

**TRABAJO DE GRADO**  
**Opción Investigación o Proyecto de Grado**

**Aprendizaje de Máquinas para calibrar un modelo hidrológico agregado en**  
**Cuencas Andinas**

Corporación Universitaria Remington.  
Facultad De Ingeniería  
Ingeniería Civil.

Eliel de Jesus Segura Perez.  
Ing. PhD. Juan Mauricio Bedoya Soto  
Opción de trabajo de grado: (Investigación).  
13/11/2024.

## **Dedicatoria**

Dedicado a mi Dios y en especial a mi madre, por su esfuerzo y sacrificio.

### **Agradecimientos**

Agradezco a mi tutor por la confianza y la oportunidad de participar en esta investigación.

## Tabla de Contenidos

Resumen.....	8
Palabras clave.....	8
Introducción, Marco teórico o de referencia.....	10
1.Introducción .....	10
2. Marco Teórico .....	11
2.1. Modelos Hidrológicos.....	11
2.2. Interacción de modelos hidrológicos por medio de modelos de machine Learning y de técnicas de optimización.....	17
2.3 Modelo de Machine Learning LSTM para predicción .....	19
2.4 Optimización de Parámetros con PSO (Particle Swarm Optimization).....	21
2.5 Métricas de calibración .....	23
Resultados y Discusión.....	35
Conclusiones .....	52
Referencias.....	54

**Lista de tablas**

Tabla 2 Metodología .....	33
Tabla 3 Metadatos de la Estaciones pluviométricas seleccionadas. ....	38
Tabla 4 Hidrología a largo paso.....	38
Tabla 5 Encabezado del DataFrame .....	45
Tabla 6 Parámetros ingresado manualmente a la simulación .....	45
Tabla 7 Validación de la simulación (observado vs real) en un periodo de simulación. .	46
Tabla 8 Parámetros optimizados ingresados manualmente a la simulación.....	48
Tabla 9 Validación de la simulación (observado vs real) en un periodo de simulación. .	48

### Lista de figuras

<b>Figura 2</b> Monitoreo RedRío, Aburrá-Medellín, desde su nacimiento hasta su confluencia. Fuente: AMVA ( <a href="https://www.metropol.gov.co/ambiental/Paginas/RedRio.aspx">https://www.metropol.gov.co/ambiental/Paginas/RedRio.aspx</a> ) ...	35
<b>Figura 3</b> Estaciones de caudal existentes en la delimitación de la cuenca Río Medellín	36
Figura 4 Estaciones pluviométricas.....	37
<b>Figura 5</b> Red de Drenaje (cuenca Río Medellín) .....	39
Figura 6 DEM (cuenca Río Medellín).....	39
Figura 7 Lluvia diaria en cada estación .....	40
Figura 8 Set de datos .....	41
<i>Figura 9 Red de Drenaje (Cuenca Río Medellín)</i> .....	42
Figura 10 Red de Drenaje (cuenca Río Medellín) .....	43
<b>Figura 11</b> Grafico de caudal observado vs simulado en la estación piloto en el periodo 1980-1984 .....	46
<b>Figura 12</b> Gráfico de dispersión Caudal Observado vs Estimado .....	47
<b>Figura 13</b> Gráfico de caudal observado vs simulado en la estación piloto en el periodo 1980-1984 .....	49
<b>Figura 14</b> Costo de la optimización de PSO por medio del indicador KGE durante el rendimiento .....	49
<b>Figura 15</b> Costo de la optimización de PSO por medio del indicador KGE durante el rendimiento .....	51

## Resumen

El presente proyecto de investigación aborda la mejora de la modelización hidrológica en las cuencas andinas de Colombia, un área caracterizada por su complejidad geográfica y climática, así como por la creciente presión sobre los recursos hídricos debido al cambio climático y al crecimiento poblacional. La necesidad de modernizar los modelos hidrológicos tradicionales se justifica por la ineficiencia de estos ante las variaciones climáticas y la heterogeneidad de los datos disponibles.

Para enfrentar este desafío, se desarrolló un modelo hidrológico automatizado en Python que integra técnicas de aprendizaje automático (ML), específicamente redes neuronales LSTM y la técnica de optimización por enjambre de partículas (PSO). La metodología incluye el preprocesamiento automatizado de datos, la selección del modelo de ML más adecuado y un sistema de calibración para mejorar la precisión de las simulaciones.

Los resultados obtenidos tras 10 épocas de entrenamiento muestran que el modelo presenta un error cuadrático medio (RMSE) estabilizado en aproximadamente 0.2, lo que indica un buen ajuste sin sobreajuste. Sin embargo, se observó que el modelo tiende a sobreestimar la lluvia, con errores que varían entre -20 y 5, y presenta picos de error en eventos extremos. La optimización mediante PSO ha demostrado ser efectiva, logrando una reducción en el costo del ajuste (KGE negativo) y mejorando la precisión en la simulación del flujo.

En conclusión, este estudio resalta la efectividad de las técnicas avanzadas de ML en la modelización hidrológica, superando las limitaciones de los modelos tradicionales. Se

sugiere que futuras investigaciones se enfoquen en el ajuste de hiperparámetros y en la exploración de técnicas alternativas de optimización y modelado, con el objetivo de seguir mejorando la precisión en la predicción de caudales en cuencas complejas como la del río Medellín, Colombia.

### **Palabras clave**

- LSTM (Long Short-Term Memory): Tipo de red neuronal recurrente diseñada para aprender patrones en secuencias de datos, capaz de retener información durante períodos prolongados, lo que la hace adecuada para tareas de predicción en series temporales.
- PSO (Optimización por Enjambre de Partículas): Algoritmo de optimización que simula el comportamiento social de grupos de partículas, utilizado para encontrar soluciones óptimas en problemas complejos mediante la exploración colaborativa del
- KGE (Kling-Gupta Efficiency): Indicador que evalúa la eficiencia de un modelo hidrológico al comparar las simulaciones con datos observados, considerando la correlación, la sesgo y la variabilidad, proporcionando una visión integral del desempeño del modelo.
- IDEAM: Instituto de Hidrología, Meteorología y Estudios Ambientales de Colombia, responsable de la recopilación y gestión de información ambiental, así como de la investigación en temas relacionados con el clima y los recursos hídricos.
- Modelo TETIS: Herramienta de simulación hidrológica que permite modelar el comportamiento del agua en cuencas, integrando diferentes variables como precipitación, caudal y evapotranspiración para predecir flujos y gestionar recursos hídricos.

Introducción, Marco teórico o de referencia

## **1.Introducción**

Este proyecto se centra en la calibración de un modelo hidrológico utilizando técnicas avanzadas de optimización y aprendizaje automático, específicamente redes neuronales LSTM (Long Short-Term Memory) y PSO (Optimización por Enjambre de Partículas). La necesidad de modernizar los modelos hidrológicos en las cuencas de los Andes colombianos surge de las complejas características geográficas y las extremas variaciones climáticas que afectan la gestión sostenible de los recursos hídricos.

El objetivo principal es desarrollar un enfoque más eficiente y escalable que permita mejorar la precisión en la simulación de caudales y crecientes. Para lograrlo, se llevará a cabo una revisión exhaustiva de la literatura existente, seguida de la automatización del preprocesamiento de datos hidrológicos y la implementación del modelo TETIS en Python. La calibración del modelo se realizará mediante el uso de PSO, buscando minimizar el error en las simulaciones. Este enfoque no solo optimiza el análisis hidrológico, sino que también reduce los errores humanos y permite un manejo más eficaz de grandes volúmenes de datos, contribuyendo así a una mejor planificación y gestión de los recursos hídricos en la región.

## 2. Marco Teórico

### 2.1. Modelos Hidrológicos

Para lograr gestionar de manera oportuna y efectiva los recursos hídricos, es necesario el uso de herramientas que comprendan, entiendan y pronostiquen el comportamiento de los ecosistemas hídricos. Los sistemas hidrológicos como las cuencas de drenaje son elementos aislados por planos horizontales y divisorias de aguas verticales, que deben ser abordados, mediante la abstracción simple de un sistema complejo, analizando el mundo real bajo ciertas condiciones que personifican los procesos dentro de una cuenca, logrando describir algunos de los componentes más importantes y básicos del sistema que simula (Pascual Aguilar & Díaz Martín, 2016). Estos modelos se definen como:

- Modelos basados en la relación de variables de entrada(input) y de salida(output), los cuales contemplan las características del medio (topografía, uso del suelo) (Mohammadi & Mehdizadeh, 2020)

- Modelos que representan matemáticamente los procesos hidrológicos de una cuenca, simulan la transferencia de lluvia-escorrentía y consideran factores como infiltración, evapotranspiración y almacenamiento del agua (Beven, 2020)

Por esto entendemos que los modelos son una representación del concepto físico de los procesos hidrológicos que suceden en una zona de estudio, y son los modelos computacionales quienes describen matemáticamente lo interpretado mediante simulaciones condicionadas a las limitaciones variables del medio.

Por consiguiente, la finalidad de un modelo hidrológico es predecir y estudiar la dinámica del sistema, enfocado a las entradas y salidas de las interacciones de los distintos planos que limitan la cuenca. Estos desempeñan un papel crucial, de acuerdo con Kratzert et al. (2019) en;

- (1) El impacto del cambio climático en los sistemas acuáticos.
- (2) La gestión de recursos hídricos.
- (3) Simulando la dinámica de sistemas hídricos en condiciones variables.
- (4) La comprensión del actuar de los componentes del ciclo hidrológico.
- (5) Conocer potenciales condiciones futuras.
- (6) La mitigación de desastres naturales y la organización de la agricultura.

Es así como los modelos hidrológicos satisfacen diversas necesidades de acuerdo con los requerimientos, existen muchas las tipologías en las que se pueden clasificar los modelos, en este trabajo seguimos los lineamientos de (Pascual Aguilar & Díaz Martín, 2016) basados en modelos matemáticos del tipo procedimental y espacial para clasificar los modelos hidrológicos en la literatura.

-Modelos materiales; Son la encarnación de un sistema por otro sistema con muchas similitudes, pero más sencillo de ejecutar que el prototipo, ya que la escala puede ser ajustada según las capacidades del modelador y al ser modelo análogo se establece correspondencia física.

-Modelos simbólicos; Es una representación abstracta de una situación idealizada, bastante simple, que comparte las características fundamentales del sistema original, expresadas de 2 maneras;

- Matemáticos: Donde el comportamiento del sistema se describe mediante un conjunto de ecuaciones, que pueden estar acompañadas de afirmaciones lógicas que establecen las conexiones entre las variables y los parámetros.

- No matemáticos: Donde un esquema de flujo muestra los componentes y procesos involucrados en el ciclo hidrológico de una cuenca hidrográfica.

- Modelos Basado en procesos (físicos); basado en ecuaciones físicas que logran sintetizar los principios hidrológicos del mundo real, usado en situaciones de condiciones cambiantes. Como el modelo SHE (System Hydrologique Europeen) Fatichi et al. (2016) Requieren datos detallados para un discernimiento de la evolución de la cuenca.

