	85.10526	66.6758	100.9704	154.2	94.8	166
114.6497	102.5504	82.5575	69.35565	97.51246	150.5	115.1	150
110.5461	94.43812	76.83733	72.23069	97.34924	147.9	100.3	164
108.059	91.0846	74.09764	77.65331	95.53538	149.1	106.2	157
104.9321	84.7838	70.85052	76.11153	93.56057	144.2	92.2	135
100.9831	78.52866	64.81051	74.86345	90.43353	138.8	80.8	127
97.53621	73.3869	60.07018	74.00077	88.25118	135	76.14	120
94.59685	67.76998	55.6656	73.58919	87.04789	131.8	73.13	112
92.92583	64.47463	53.36536	73.30829	86.08074	129.5	71.34	105
91.82774	61.05896	51.7055	73.04053	86.97585	129.2	70.83	104
92.29168	59.51965	51.60494	73.62952	90.42778	132.5	69.88	118
91.39909	59.82083	51.71494	73.02922	87.52603	129.8	68.38	111
90.17447	55.39388	49.25466	72.85387	85.79091	126.8	66.59	113
89.22091	57.88956	49.65063	73.42004	86.69257	129.1	70.55	122
90.82865	60.27014	49.44144	73.97538	91.11924	130	72.9	114
91.23486	64.10194	52.4655	74.61913	90.15313	131.7	67.81	121
93.45328	75.24165	59.78233	74.8019	90.22791	133.5	65.55	119
92.185	72.64356	58.43583	75.74649	88.33015	134.4	64.51	115

99.94289	67.60864	54.18125	74.74985	91.6082	133.3	63.6	110
93.61803	58.97106	48.94225	74.74152	90.36285	129.7	62.41	107
93.47778	56.67123	46.99133	74.02111	88.65153	128.6	64.35	112
91.76743	57.15899	47.44876	72.3749	89.36042	125.9	69.33	113
93.36018	55.4561	46.48311	72.52775	91.29758	126.3	66.15	108
92.60788	54.82286	45.86248	72.60194	93.81709	127	68.14	121
101.3682	59.93346	49.24917	74.61456	103.4552	140.3	74.34	140
109.8883	84.79782	65.73879	77.79214	101.1001	150.1	69.91	208
164.5463	116.629	90.95377	86.72569	142.055	200.3	103.9	162
130.3426	99.82892	77.86787	78.8159	120.7997	159.1	95.1	178
125.7779	88.56581	70.84675	81.86456	110.8961	160.7	78.8	180
118.3309	87.11256	69.01258	77.78658	105.7964	153.1	72.88	148
109.2132	78.81345	61.26963	74.88629	101.1734	143.5	67.75	127
102.0885	71.99358	56.14822	74.00121	97.42554	137.1	63.93	116
98.23499	66.48078	52.68645	72.56688	93.79552	131.8	61.89	117
95.60204	62.70776	50.45842	71.67157	92.22594	128.5	64.72	122
95.1152	62.92091	49.9724	73.42716	92.90742	130.3	64.67	119
96.20121	65.66915	51.94473	67.72617	91.10783	126.1	65.74	121
97.0763	71.44614	55.0667	69.54854	92.02128	129.3	71.64	124
96.37885	69.64554	54.14679	67.11829	92.78945	127.1	80.6	120
96.39518	71.52534	55.12458	68.02552	92.40285	127.9	78.84	112
95.19772	69.00783	53.83817	68.29411	91.79057	126.9	68.5	149