- Modelos Empíricos; basado en los lazos estadísticos que surgen directamente de las observaciones, fundamentado en cálculos numéricos derivados de la investigación experimental, donde los datos históricos establecen relaciones entre la lluvia y la escorrentía sin tener en cuenta explícitamente los procesos físicos subyacentes. (Demirezen & Kazezyilmaz-Alhan, 2022)

- Modelos Conceptual; basado en la abstracción simple de los procesos hidrológicos a través de reservorios (tanques) interconectados que simulan el almacenamiento y el flujo del agua en una cuenca, modelos intermedios entre los empíricos y teóricos. Como el modelo HBV Beck et al. (2020)

- Modelos de Aprendizaje Automático; ha surgido como una nueva tendencia en la modelización hidrológica, basado en algoritmos de aprendizaje automático, que establecen relaciones entre entradas y salidas sin especificar explícitamente los procesos físicos, como los modelos LSTM (Long Short-Term Memory) aplicados a hidrología (Kratzert y otros, 2021)

- Modelo Lineal; Un ejemplo clásico de un modelo lineal es el hidrograma unitario, que simplifica los procesos mediante una secuencia lineal entre las entradas, los procesos intermedios y las salidas. Asumen una relación lineal entre las variables de entrada y salida. Aunque menos comunes en la hidrología moderna, Tyralis et al. (2021) demostraron que los modelos lineales simples pueden servir como líneas de base útiles para evaluar modelos más complejos.

- Modelo No Lineal; Por otro lado, cuando se intenta abordar la mayor complejidad de los procesos físico-hidrológicos, se recurre a modelos no lineales. Capturan relaciones más complejas entre variables. La mayoría de los modelos hidrológicos modernos son no lineales. Por ejemplo, Frame et al. (2022) utilizaron técnicas de aprendizaje profundo no lineal para mejorar las predicciones de caudales en cuencas no aforadas, superando significativamente a los modelos lineales tradicionales.

- Modelo Temporal Variante o Eventos; Consideramos que un modelo es invariante en el tiempo cuando su relación entre entradas y salidas no cambia con el paso del tiempo, donde la forma de la salida depende únicamente de la forma de la entrada y no del momento en que se aplica la entrada. Se centran en simular eventos individuales de lluvia-escorrentía. Zhu et al. (2020) desarrollaron un modelo de eventos basado en

redes neuronales para predecir inundaciones repentinas en áreas urbanas, logrando una precisión significativamente mayor que los métodos tradicionales.

- Modelo Temporal No Variante o Continuos; Por el contrario, los modelos que no cumplen con esta propiedad se consideran variables en el tiempo. La mayoría de los sistemas hidrológicos son sensibles a las variaciones temporales, aunque para simplificar, algunos modelos asumen condiciones de invariabilidad temporal. Simulan períodos más largos, capturando la variabilidad temporal de los procesos hidrológicos. Rakovec et al. (2016) utilizaron un modelo continuo para evaluar los impactos del cambio climático en los recursos hídricos a escala continental, destacando la importancia de las simulaciones a largo plazo para la gestión adaptativa del agua.

- Modelos Espaciales Agregados; Un modelo es agregado cuando, sin importar los procesos que considere, los parametriza como valores generales para la cuenca en su conjunto, sin considerar su diversidad espacial. Tratan toda la cuenca como una unidad homogénea. Gao et al. (2019) demostraron que los modelos agregados pueden ser sorprendentemente efectivos en cuencas de tamaño medio, logrando un rendimiento comparable a modelos más complejos en términos de NSE y KGE.

- Modelos Espaciales Distribuidos; Se considera distribuido al modelo que busca representar con precisión los procesos y parámetros en cuestión, considerando su variabilidad espacial y temporal. Dividen la cuenca en unidades más pequeñas, permitiendo una representación más detallada de la variabilidad espacial. Fatichi et al. (2016) argumentan que los modelos distribuidos son esenciales para capturar los impactos localizados del cambio climático y los cambios en el uso del suelo. Sin

embargo, también señalan que el aumento de la complejidad no siempre se traduce en mejores predicciones, especialmente en cuencas con escasez de datos

- Modelos Espaciales Semidistribuidos; El modelo semidistribuido es aquel que ofrece una representación intermedia de la diversidad espacial, dividiendo la cuenca en zonas de comportamiento similar que se consideran como unidades agregadas.

- Modelos Determinísticos; No consideran la incertidumbre en la formulación matemática, una entrada determinada produce siempre una salida predecible. Producen una única salida para un conjunto dado de entradas y parámetros. Shin et al. (2015) desarrollaron un modelo determinístico basado en procesos para simular la dinámica del carbono en ecosistemas terrestres, demostrando su utilidad en la predicción de impactos del cambio climático.

- Modelos Estocásticos; Consideran que presentan una componente aleatoria, se utiliza cuando la variabilidad es significativa. Incorporan elementos aleatorios para representar la incertidumbre. Flores et al. (2019) propusieron un enfoque estocástico para la predicción de caudales, combinando modelos físicos con técnicas de aprendizaje automático para cuantificar y reducir la incertidumbre en las predicciones.

Es importante destacar que estas categorías no son mutuamente excluyentes, y muchos modelos hidrológicos modernos combinan características de múltiples categorías, por lo cual, en los últimos años, ha habido una tendencia hacia la integración de diferentes enfoques de modelado, buscando combinar las fortalezas de distintos tipos de modelos para mejorar la comprensión y predicción de los procesos hidrológicos. Nearing et al.

(2021) proponen un enfoque híbrido que integra modelos basados en procesos con técnicas de aprendizaje automático, buscando aprovechar las fortalezas de ambos enfoques para mejorar las predicciones hidrológicas.

De igual manera es notario el crecimiento hacia el uso de modelos más complejos y distribuidos, impulsada por la mayor disponibilidad de datos y el aumento de la capacidad computacional. Sin embargo, también se ha reconocido la importancia de la parsimonia en el modelado hidrológico, lo que ha llevado a un renovado interés en modelos conceptuales más simples pero robustos, (Fowler et al. 2018). El uso de estos modelos ha evolucionado significativamente, incorporando avances en tecnología computacional, técnicas de aprendizaje automático y una comprensión más profunda de los procesos hidrológicos (Addor & Melsen, 2019).

## **2.2. Interacción de modelos hidrológicos por medio de modelos de machine Learning y de técnicas de optimización.**

El progreso significativo en la integración de modelos hidrológicos tradicionales con técnicas avanzadas de Machine Learning (ML), ha hecho que los modelos lluvia escorrentía evolucionen significativamente. Bouaziz et al. (2020) realizaron un estudio comparativo de varios modelos conceptuales, incluyendo versiones actualizadas de modelos de tanques como el utilizado en TETIS. Los autores destacan la importancia de la representación adecuada de los procesos hidrológicos en diferentes escalas temporales y espaciales, enfatizando la necesidad de modelos flexibles que puedan adaptarse a condiciones cambiantes.

Es así como Gauch et al. (2021) introdujeron el concepto de "Redes Neuronales Basadas en Procesos" (PBNN, por sus siglas en inglés), que combinan la flexibilidad de las redes neuronales con la interpretabilidad de los modelos conceptuales. Este enfoque híbrido permite incorporar conocimientos hidrológicos previos en la estructura de la red neuronal, mejorando tanto el rendimiento como la interpretabilidad del modelo.

En la revisión de Tikhamarine et al. (2020) se propusieron métodos de inteligencia artificial, integrando optimización por enjambre de partículas (PSO) y el optimizador de halcones de Harris (HHO) para mejorar la precisión en la simulación de la escorrentía. Este enfoque permitió seleccionar las mejores combinaciones de entrada y preprocesar los datos de manera efectiva, reduciendo los errores en las predicciones.

En el ámbito del aprendizaje automático, las redes neuronales han revolucionado el modelado de procesos hidrológicos. En particular, Kratzert et al. (2018) demostraron la superioridad de las redes LSTM en la predicción de caudales en comparación con los modelos hidrológicos tradicionales, utilizando un conjunto de datos de cuencas de los Estados Unidos. Su trabajo sentó las bases para una nueva generación de modelos hidrológicos basados en datos.

Es por eso por lo que, para alcanzar la calibración más versátil, debemos emplear la sinergia que preside en la optimización y las predicciones, ya que la optimización de los parámetros del modelo es crucial para mejorar su rendimiento. El algoritmo de optimización por enjambre de partículas (PSO) ha sido ampliamente utilizado debido a su simplicidad y eficacia. PSO simula el comportamiento de un grupo de partículas que exploran el espacio de búsqueda para encontrar la solución óptima (Tikhamarine y otros,

2020). En la predicción del flujo de corriente, el uso de PSO ha demostrado mejorar la precisión de los modelos al optimizar los parámetros clave. Rafael-Miñope et al. (2022) también mencionan la importancia de la hibridación de algoritmos, combinando RNA y PSO para generar modelos que ofrecen mejores estimaciones en el modelado hidrológico.

### **2.3 Modelo de Machine Learning LSTM para predicción**

Entre las técnicas avanzadas de aprendizaje automático, las Redes Neuronales Recurrentes (RNN) y particularmente las Long Short-Term Memory (LSTM), han emergido como herramientas de vanguardia en la predicción de caudales. La eficacia de las LSTM en este dominio radica en su capacidad para desentrañar las inconvenientes dependencias temporales inherentes a los procesos hidrológicos, un atributo crucial para la comprensión y pronóstico de sistemas fluviales complejos (Frame y otros, 2022)

La arquitectura LSTM, un refinamiento sofisticado de las RNN convencionales, se distingue por su estructura modular compuesta por celdas de memoria interconectadas. Estas unidades funcionales operan como repositorios dinámicos de información, dotadas de mecanismos de compuerta que regulan el flujo y la persistencia de datos relevantes. Wang et al. (2017) elucidan cómo esta configuración permite a las LSTM discernir y preservar patrones significativos a múltiples escalas temporales, desde fluctuaciones efímeras hasta tendencias seculares en las series hidrológicas.

La superioridad de las LSTM en el modelado hidrológico no es meramente teórica. Investigaciones empíricas han corroborado su eficacia en diversos contextos geoclimáticos. La versatilidad de las LSTM se manifiesta en su capacidad para procesar

secuencias de longitud variable, una característica particularmente ventajosa en el análisis de series temporales hidrológicas irregulares o con lagunas de datos. Además, su resistencia intrínseca al problema del desvanecimiento del gradiente, un obstáculo común en el entrenamiento de redes neuronales profundas las posiciona como una opción preferencial para modelar procesos hidrológicos con memoria a largo plazo (Wang y otros, 2017)

En síntesis, la confluencia de una arquitectura neuronal sofisticada con la complejidad inherente de los sistemas hidrológicos ha catapultado a las LSTM al frente de la investigación en modelación de caudales. Su capacidad para capturar y representar las sutilezas de los procesos hidrológicos promete no solo mejorar la precisión de las predicciones, sino también profundizar nuestra comprensión de los mecanismos subyacentes que gobiernan el comportamiento de los sistemas fluviales.

Es así como las redes neuronales (LSTM) han mostrado un rendimiento excepcional en la predicción de series temporales hidrológicas (Mohammadi B. , 2021). LSTM es capaz de aprender patrones de largo plazo y manejar la no linealidad inherente a los procesos hidrológicos. El estudio de Mohammadi & Mehdizadeh (2020) aplicó un modelo LSTM junto con un algoritmo de optimización de ballenas (WOA) para la estimación de la evapotranspiración diaria. Los resultados mostraron una mejora significativa en la precisión de las predicciones en comparación con métodos tradicionales, destacando la capacidad del modelo LSTM para manejar la complejidad y la variabilidad en los datos.

## **2.4 Optimización de Parámetros con PSO (Particle Swarm Optimization)**

La calibración de modelos hidrológicos constituye un proceso fundamental para garantizar la fidelidad y aplicabilidad de las simulaciones en el ámbito de la gestión hídrica. En este contexto, el algoritmo de Optimización por Enjambre de Partículas (PSO, por sus siglas en inglés) ha emergido como un paradigma de vanguardia en la optimización paramétrica. Este método, inspirado en la dinámica social de agregaciones animales, como las bandadas aviares o los cardúmenes ictiológicos, ha demostrado una singular eficacia en la resolución de problemas de optimización caracterizados por su no linealidad y multimodalidad, atributos intrínsecos a la calibración de modelos hidrológicos complejos (Gad, 2022).

El modus operandi del PSO se fundamenta en la evolución dinámica de un conjunto de soluciones potenciales, metafóricamente denominadas "partículas", que transitan por un hiperespacio de búsqueda multidimensional. Cada entidad algorítmica modula su trayectoria basándose en una sinergia entre su historial individual de desempeño y la información colectiva del enjambre. Esta heurística de búsqueda exhibe una notable eficiencia computacional y una robustez significativa frente a la convergencia prematura en óptimos locales, cualidades que lo posicionan como una herramienta idónea para la calibración de modelos hidrológicos con una plétora de parámetros interrelacionados (Zhang y otros, 2021).

La implementación del PSO en el dominio de la modelación hidrológica ha catalizado avances significativos en la precisión y confiabilidad de las predicciones. Gad (2022)

elucida cómo la naturaleza estocástica del algoritmo facilita una exploración exhaustiva del espacio paramétrico, permitiendo la identificación de configuraciones óptimas que podrían eludir métodos de optimización más convencionales. Esta capacidad resulta particularmente valiosa en la calibración de modelos que simulan procesos hidrológicos altamente no lineales, donde la topografía de la función objetivo puede ser extremadamente compleja.

La versatilidad del PSO se manifiesta en su adaptabilidad a diversos esquemas de modelación hidrológica. Zhang et al. (2021) detallan cómo el algoritmo puede ser eficazmente aplicado tanto a modelos conceptuales simples como a representaciones físicamente basadas de alta complejidad. Los investigadores subrayan la flexibilidad del PSO para incorporar múltiples criterios de desempeño, permitiendo una calibración multiobjetivo que balancee diferentes aspectos del comportamiento hidrológico, como los flujos base y los eventos de crecida.

En síntesis, la integración del PSO en el proceso de calibración de modelos hidrológicos representa un avance significativo en la ciencia hidrológica computacional. Su capacidad para navegar eficientemente espacios de búsqueda complejos, combinada con su robustez y eficiencia algorítmica, lo posiciona como una herramienta invaluable en la optimización de modelos predictivos del ciclo hidrológico. La adopción generalizada de esta técnica promete no solo mejorar la precisión de las simulaciones hidrológicas, sino también profundizar nuestra comprensión de los sistemas hídricos a través de una parametrización más refinada y representativa de los procesos físicos subyacentes.

## 2.5 Métricas de calibración

La cuantificación del desempeño en la modelación hidrológica sea esta basada en enfoques tradicionales o en paradigmas de aprendizaje automático, constituye un aspecto crítico en la validación y optimización de los modelos predictivos. Este proceso de evaluación, que abarca tanto la calibración de modelos conceptuales como la valoración de algoritmos de aprendizaje automático, pivota fundamentalmente sobre la selección e implementación judiciosa de métricas de rendimiento. Estas métricas no solo funcionan como indicadores cuantitativos de la precisión y fiabilidad de las predicciones, sino que también desempeñan un papel instrumental en la guía del proceso de optimización paramétrica y en la validación holística de los modelos (Kratzert y otros, 2019)

En el contexto de los modelos hidrológicos tradicionales, las métricas de rendimiento sirven como funciones objetivo en el proceso de calibración, guiando la optimización de los parámetros del modelo hacia configuraciones que minimicen la discrepancia entre las predicciones y las observaciones. Este proceso iterativo de ajuste paramétrico se beneficia enormemente de la selección de métricas que reflejen adecuadamente las prioridades de modelación. Paralelamente, en el dominio de los modelos basados en aprendizaje automático, las métricas de rendimiento desempeñan un papel dual: por un lado, guían el proceso de entrenamiento a través de funciones de pérdida; por otro, proporcionan una base para la evaluación comparativa y la validación del modelo. La complejidad inherente a muchos algoritmos de ML, particularmente aquellos basados en redes neuronales profundas, acentúa la importancia de emplear un conjunto diverso de métricas que capture diferentes aspectos del rendimiento del modelo. Esto puede incluir

métricas específicas para evaluar la capacidad del modelo para generalizar a condiciones no vistas o para capturar la incertidumbre en sus predicciones.

La calibración de modelos predictivos se fundamenta en un conjunto de métricas estadísticas que permiten cuantificar la concordancia entre las simulaciones y las observaciones empíricas. Estas métricas, lejos de ser meros indicadores numéricos, constituyen el lenguaje formal a través del cual se articula la eficacia y fiabilidad de los modelos hidrológicos. Deng et al. (2021) identifican un cuarteto de métricas que han alcanzado una prominencia significativa en la praxis hidrológica contemporánea, cada una de ellas iluminando facetas distintivas del rendimiento modelístico.

El Coeficiente de Eficiencia de Nash-Sutcliffe (NSE) se erige como un paradigma evaluativo que cuantifica la capacidad predictiva del modelo en relación con la simplicidad de un predictor basado en la media observacional. Este coeficiente, oscilando en un rango de  $-\infty$  a 1, proporciona una medida normalizada de la discrepancia entre las predicciones y las observaciones, donde valores cercanos a la unidad indican una concordancia óptima. La universalidad de su aplicación en hidrología se debe, en parte, a su sensibilidad a los eventos de alto caudal, una característica particularmente relevante en estudios de inundaciones y gestión de recursos hídricos.

El Coeficiente de Determinación ( $R^2$ ) ofrece una perspectiva complementaria, cuantificando la proporción de variabilidad en los datos observados que es explicada por el modelo. [Este estadístico, acotado entre 0 y 1, proporciona una medida intuitiva de la fuerza de la relación entre las predicciones y las observaciones. No obstante, su

insensibilidad a los sesgos sistemáticos exige una interpretación cautelosa y su uso en conjunción con otras métricas para una evaluación comprehensiva.]

El Error Cuadrático Medio (RMSE), por su parte, ofrece una cuantificación directa de la magnitud promedio de los errores de predicción. Esta métrica, expresada en las mismas unidades que la variable de interés, penaliza desproporcionadamente los errores grandes debido a su naturaleza cuadrática. Su utilidad es particularmente notable en la comparación de modelos aplicados a la misma cuenca y escala temporal, proporcionando una medida absoluta de la precisión predictiva.

La Kling-Gupta Efficiency (KGE) es una métrica estadística diseñada para evaluar el desempeño de modelos hidrológicos, superando algunas limitaciones de métricas tradicionales como el Coeficiente de Eficiencia de Nash-Sutcliffe (NSE) y el Coeficiente de Determinación ( $R^2$ ). Introducida por Harald Kling y Hoshin Vijai Gupta en 2009, la KGE proporciona una evaluación más completa al considerar no solo la precisión de las predicciones, sino también su variabilidad y sesgo.

La aplicación sinérgica de estas métricas permite una evaluación holística del rendimiento de los modelos capturando tanto la precisión global como aspectos específicos de la dinámica hidrológica. Este enfoque facilita una calibración más robusta y una validación más rigurosa de los modelos y también proporciona una base sólida para la comparación entre diferentes enfoques de modelación.

### **Planteamiento del problema.**

La modelización hidrológica en las cuencas andinas representa un paradigma de complejidad creciente en la ciencia hidrológica contemporánea. La intrincada interacción entre la variabilidad climática y la diversidad geográfica de estas regiones engendra una serie de desafíos metodológicos y epistemológicos en la conceptualización y operacionalización de modelos lluvia-escorrentía fidedignos. La fluctuación inherente a los resultados obtenidos y la incertidumbre asociada a las predicciones constituyen obstáculos fundamentales que socavan la eficacia de la planificación y gestión de los recursos hídricos en estos ecosistemas críticos (Beven, 2019).

La escasez de series de datos completas y la imprecisión intrínseca a la medición remota de flujos hídricos exacerban la complejidad del proceso de calibración de estos modelos, lo que resalta la necesidad de enfoques innovadores que integren técnicas avanzadas de aprendizaje automático (Rafael-Miñope, 2022). En respuesta a esta problemática, la comunidad científica ha adoptado metodologías innovadoras como GLUE (Estimación de Incertidumbre de Verosimilitud Generalizada), que propugna la identificación de múltiples conjuntos de parámetros congruentes con los datos observados, en contraposición a la búsqueda de una solución única y determinista. Este enfoque paradigmático subraya la imperativa necesidad de desarrollar estrategias de calibración más robustas y adaptativas (Beven, 2019).

La heterogeneidad topográfica y la diversidad climática características de las regiones andinas introducen variabilidad en los resultados de los modelos hidrológicos formando capas adicionales de incertidumbre que permiten la equifinalidad, donde diferentes

conjuntos de parámetros pueden producir resultados similares, son obstáculos que requieren estrategias de calibración efectivas (Beven, 2019; Mohammadi, 2021).

Las técnicas tradicionales de modelado han evidenciado limitaciones significativas al enfrentarse a la dinámica no lineal, en este contexto, (Mohammadi B. A., 2021) postulan que la integración del aprendizaje automático tiene el potencial de mejorar sustancialmente la capacidad de capturar relaciones no lineales y manipular grandes volúmenes de datos. También logran mejorar la precisión de las predicciones (Rafael-Miñope et al. 2022). No obstante, la implementación eficaz de estas metodologías requiere la aplicación de métodos sofisticados de análisis y calibración para optimizar la precisión y eficiencia de las simulaciones (Beven, 2019).

Según Mohammadi et al. (2020) argumentan de manera convincente que la sinergia entre las técnicas de aprendizaje automático y los modelos hidrológicos tradicionales permite la identificación y caracterización de patrones complejos que eluden la capacidad de detección de las metodologías convencionales. Enfrentando desafíos limitantes, en los cuales se encuentran:

1. **Preprocesamiento de Datos:** La heterogeneidad de los datos, que provienen de diversas fuentes y formatos, dificulta su integración y uso en modelos de aprendizaje automático. Esta falta de automatización en esta etapa puede introducir errores y reducir la eficiencia del modelo, como señala Mohammadi et al. (2020). Además, Tikhamarine et al. (2020) destaca que la inconsistencia en los datos puede llevar a resultados imprecisos y a una interpretación errónea de los patrones subyacentes.

2. **Selección de Modelos:** La elección del modelo de aprendizaje automático más adecuado es fundamental y no existe un consenso claro sobre cuál es el mejor para la descripción del modelo lluvia-escorrentía (Tikhamarine et al. 2020); Mohammadi, et al. 2020).
3. **Calibración de Modelos:** Los métodos tradicionales de calibración son laboriosos y a menudo no capturan la complejidad de los sistemas hidrológicos, especialmente en cuencas montañosas (Humphrey y otros, 2016). La implementación de técnicas de aprendizaje automático para la calibración automatizada podría mejorar la precisión de las simulaciones, aunque la falta de un enfoque estandarizado limita su adopción (Tikhamarine, y otros, 2020).