95.15811	69.49869	54.21341	67.57094	91.48114	126.6	71.71	128
96.74855	77.38838	59.50761	69.52049	91.23171	129.7	115.9	158
97.67139	84.67818	67.36932	70.94829	93.98622	134.8	118.5	131
99.7816	93.17199	72.48958	73.63174	93.49624	140.1	112.5	135
97.5798	83.6273	65.87193	63.88562	91.79082	128.6	87.2	136
95.02916	77.75672	60.04103	63.94122	89.61362	124.3	88.5	128
92.26678	72.87834	56.40387	63.88369	87.94727	121.8	83.3	120
91.7661	71.67505	52.93584	62.17449	84.34656	116.9	71.14	106
89.27263	68.94425	50.51851	61.31463	77.06299	114	66.37	100
87.4549	65.72391	48.80701	59.74454	68.31059	110.3	63.72	90
86.13439	62.61792	47.32577	59.62115	62.18914	108.3	62.07	93
86.03241	61.82596	46.81386	60.04858	56.54266	108.4	62.08	98
84.93051	60.31282	46.35747	59.10063	55.18111	107.7	64.68	92
85.65953	81.34078	58.25747	72.11806	55.956	124.2	63.61	127
88.50878	87.18999	64.8392	71.44313	57.94738	131.3	69.06	126
88.3945	75.35075	55.19667	63.84039	52.9873	118.4	75.3	115
87.00708	70.20306	52.1941	62.16647	48.84821	114.3	75.41	97
95.56597	75.7178	56.11455	61.69439	57.92402	125.3	71.94	136
113.6178	92.90509	71.46655	65.58956	60.81285	147.3	77.5	132
118.8632	108.786	84.16022	75.47395	107.5916	154.8	72.5	125
104.9366	93.86889	72.13156	66.05097	104.3674	143	72.98	115
97.94761	81.53904	59.14572	58.91294	94.84529	124.3	69.16	133

97.48924	78.57802	56.57999	59.7076	92.95947	124.5	89	120
112.0046	81.61321	60.03775	62.73353	98.60773	141.6	90.2	142
140.669	115.4913	87.33751	69.16062	108.5373	170.9	88.2	150
157.1805	114.6189	85.92728	64.0844	115.8557	178.4	122.1	172
150.7661	134.4773	106.6266	80.4178	111.9992	195.8	106.4	196
149.1148	150.6688	106.6669	77.7262	110.7264	195.6	111.5	188
151.4939	149.7461	105.1337	74.94566	108.6267	194	95.9	180
149.1204	145.5114	98.69981	69.05912	104.6544	187.4	96.3	175
136.706	133.321	82.9493	65.07187	95.13969	172.6	93	156
130.5036	127.6644	84.21761	68.21136	87.56266	165.1	81.4	184
133.2783	127.3739	84.72379	74.91879	88.14709	168.5	86	185
124.7772	117.969	79.64953	73.15897	86.03653	156.2	117	150
120.0189	114.3583	77.8533	72.34713	83.53189	160.7	91.9	154
123.9535	107.8807	71.94264	68.8121	83.24751	153.8	88	136
109.4001	98.69344	66.1601	66.89881	76.11464	139.9	74.6	130
103.7254	92.18127	62.32502	66.06579	72.28212	133.4	70.2	127
101.3557	89.02161	61.45467	66.22717	71.57229	131.3	68.92	148
101.461	87.55437	60.7571	66.10724	72.56872	130.7	111.4	300
112.9983	116.3246	75.92193	72.63109	75.06937	150.2	241.7	164
108.9991	107.8659	71.24238	67.49025	71.28139	146.1	108.3	148
104.6743	98.67454	65.60072	63.32775	69.05839	136.8	99	160
101.6668	94.95628	63.0285	62.34247	70.54465	132.8	124.8	174

104.9557	94.4642	62.85295	62.71914	74.85609	133.5	129.7	165
106.2019	94.89555	63.71981	64.94886	72.29677	143	111.6	152
117.919	101.024	67.21614	66.27249	77.55746	148.5	111.2	142
110.0316	132.5837	110.3985	92.14265	130.3672	195.2	121.7	190
82.95216	92.08566	110.3004	80.23158	137.4456	144.9	112.7	148
81.0686	87.60767	110.3335	79.61653	129.5942	140.8	104.7	147
80.15159	88.42578	110.4573	79.4296	124.3683	139.9	107.3	139
79.78111	89.12861	110.5497	80.24985	129.7523	141	113	325
89.51426	123.8088	110.3667	84.93551	137.9576	162	236.8	335
112.2381	168.1268	110.2769	111.3561	145.7925	220	256.1	221
100.6153	151.3319	110.2861	106.5227	138.6072	201.9	147.6	196
92.46735	137.4633	110.1459	102.5714	138.3057	187.6	121.4	172
92.5237	125.1753	110.1441	94.71798	137.6321	172.2	104.7	168
93.17118	115.3948	110.1484	87.94151	133.5745	159.1	97.2	152

Código en R.