Las innovaciones recientes en el ámbito de los algoritmos de optimización y aprendizaje automático han desbloqueado nuevos horizontes en la superación de las limitaciones impuestas por los modelos hidrológicos clásicos. En una demostración empírica, Tikhamarine et al. (2020) probó que la integración de metodologías avanzadas, como el Optimizador de Harris Hawks y la Optimización Gray Wolf, propicia un aumento significativo en la precisión de los modelos de lluvia-escorrentía. Estos resultados evidencian la necesidad imperativa de adoptar enfoques innovadores para el análisis de cuencas complejas, fundamentales en la realidad andina.

De manera similar, por otra parte, en una línea de investigación paralela pero complementaria, Mohammadi & Mehdizadeh (2020) aplicaron una metodología combinada del mismo algoritmo con la regresión por soporte vectorial y lograron desarrollar una solución eficaz. De hecho, esa metodología híbrida había revolucionado

el campo de la modelización de la evapotranspiración, un aspecto central del balance hídrico de las cuencas. Dada la probada eficacia de las metodologías híbridas, se podía inferir un futuro prometedor para el campo de la modelización hidrológica de cuencas de alta complejidad.

La necesidad de mejorar la precisión en la predicción de caudales y escorrentías iba más allá de las fronteras de las cuencas andinas, constituyendo un desafío de alcance global. Mohammadi B. A. (2021) realizó una revisión exhaustiva sobre las aplicaciones del aprendizaje automático en la modelización de la escorrentía, destacando la implementación de estos métodos en diversas regiones del mundo para abordar problemas similares de incertidumbre y variabilidad en los modelos hidrológicos. Esta revisión sistemática reforzaba la importancia de continuar desarrollando y adaptando estas técnicas avanzadas para optimizar la gestión de recursos hídricos en entornos de alta complejidad a nivel mundial.

Hasta el momento, la selección del método más adecuado para cada entorno específico sigue siendo un problema vertiginoso debido a la variabilidad de los resultados entregados por las técnicas. La investigación inaugurada por Tikhamarine et al. (2020) sobre la optimización de los algoritmos de IA para el pronóstico de las descargas mensuales a través de la aplicación de Grey Wolf ha dejado en claro la dependencia crítica de las predicciones realizadas sobre la elección del algoritmo, que debe hacerse según la propia naturaleza de los datos de entrada y los rasgos característicos de las cuencas individuales.

Esta necesidad de personalización y ajuste fino de los modelos se erige como un requisito para mejorar la precisión y confiabilidad de las predicciones, especialmente en regiones montañosas donde la escasez y dificultad de obtención de datos constituyen obstáculos adicionales.

En el contexto de las cuencas andinas, la implementación de modelos hidrológicos de alta precisión se volvió crucial para la gestión sostenible de los recursos hídricos. De acuerdo con Mohammadi et al. (2020) la simulación de las fluctuaciones en el nivel del agua del lago Titicaca, mediante técnicas híbridas de aprendizaje automático y optimización, había permitido una representación más precisa de las dinámicas hidrológicas en cuencas de alta montaña. Estos hallazgos eran fundamentales para el desarrollo de herramientas de predicción más robustas, capaces de adaptarse a las condiciones cambiantes del clima y la geografía características de la región andina.

A medida que la variabilidad climática y las demandas sobre los recursos hídricos aumentan, la necesidad de herramientas más precisas y eficientes para la gestión de cuencas se vuelve cada vez más urgente. El desarrollo de un modelo hidrológico automatizado en Python, como se propone en esta investigación, tiene el potencial de mejorar significativamente la predicción de caudales proporcionando una herramienta más robusta para la gestión de recursos hídricos en cuencas complejas (Mohammadi B. A., 2021). La pregunta de investigación que guía este estudio es: ¿Qué modelos de ML arrojan la mejor descripción del modelo lluvia-escorrentía?, y se enmarca en la necesidad de identificar las técnicas de ML que ofrezcan la mayor precisión y aplicabilidad en la

simulación hidrológica colombiana (Tikhamarine et al., 2020; Mohammadi et al., 2020; Mohammadi, 2021).

## **Objetivos**

### **Objetivo General**

Desarrollar un modelo hidrológico automatizado en Python para mejorar la predicción de caudales superficiales mediante técnicas de Machine Learning (ML).

### **Objetivos específicos**

- Crear un modelo hidrológico TETIS agregado de tanques en Python
- Crear un algoritmo en Python que automatice el preprocesamiento de datos hidrológicos para simplificar la entrada de información al modelo.
- Seleccionar el modelo más adecuado conforme a la naturaleza de los datos de entrada para la optimización de parámetros
- Implementar un sistema de calibración del modelo utilizando técnicas de Machine Learning para mejorar la precisión en la simulación de caudales

## Metodología

Este proyecto tiene como objetivo la calibración de un modelo hidrológico utilizando técnicas de optimización y aprendizaje automático, específicamente redes neuronales LSTM (Long Short-Term Memory) y PSO (Optimización por Enjambre de Partículas). A continuación, se detallan los métodos empleados para alcanzar los objetivos generales y específicos del proyecto en la tabla 1.

*Tabla 1*  
*Metodología*

Actividad/Paso	Métodos	Técnicas de Recolección de Datos	Producto
<b>Actividad 1: Revisión de la literatura</b>	Revisión exhaustiva de la literatura sobre modelos de calibración hidrológica, optimización y aprendizaje automático. Análisis de técnicas de imputación de datos y algoritmos de optimización.	Revisión de bases de datos científicas y reportes técnicos.	Definición de conceptos clave, selección de modelos (LSTM, PSO) y elección de métricas relevantes (RMSE).
<b>Paso 1: Algoritmo de preprocesamiento</b>	Uso de librerías (`numpy`, `pandas`, `sklearn`, `matplotlib`) para procesar datos de precipitación. Entrenamiento de red LSTM para imputar valores faltantes. Evaluación mediante RMSE.	Procesamiento de datos en conjuntos de entrenamiento, validación y prueba.	Entrenamiento y evaluación de red LSTM, gráficas de pérdida del modelo.
<b>Actividad 2: Automatización de preprocesamiento</b>	Desarrollo de un algoritmo automatizado en Python para extracción de series de precipitación y caudal de bases de datos.	Extracción de datos de bases IDEAM y AMVA, selección de series clave.	Algoritmos desarrollados, red LSTM entrenada, archivo `.xlsx` con datos confiables.
<b>Paso 2: Creación del modelo TETIS</b>	Implementación del modelo TETIS en Python. Definición de parámetros y simulación de caudales, visualización mediante `matplotlib`.	Procesamiento y simulación de datos hidrológicos.	Modelo TETIS implementado, gráficas de resultados de simulación.

<b>Actividad 3: Calibración del modelo TETIS</b>	Uso de PSO para calibración del modelo TETIS, minimización del error KGE. Optimización de parámetros mediante `pyswarms`.	Validación con caudales simulados y observados, utilizando el KGE como métrica de evaluación.	Gráficas de simulación del modelo TETIS, validación detallada del método de calibración.
<b>Paso 3: Optimización de parámetros del modelo</b>	Definición de la arquitectura del algoritmo PSO para optimización de parámetros, visualización de resultados.	Optimización de parámetros del modelo TETIS, minimización del error KGE.	Parámetros optimizados, gráficas que visualizan resultados de la optimización.

Nota: Elaboración propia

## Resultados y Discusión

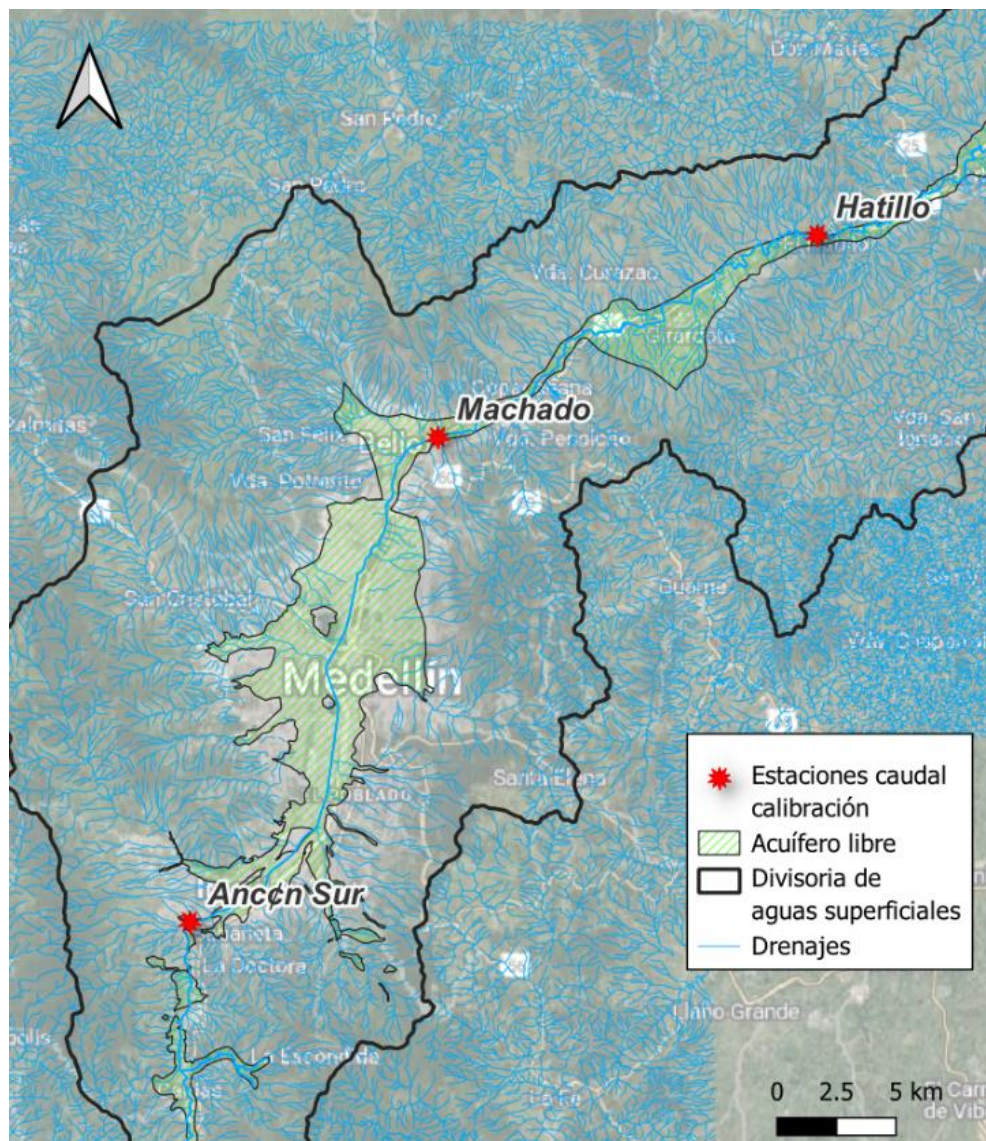
El modelo agregado conceptual TETIS, ha sido creado en el IDE de Colab con el lenguaje de programación Python, la simulación se implementó por efectos académicos en una sola cuenca, la cuenca de Río Medellín ubicado en el corazón de Colombia, en el departamento de Antioquia, convertido en un referente a nivel nacional e internacional en cuanto a la aplicación de modelos hidrológicos y simulaciones por su compleja dinámica, influenciada por factores naturales y antropogénicos. Escenario ideal para el desarrollo y validación de herramientas de predicción y gestión de recursos hídricos, esta cuenta con una orografía compleja andina y posee una instrumentación bastante completa, que facilita la obtención de las principales variables de entrada para la simulación, De acuerdo con la figura 1.



**Figura 1** Monitoreo RedRío, Aburrá-Medellín, desde su nacimiento hasta su confluencia.

Fuente: AMVA (<https://www.metropol.gov.co/ambiental/Paginas/RedRio.aspx>)

Se selecciono la estación Ancón-sur como estación de caudal, porque de acuerdo con la figura 2, esta cuenta con un porcentaje menor en la cobertura urbana.



**Figura 2** Estaciones de caudal existentes en la delimitación de la cuenca Río Medellín

Se visualizan las estaciones pluviométricas más afluentes para la cuenca piloto de acuerdo con la figura 3.

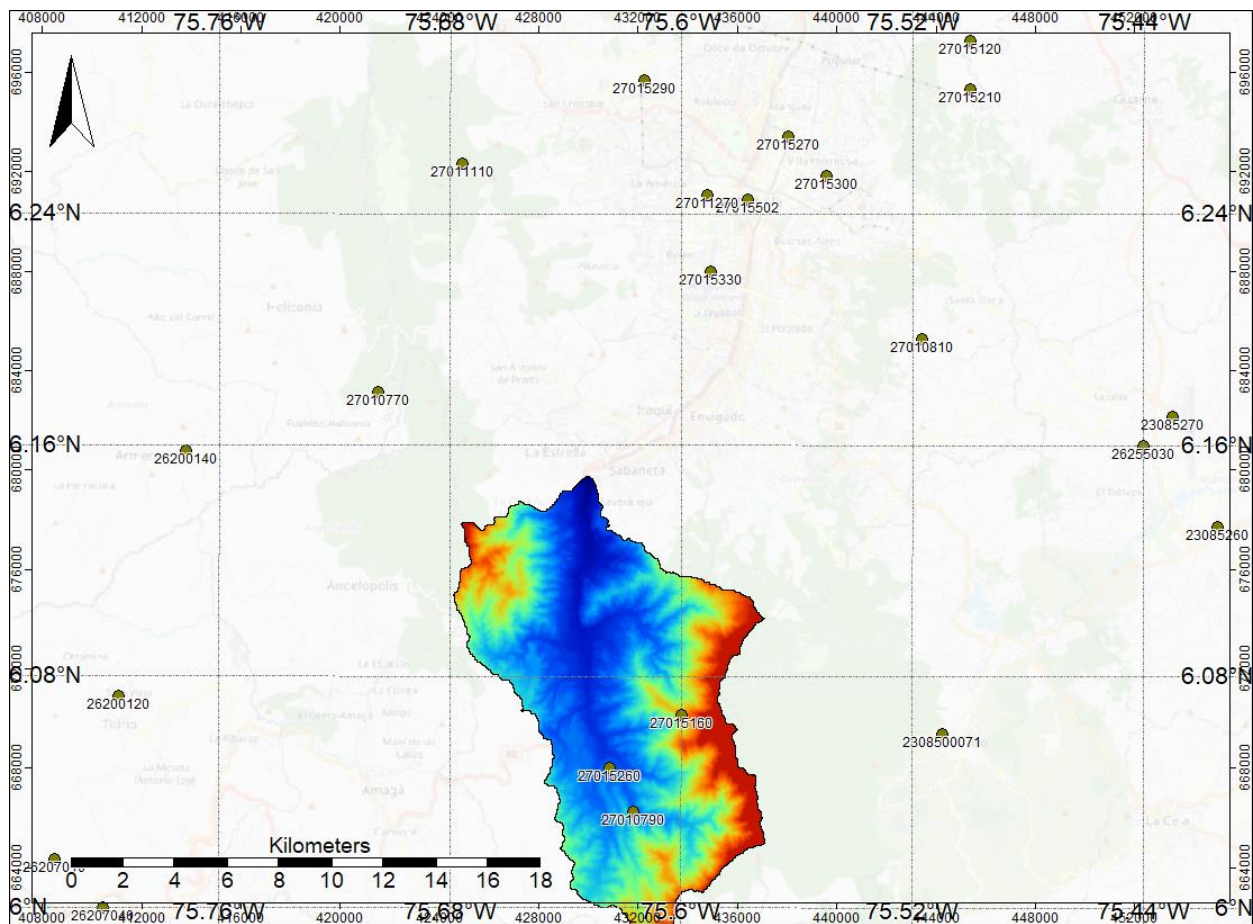


Figura 3 Estaciones pluviométricas

Se obtuvieron los datos del balance hídrico por medio del informe suministrado por la AMVA (2016), presentados en la tabla 2. Posteriormente se implementó un algoritmo que automatice la selección y extracción de los metadatos en las estaciones limnimétricas circundantes a través de líneas rectas que miden la proximidad al punto de caudal, en este caso la estación de caudal Ancon-sur. Se presenta el resumen de esta información en la tabla 3.

Tabla 2

Metadatos de la Estaciones pluviométricas seleccionadas.

CODIGO	Nombre	CATEGORIA	Altitud	Latitud	Longitud
27010790	MESETA LA-SENA [27010790]	PM	1700.0	6.03333	-75.667
27015160	CLARA LA [27015160]	AM	2500.0	6.06666667	-75.666
2308500071	RETIRO - AUT [2308500071]	CP	2590.0	6.08211	-75,567
27015260	SALADA LA [27015260]	AM	1923.0	6.04836111	-7.6667

Nota: Tomada de IDEAM (<http://dhime.ideam.gov.co/atencionciudadano/> )

Tabla 3

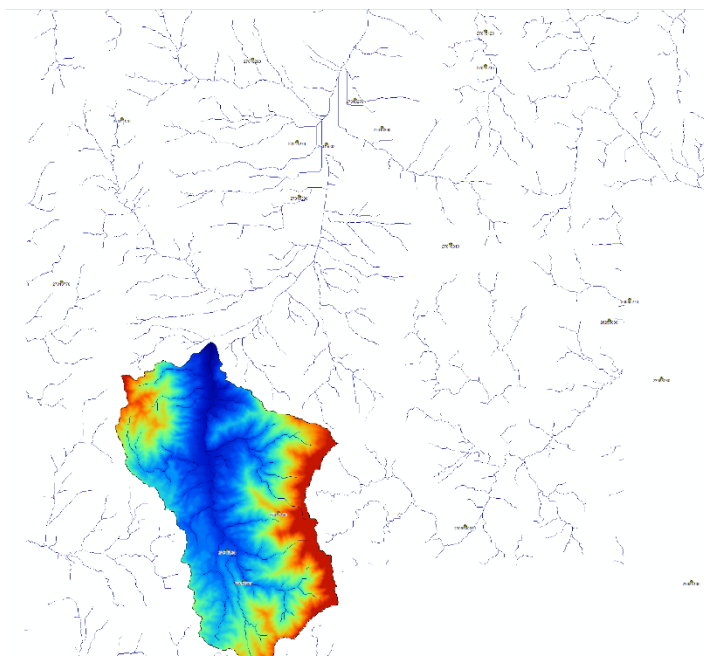
Hidrología a largo paso.

**Tabla 61. Estaciones de caudal a lo largo del río Medellín – Aburrá (Componente Hidrológica e Hidráulica Red Río Fase IV)**

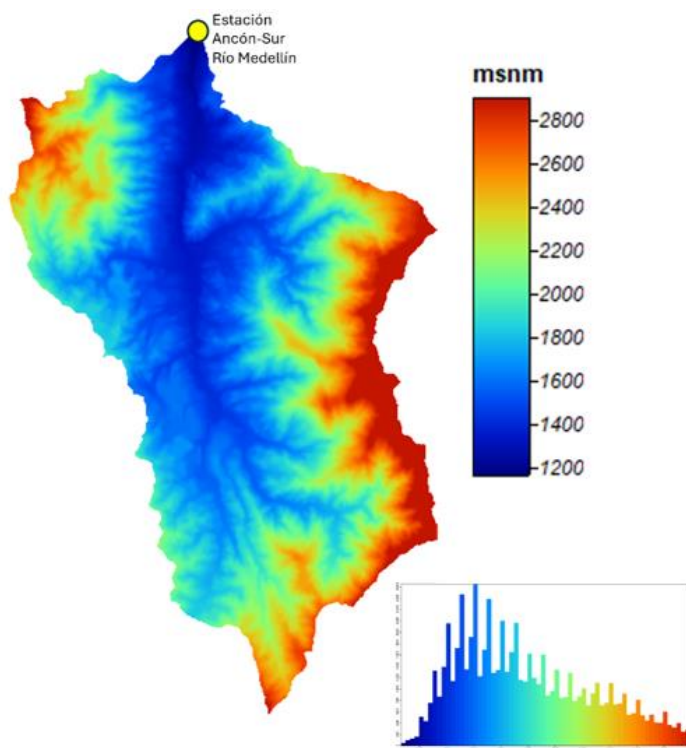
CÓDIGO	NOMBRE	ÁREA DE LA CUENCA (KM <sup>2</sup> )	PRECIPITACIÓN MEDIA (MM/AÑO)	EVAPOTRANSPIRACIÓN POTENCIAL (MM/AÑO)
2701733	SALADA LA RMS 11	42.50	2304.49	1101.07
2701716	CALDAS RM 16	98.59	2295.38	1113.91
2701727	ANCON SUR RMS 17	119.36	2282.98	1117.15
2701734	MACHADO RMS.12	634.41	1944.60	1128.73
2701781	ANCON NORTE RMS 20	731.48	1917.39	1129.02
2701803	GIRARDOTA	775.74	1906.92	1130.80
2701738	HATILLO EL RMS.13	855.70	1900.74	1134.81
2701735	YARUMITO RMS_14	1087.34	1983.45	1136.31

Nota: Balance hídrico, AMVA (2016)

A través de la figura 4 y 5, conocemos los datos geomorfológicos de la cuenca piloto.



*Figura 4 Red de Drenaje (cuenca Río Medellín)*



*Figura 5 DEM (cuenca Río Medellín)*

Se graficaron en la figura 6 las series de precipitación y se observaron las inconsistencias en las series temporales de lluvia correspondiente a cada estación limnimétrica, productos de las mediciones remotas en diversos puntos y reflejadas en la escasez de datos.