Código para generar el modelo.

```
# Tratamiento de datos y graficos
```

```
#
```

```
=====
```

```
#install.packages("tidymodels")
```

```
#install.packages("tidyverse")
```

```
#install.packages("skimr")
```

```
#install.packages("DataExplorer")
```

```
#install.packages("ggpubr")
```

```
#install.packages("mosaicData")
```

```
#install.packages("Metrics")
```

```
library(tidymodels)
```

```
library(tidyverse)
```

```
library(skimr)
```

```
library(DataExplorer)
```

```
library(ggpubr)
```

```
library(mosaicData)
```

```
library(readxl)
```

```
library(Metrics)
```

```
# Modelado
```

```
#
```

```
=====
```

```
=====
```

```
=====
```

```
library(h2o)
```

```
# Datos
```

```
#
```

```
=====
```

```
=====
```

```
=====
```

```
datos <- read_excel("C:/Users/user/Desktop/TESIS/Codigos/niveles.xlsx")
```

```
# Tabla resumen
```

```
#
```

```
=====
```

```
=====
```

```
=====
```

```
skim(datos)
```

```
# Numero de datos ausentes por variable
```

```

#
=====
=====
=====

datos %>% map_dbl(.f = function(x){sum(is.na(x))})

# Porcentaje de valores ausentes

plot_missing(
  data = datos,
  title = "Porcentaje de valores ausentes",
  ggtheme = theme_bw(),
  theme_config = list(legend.position = "none")
)

# Distribucion variable respuesta

#
=====
=====
=====

ggplot(data = datos, aes(x = nivel)) +
  geom_density(fill = "steelblue", alpha = 0.8) +
  geom_rug(alpha = 0.1) +
  scale_x_continuous(labels = scales::comma) +
  labs(title = "Distribucion original") +

```

```

theme_bw()

# Tabla de estadisticos principales de la variable a predecir
summary(datos$nivel)

# Grafico de distribucion para cada variable numerica
#
=====
=====
=====

plot_density(
  data = datos %>% select(-nivel),
  ncol = 3,
  title = "Distribucion variables continuas",
  ggtheme = theme_bw(),
  theme_config = list(
    plot.title = element_text(size = 14, face = "bold"),
    strip.text = element_text(colour = "black", size = 12, face = 2)
  )
)

# Reparto de datos en train y test
#
=====