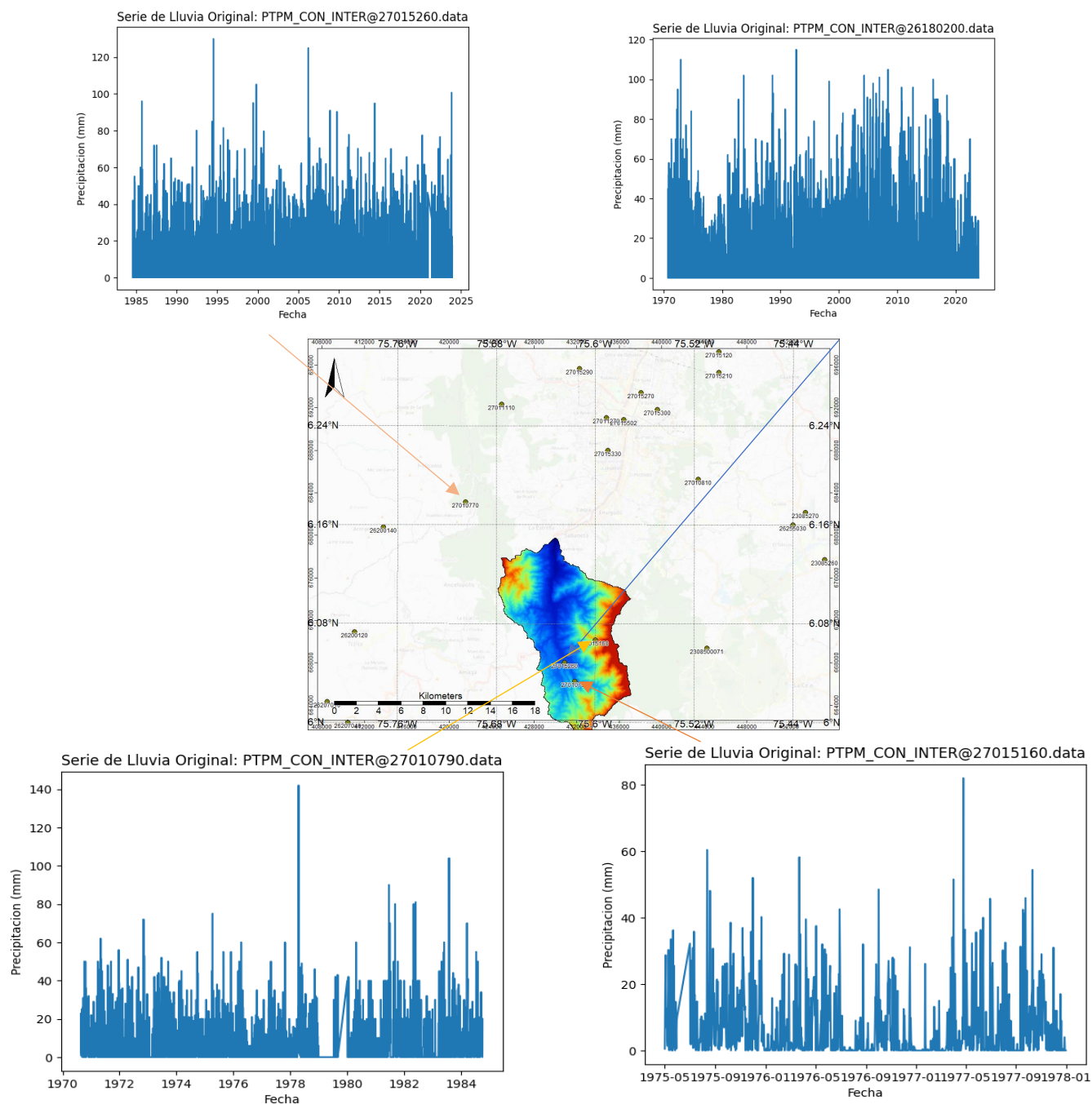


Figura 6 Lluvia diaria en cada estación

En la figura 7 se visualiza la división de las series de lluvia en los sets de entrenamiento, validación y prueba. Asignando el 80, 10 y 10 por ciento respectivamente de los datos.

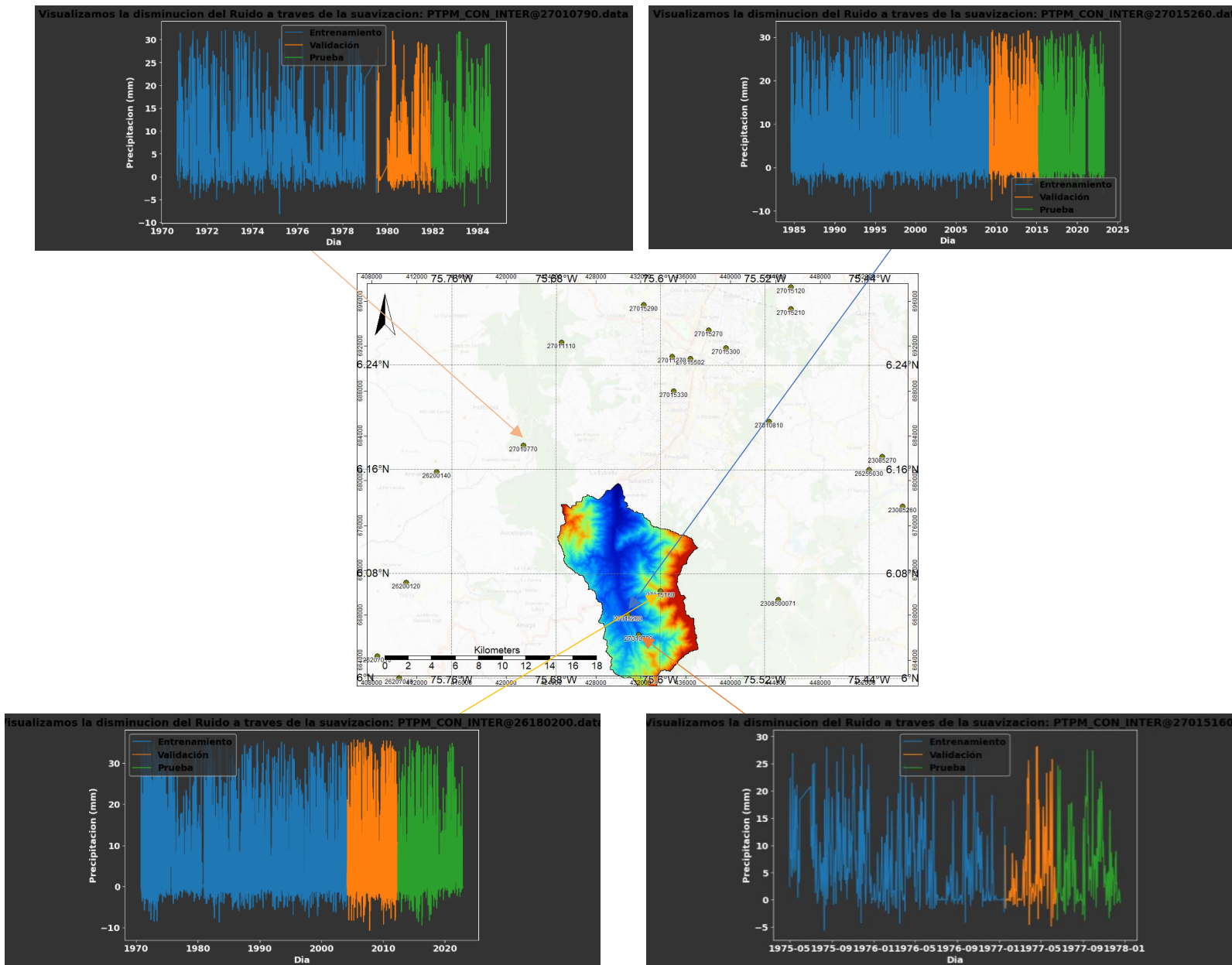
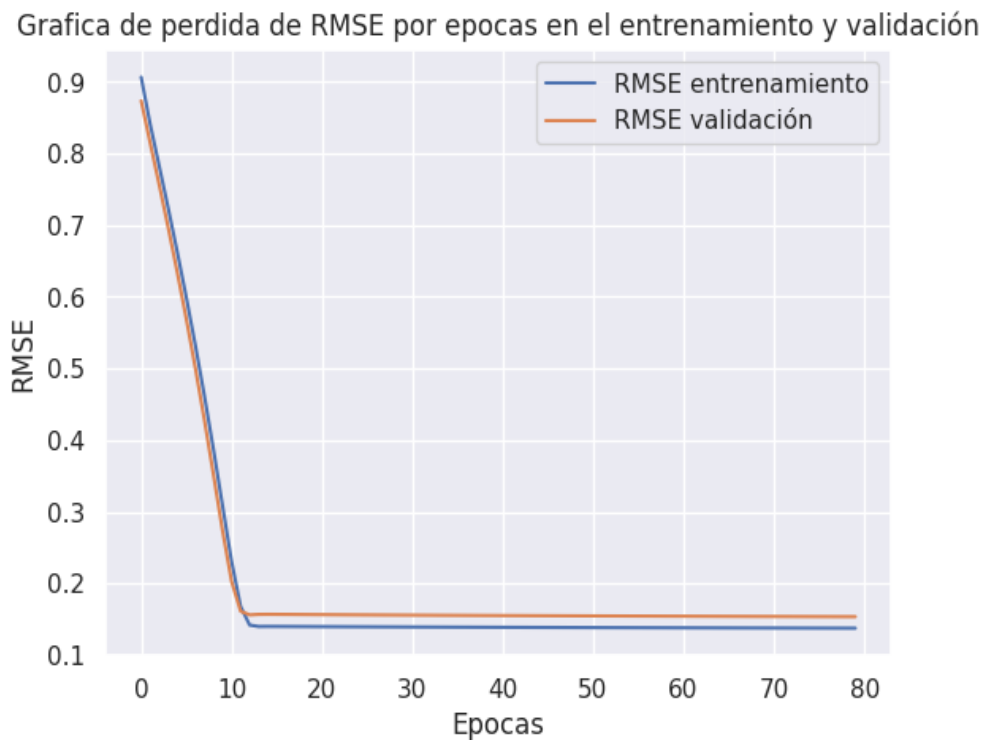


Figura 7 Set de datos

De acuerdo con la figura 8 se verifica que la red aprendió satisfactoriamente, demostrando la evolución de la raíz cuadrada del error cuadrático medio (RMSE) durante el proceso de entrenamiento y validación de la red neuronal LSTM utilizando el conjunto de datos de lluvia diaria.



*Figura 8 Red de Drenaje (Cuenca Río Medellín)*

Entre algunos de los resultados claves como el comportamiento estable y eficiente, con un ajuste adecuado a los datos tanto en la fase de entrenamiento como en la fase de validación, se puede destacar que:

- Convergencia rápida del error: Tanto el RMSE del conjunto de entrenamiento como el de validación disminuyen rápidamente durante las primeras 10 épocas, indicando que la red neuronal ajusta sus parámetros de manera eficiente en las primeras etapas del entrenamiento.

- Estabilidad posterior: Después de la época 10, el RMSE de ambos conjuntos se estabiliza, manteniéndose en un valor cercano a 0.2. Esto sugiere que la red ha alcanzado un buen ajuste y no hay signos evidentes de sobreajuste (overfitting), ya que los errores en ambos conjuntos se comportan de manera similar.
- Buen desempeño del modelo: Los valores de RMSE finales, alrededor de 0.2 para tanto entrenamiento como validación, indican un rendimiento adecuado del modelo al predecir los valores de lluvia diaria.

La figura 9, representa en un gráfico el error calculado como la diferencia entre los valores observados y los valores simulados (predichos) por la red neuronal LSTM en el conjunto de datos de prueba, es decir, aquellos datos que la red no ha visto previamente durante el entrenamiento.

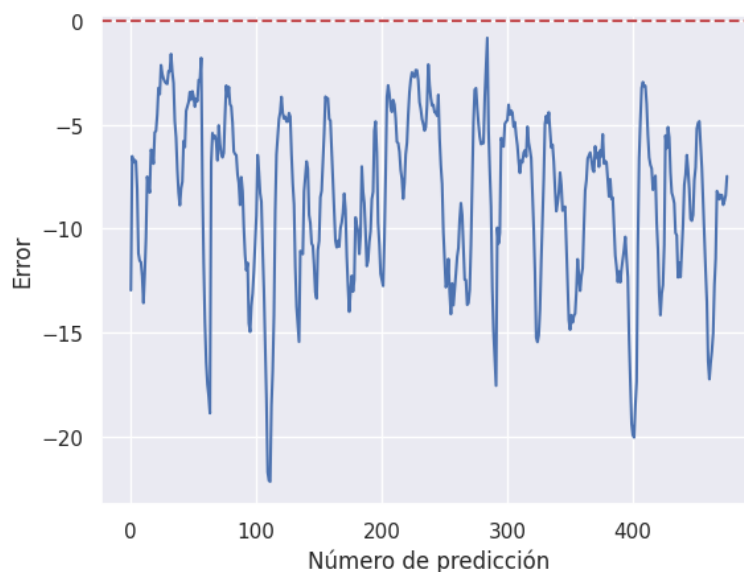


Figura 91 Red de Drenaje (cuenca Río Medellín)

A partir de esta gráfica se pueden obtener los siguientes resultados:

- Oscilación del error: El error varía significativamente a lo largo de las predicciones, con valores que oscilan entre aproximadamente -20 y 5. Esto indica que, aunque la red neuronal ha aprendido a generalizar, no siempre logra predecir con alta precisión para todos los puntos de datos.
- Error negativo predominante: La mayor parte del tiempo, el error se encuentra en el rango negativo, lo que sugiere que, en general, las predicciones del modelo tienden a sobreestimar los valores observados. Es decir, el modelo predice más lluvia de la que se observa realmente.
- Picos de error: Se observan varios picos donde el error se amplifica, llegando a ser mayor que -20 en algunos casos. Esto podría indicar situaciones en las que el modelo tiene dificultad para captar patrones en los datos, posiblemente en eventos de precipitación extremos o atípicos.
- Capacidad de generalización: A pesar de los errores, el modelo parece capturar correctamente la tendencia general de los datos, manteniendo el error dentro de un rango moderado en la mayoría de las predicciones. No obstante, los picos de error sugieren que hay margen para mejorar la capacidad del modelo para generalizar, especialmente en condiciones particulares.

Se realiza la simulación inicial con el modelo de tanques agregados escrito en lenguaje Python como una función que recibe un archivo .xlsx, que contiene los datos sólidos de la precipitación diaria en cada estación, del caudal diario y la evapotranspiración diaria, de acuerdo con la tabla 4, la cual presenta el encabezado del DataFrame generado en el proceso de preprocesamiento.

*Tabla 4*

*Encabezado del DataFrame*

Índices (fechas)	Caudal Observado	Estación 1	Estación 2	Estación 3	Temperatura en °C
------------------	------------------	------------	------------	------------	-------------------

Se inicializo la simulación con los parámetros iniciales propuestos en la tabla 5, para un periodo temporal de 5 años.

*Tabla 5*

*Parámetros ingresado manualmente a la simulación*

Parámetros		Mínimo Sugerido	Máximo Sugerido
Almacenamiento Capilar	500,0	20	600
Conductividad Capa Sup (mm/día)	20,0	1	100
Conductividad Capa Inf (mm/día)	4,0	0,01	10
Perdidas Subterráneas (mm)	0	0	10
Tiempo Medio de Residencia Flujo Superficial (días)	7,0	1	10
Tiempo Medio de Residencia Flujo Subsuperficial (días)	8,0	1	10
Tiempo Medio de Residencia Flujo Base (días)	100,0	50	200

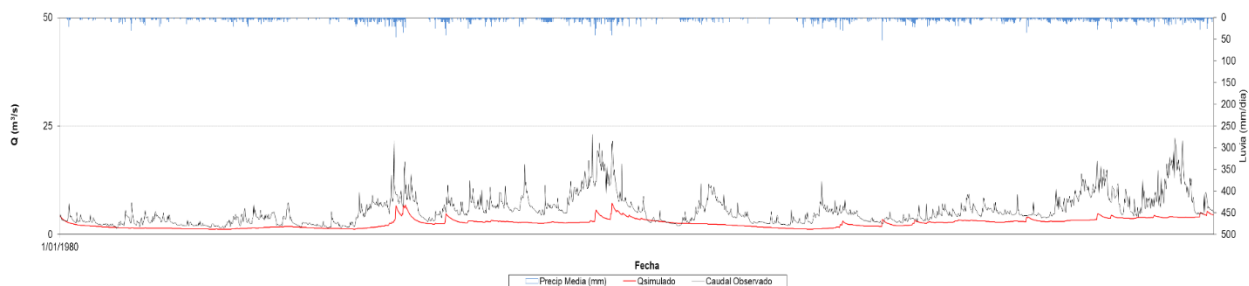
La simulación resulta en valores alejados de la realidad, reflejados en la tabla 6, donde se compara el rendimiento de la simulación a través de la precipitación y caudal.

Tabla 6

Validación de la simulación (observado vs real) en un periodo de simulación.

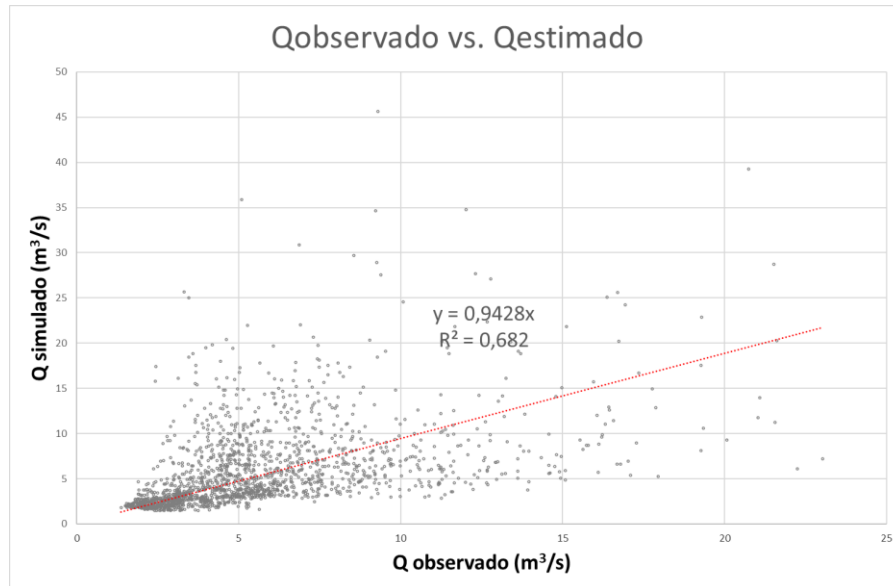
<b>Años de Simulación</b>	5
<b>Prec. Estimada</b>	1398
<b>Prec. Real</b>	2283
<b>ERROR</b>	885
<b>Qreal</b>	5,5
<b>Qsimulado</b>	2,6
<b>Parámetros</b>	<b>Calibración</b>
<b>Años</b>	<b>1980-1984</b>
<b>RMSE</b>	4,1
<b>NASH</b>	-0,472337946

La figura 10 representa gráficamente el comportamiento de la simulación en el punto de medición Ancón-sur en el río Medellín.



**Figura 10** Gráfico de caudal observado vs simulado en la estación piloto en el periodo 1980-1984

La figura 11 representa gráficamente qué tan bien el modelo matemático utilizado reproduce la realidad, con los parámetros no optimizados.



**Figura 11** Gráfico de dispersión Caudal Observado vs Estimado

Se realiza la Optimización aplicando la técnica PSO sobre los parámetros haciendo uso de la función objetivo, resumida en la ecuación 1 que disminuye el error KGE, obteniendo los siguientes resultados de la optimización.

$$\text{Ecuación (1)} \quad \mathbf{J}(\mathbf{p}) = -\frac{1}{n} \sum_{i=1}^n \mathbf{KGE}(\mathbf{Q}_{obs}^i, \mathbf{Q}_{sim}^i)$$

Donde:

- $\mathbf{J}(\mathbf{p})$  es la función objetivo a minimizar, basada en la eficiencia del modelo.
- $\mathbf{p}$  representa el vector de parámetros optimizados.
- $n$  es el número de partículas (modelos simulados con diferentes parámetros).
- $\mathbf{Q}_{obs}(i)$  es el caudal observado en el instante  $i$
- $\mathbf{Q}_{sim}(i)$  es el caudal simulado en el instante  $i$
- $\mathbf{KGE}(\mathbf{Q}_{obs}, \mathbf{Q}_{sim})$  el coeficiente de eficiencia Kling-Gupta.

La meta de la optimización es minimizar el valor negativo de KGE para encontrar los mejores parámetros del modelo. Posteriormente se inicializo la simulación con los parámetros encontrados en el proceso de optimización, propuestos en la tabla 7, para un periodo temporal de 5 años.

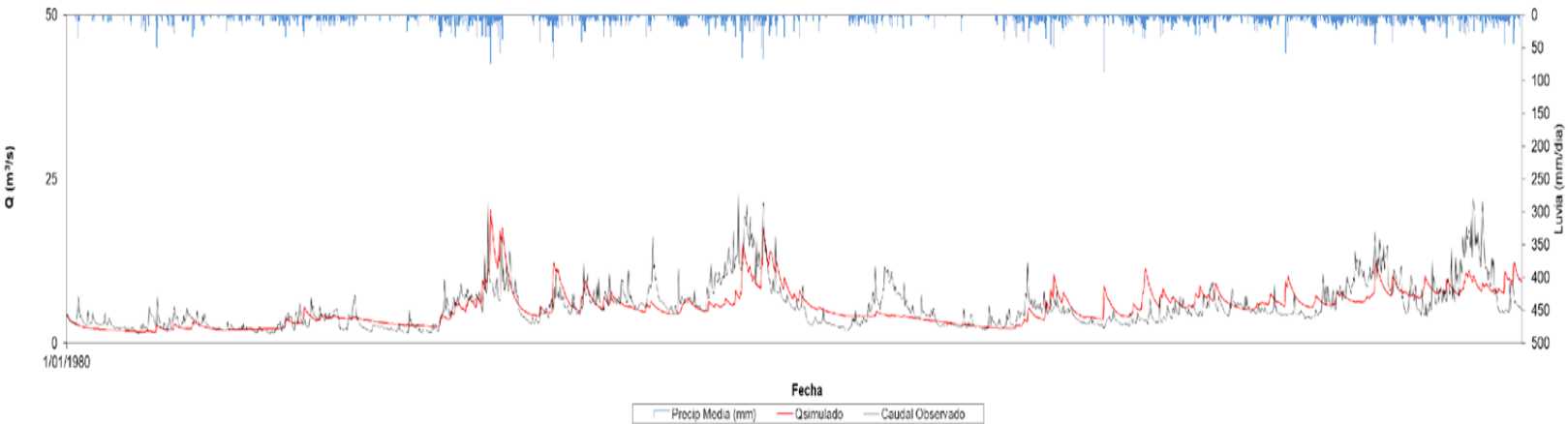
*Tabla 7*  
*Parámetros optimizados ingresados manualmente a la simulación*

Parámetros		Mínimo Sugerido	Máximo Sugerido
Almacenamiento Capilar	324,0	20	600
Conductividad Capa Sup (mm/día)	57,0	1	100
Conductividad Capa Inf (mm/día)	10,0	0,01	10
Perdidas Subterráneas (mm)	0	0	10
Tiempo Medio de Residencia Flujo Superficial (días)	5,0	1	10
Tiempo Medio de Residencia Flujo Subsuperficial (días)	10,0	1	10
Tiempo Medio de Residencia Flujo Base (días)	185,4	50	200

Se valida la mejoría en la simulación de acuerdo con la figura 12 y a la tabla 8, donde los valores estimados son muy aproximados a los observados.