```

```
=====
=====

# Con el objetivo de poder estimar el error que comete el modelo al predecir
# nuevas observaciones, se dividen los datos en dos grupos, uno de entrenamiento
# y otro de test (80%, 20%).
```

```
set.seed(123)
```

```
split_inicial <- initial_split(
```

```
  data = datos,
```

```
  prop = 0.8,
```

```
  strata = nivel
```

```
)
```

```
datos_train <- training(split_inicial)
```

```
datos_test <- testing(split_inicial)
```

```
# Tras realizar el reparto, se verifica que los dos grupos son similares.
```

```
summary(datos_train$nivel)
```

```
summary(datos_test$nivel)
```

```
# Preprocesado
```

```
#
```

```
=====
```

```
=====
```

```
=====
```

```
transformer <- recipe(
```

```
  formula = nivel ~ .,
```

```
  data = datos_train
```

```
) %>%
```

```
  step_naomit(all_predictors()) %>% #crea una especificacion de un paso de receta que eliminar  
  las
```

```
  #observaciones (filas de datos) si contienen valores NA o NaN.
```

```
  step_nzv(all_predictors()) %>% #crea una especificacion de un paso de receta que eliminar  
  potencialmente las
```

```
  #variables que son muy escasas y desequilibradas.
```

```
  step_center(all_numeric(), -all_outcomes()) %>% #crea una especificacion de un paso de receta  
  que normaliza
```

```
  #los datos numericos para tener una media de cero.
```

```
  step_scale(all_numeric(), -all_outcomes()) %>% #crea una especificacion de un paso de receta  
  que normalizar los
```

```
  #datos numericos para tener una desviacion estandar de uno.
```

```
  step_dummy(all_nominal(), -all_outcomes()) #crea una especificacion de un paso de receta que  
  convertira datos
```

```
#nominales (por ejemplo, caracteres o factores) en uno o mas terminos de modelo binario
numerico para los
```

```
#niveles de los datos originales.
```

```
transformer
```

```
# Se entrena el objeto recipe
```

```
transformer_fit <- prep(transformer)
```

```
datos_train_prep <- bake(transformer_fit, new_data = datos_train)
```

```
datos_test_prep <- bake(transformer_fit, new_data = datos_test)
```

```
glimpse(datos_train_prep)
```

```
# Modelado
```

```
# Inicializacion del cluster
```

```
#
```

```
=====
=====
=====
```

```
h2o.init(
```

```
  nthreads = -1,
```

```

max_mem_size = "4g"
)

# Se eliminan los datos del cluster por si ya habia sido iniciado.
h2o.removeAll()

h2o.no_progress()

# Se transfieren los datos al cluster de H2O.....

datos_train <- as.h2o(datos_train_prep, key = "datos_train")
datos_test  <- as.h2o(datos_test_prep, key = "datos_test")

# Espacio de busqueda de cada hiperparametro
#
=====
=====
=====

hiperparametros <- list(
  activation = c("Rectifier", "Maxout", "Tanh", "RectifierWithDropout"),
  #hidden = list(c(5), c(10), c(50), c(10, 10
  hidden = c(50),
  l1 = c(0, 0.00001, 0.0001),
  l2 = c(0, 0.00001, 0.0001),

```

```

rate = c(0, 0.01, 0.005, 0.001),
rate_annealing = c(1e-8, 1e-7, 1e-6),
rho = c(0.9, 0.95, 0.99, 0.999),
epsilon = c(1e-10, 1e-8, 1e-6, 1e-4),
momentum_start = c(0, 0.5),
momentum_stable = c(0.99, 0.5, 0),
input_dropout_ratio = c(0, 0.1, 0.2),
max_w2 = c(10, 100, 1000, 3.4028235e+38)
)

# Al ser una busqueda aleatoria, hay que indicar criterios de parada.
search_criteria <- list(
  strategy = "RandomDiscrete",
  max_runtime_secs = 5*60, # Tiempo maximo de busqueda (5 minutos)
  max_models = 100, # Numero maximo de modelos
  stopping_tolerance = 0.01,
  stopping_rounds = 5,
  seed = 1234
)

# Busqueda por validacion cruzada

```

```

#
=====
=====
=====

variable_respuesta <- 'nivel'

predictores <- setdiff(colnames(datos_train), variable_respuesta)

grid <- h2o.grid(
  algorithm = "deeplearning",
  epochs    = 5000,
  x         = predictores,
  y         = variable_respuesta,
  training_frame = datos_train,
  nfolds    = 3, #validacion cruzada
  standardize = TRUE,
  hyper_params = hiperparametros,
  search_criteria = search_criteria,
  seed       = 123,
  grid_id    = "grid"
)

# Resultados del grid

#
=====

```

```
=====  
=====  
resultados_grid <- h2o.getGrid(  
  sort_by = 'rmse',  
  grid_id = "grid",  
  decreasing = FALSE  
)  
data.frame(resultados_grid@summary_table)
```

```
# Mejor modelo encontrado
```

```
#
```

```
=====  
=====  
=====  
modelo_final <- h2o.getModel(resultados_grid@model_ids[[1]])
```

```
# PREDICCIONES
```

```
#
```

```
=====  
=====  
=====  
predicciones <- h2o.predict(  
  object = modelo_final,  
  newdata = datos_test
```

```
)
```

```
predicciones <- predicciones %>%
```

```
  as_tibble() %>%
```

```
  mutate(valor_real = as.vector(datos_test$nivel))
```

```
predicciones %>% head(5)
```

```
# Se exporta en .txt en la carpeta Documentos
```

```
write.table(predicciones, "predicciones.txt")
```

```
# CONCLUSION
```

```
#
```

```
=====
=====
=====
```

```
modelo_final@allparameters
```

```
# ERROR DE TEST
```

```
#
```

```
=====
```

```
=====
=====

# Error porcentual medio
```

```
#
```

```
=====
=====

EPM <- mape(predicciones$valor_real, predicciones$predict)*100
```

```
EPM
```

```
rmse(predicciones$valor_real, predicciones$predict)
```

```
# Se apaga el cluster H2O
```

```
#
```

```
=====
=====

#h2o.shutdown(prompt = FALSE)
```

```
obs <- predicciones$valor_real
```

```
pred <- predicciones$predict
```

Código para obtener los parámetros de la curva de calibración de caudales

```
#PARAMETROS DE CURVA DE DESCARGA
```

```
#Librerias requeridas
```

```
library(readxl) #lee archivos de excel
```

```
library(minpack.lm) #Obtener regresiones no lineales
```

```
#Lectura de datos
```

```
datos<- read_excel("C:/Users/user/Desktop/TESIS/Codigos/curva_de_descarga.xlsx",1,na="-99.9")
```

```
aforos<- data.frame(datos[,2], datos[,3]) #en la columna 2 se encuentra el nivel y en la 3 el caudal
```

```
#Omitir los valores faltantes identificados en la data como -99.9
```

```
aforos<- na.omit(aforos)
```

```
#cambio de nombres en las columnas nivel y caudal
```

```
colnames(aforos)<- c("H", "Q") # H = Nivel (m) ----- Q = Caudal (m3/s)
```

```
head(aforos)
```

```
#Escribir el modelo
```

```
model.cd <- nlsLM(Q~K*(H-H0)^n, data = aforos, start = list(K=60, H0=0, n=3))
```

```
summary(model.cd)
```

```
#Graficos
```

```
G1<- plot(aforos[,2], aforos[,1], xlab = "Caudal (m3/s)", ylab = "Nivel (m)", main = "Curva de  
descarga",col="blue")
```

```
G2<- points(predict(model.cd), aforos[,1], col="red")
```

```
legend( x = "topleft",
```

```
       legend = c("Caudales aforados","Caudales calculados"),
```

```
       col = c("blue","red"), lwd = 2, lty = c(0,0),
```

```
       pch = c(1,1) )
```

Información geográfica de la cuenca hidrográfica.

Se adjunta en un CD lo pertinente a la elaboración de la información geográfica de la cuenca hidrográfica.

AUTORIZACION DE PUBLICACION EN EL REPOSITORIO INSTITUCIONAL

Yo Juder Antonio Castillo González, portador de la cédula de ciudadanía N.º 1105377004. En calidad de autor y titular de los derechos patrimoniales del trabajo de titulación **“EVALUACION DE MODELOS DE PRONOSTICO PARA ESTIMAR EL INGRESO DE CAUDALES AL EMBALSE MAZAR: APLICACIÓN CON REDES NEURONALES.”** de conformidad a lo establecido en el artículo 114 Código Orgánico de la Economía Social de los Conocimientos, Creatividad e Innovación, reconocemos a favor de la Universidad Católica de Cuenca una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos, Así mismo; autorizamos a la Universidad para que realice la publicación de este trabajo de titulación en el Repositorio Institucional de conformidad a lo dispuesto en el artículo 144 de la Ley Orgánica de Educación Superior.

Cuenca, **28 de febrero de 2023.**

F: 

Juder Antonio Castillo González.

1105377004