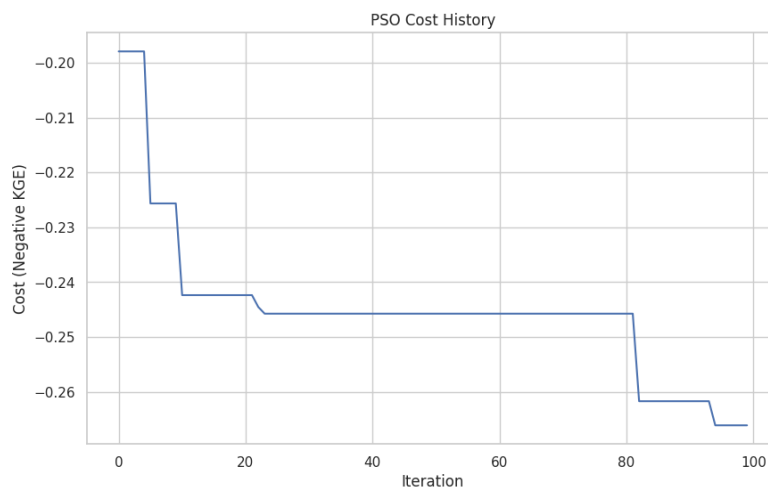
*Tabla 8*  
*Validación de la simulación (observado vs real) en un periodo de simulación.*

<b>Años de Simulación</b>	5
<b>Prec. Estimada</b>	2292
<b>Prec. Real</b>	2283
<b>ERROR</b>	9
<b>Qreal</b>	5,5
<b>Qsimulado</b>	5,4
<b>Parámetros</b>	<b>Calibración</b>
<b>Años</b>	<b>1980-1984</b>
<b>RMSE</b>	2,5
<b>NASH</b>	0,432972693



**Figura 12** Gráfico de caudal observado vs simulado en la estación piloto en el periodo 1980-1984

Para conocer los detalles sobre el proceso de optimización de parámetros, utilizando la técnica de Optimización por Enjambre de Partículas (PSO), a través de la figura 13, visualizamos el historial de costo, en este caso, está representado por el Kling-Gupta Efficiency (KGE) negativo, lo que indica que el objetivo de la optimización es minimizar este valor para mejorar el desempeño del modelo hidrológico.

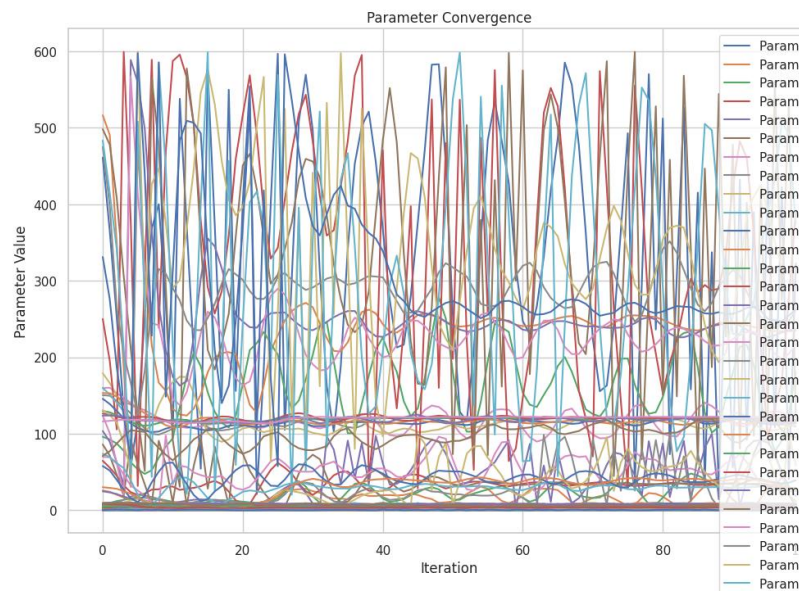


**Figura 13** Costo de la optimización de PSO por medio del indicador KGE durante el rendimiento

Los resultados clave que se pueden destacar son:

- Mejora inicial rápida: En las primeras iteraciones, el costo disminuye de forma significativa, pasando de un valor cercano a -0.20 a aproximadamente -0.23 en las primeras 20 iteraciones. Esto indica que el PSO ajusta rápidamente los parámetros del modelo, encontrando configuraciones que mejoran el desempeño inicial.
- Estancamiento en el medio: Entre las iteraciones 20 y 80, el costo se mantiene relativamente constante, alrededor de -0.24. Esto sugiere que, en esta fase, el algoritmo no encontró mejoras significativas en el espacio de búsqueda de los parámetros, alcanzando una especie de "meseta".
- Mejora tardía: Después de la iteración 80, el costo vuelve a disminuir, alcanzando un valor de alrededor de -0.26 en la iteración 100. Esto indica que el algoritmo logró escapar de la meseta y encontró configuraciones de parámetros que mejoraron aún más el rendimiento del modelo.
- Convergencia: Hacia el final del proceso de optimización (iteraciones 90-100), el costo parece estabilizarse, lo que sugiere que el PSO ha convergido a una solución óptima o cercana al óptimo para los parámetros del modelo.

Por último, visualizamos en la figura 14 el rango de valores que toma el enjambre en cada época de acuerdo con el parámetro.



**Figura 14** Costo de la optimización de PSO por medio del indicador KGE durante el rendimiento

## Conclusiones

Se desarrolló un algoritmo que permite extraer metadatos de precipitación de las estaciones más cercanas al caudalímetro seleccionado, generando un conjunto de datos en formato .xlsx para su uso en el modelo TETIS. Este avance facilita la automatización en la recopilación de datos hidrológicos, optimizando el proceso de modelación y asegurando la disponibilidad de información precisa y actualizada.

La implementación del modelo de optimización por PSO logró reducir significativamente el costo de simulación, mejorando la eficiencia del modelo hidrológico. La disminución del KGE negativo indica un mejor ajuste entre los datos simulados y observados, alcanzando una mayor precisión en la simulación del flujo, lo que se traduce en un RMSE estabilizado en aproximadamente 0.2.

El enfoque sinérgico que combina redes neuronales LSTM con el método de optimización por PSO ha demostrado ser efectivo para calibrar el modelo, reduciendo la incertidumbre en la variabilidad de la lluvia. Este método ha permitido calibrar el modelo de manera aceptable, lo que resalta el potencial de las técnicas de aprendizaje automático en la mejora de la precisión de las predicciones hidrológicas.

En conclusión, este estudio establece un marco sólido para futuras investigaciones en hidrología computacional, evidenciando que la integración de técnicas avanzadas de aprendizaje automático y optimización puede mejorar significativamente la precisión de las predicciones y la gestión de recursos hídricos en cuencas complejas como las andinas. La automatización y la optimización no solo aumentan la eficiencia del análisis, sino que

también permiten un manejo más eficaz de grandes volúmenes de datos, lo que es crucial en el contexto actual de cambio climático y presión sobre los recursos hídricos.

## Referencias

- Abudu, S., Cui, C.-l., King, J. P., & Abudukadeer, K. (2010). Comparison of performance of statistical models in forecasting monthly streamflow of Kizil River, China. *Water Science and Engineering*, 269--281.
- Adamowski, J., Chan, H. F., Prasher, S. O., & Sharda, V. N. (2012). Comparison of multivariate adaptive regression splines with coupled wavelet transform artificial neural networks for runoff forecasting in Himalayan micro-watersheds limited data. *Journal of hydroinformatics*, 731--744.
- Addor, N., & Melsen, L. (2019). Legacy, rather than adequacy, drives the selection of hydrological models. *Water resources research*, 55(1), 378--390.
- Babak , M., Nguyen Thi Thuy , L., Quoc Bao , P., Ali Najah , A., Jana , V., S.I. , A., . . . Ahmed , E.-S. (2020). Adaptive neuro-fuzzy inference system coupled with shuffled frog leaping algorithm for predicting river streamflow time series. *Hydrological Sciences Journal*, 1738--1751.
- Beck, H. E., Pan, M., Lin, P., Seibert, J., van Dijk, A. I., & Wood, E. F. (2020). Global fully distributed parameter regionalization based on observed streamflow from 4,229 headwater catchments. *Journal of Geophysical Research: Atmospheres*, 125(17), e2019JD031485.
- Beven, K. (2019). How to make advances in hydrological modelling. *Hydrology Research*, 1481--1494.

- Beven, K. J. (2020). A history of the concept of time of concentration. *Hydrology and Earth System Sciences*, 24(5), 2655--2670.
- Bouaziz, L. J., Thirel, G., de Boer-Euser, T., Melsen, L. A., Buitink, J., Brauer, C. C., . . . others. (2020). Behind the scenes of streamflow model performance. *Hydrology and Earth System Sciences Discussions*, 2020, 1--38.
- Cabrera, J. (2012). Calibración de modelos hidrológicos. *Instituto para la Mitigación de los Efectos del Fenómeno El Niño, Universidad Nacional de Ingeniería, Facultad de Ingeniería Civil, Perú*, 1(1).
- Chadalawada, J., Herath, H., & Babovic, V. (2020). Hydrologically informed machine learning for rainfall-runoff modeling: A genetic programming-based toolkit for automatic model induction. *Water Resources Research*, e2019WR026933.
- Demirezen, K. İ., & Kazezyılmaz-Alhan, C. M. (2022). Evaluation of the hydrological performance of infiltration trench with rainfall-watershed-infiltration trench experimental setup. *Journal of Hydrologic Engineering*, 27(3), 04021050.
- Fatichi, S., Vivoni, E. R., Ogden, F. L., Ivanov, V. Y., Mirus, B., Gochis, D., . . . others. (2016). An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. *Journal of Hydrology*, 537, 45--60.
- Fenicia, F., Kavetski, D., Savenije, H. H., & Pfister, L. (2016). From spatially variable streamflow to distributed hydrological models: Analysis of key modeling decisions. *Water Resources Research*, 954--989.

- Flores Quispe, E., Delgado, C., & Condori, E. (2019). PREDICCIÓN DE CAUDALES DIARIOS DE LA ESTACIÓN HIDROMÉTRICA TUMILACA APLICANDO MODELOS ESTOCÁSTICOS AUTOREGRESIVOS. *Ciencia & Desarrollo*, 68-73. <https://doi.org/10.33326/26176033.2015.19.489>
- Fowler, K., Peel, M., Western, A., & Zhang, L. (2018). Improved rainfall-runoff calibration for drying climate: Choice of objective function. *Water Resources Research*, 54(5), 3392--3408.
- Frame, J. M., Kratzert, F., Klotz, D., Gauch, M., Shalev, G., Gilon, O., . . . Nearing, G. S. (2022). Deep learning rainfall--runoff predictions of extreme events. *Hydrology and Earth System Sciences*, 26(13), 3377--3392.
- Gad, A. G. (2022). Particle swarm optimization algorithm and its applications: a systematic review. *Archives of computational methods in engineering*, 29(5), 2531--2561.
- Gao, H., Birkel, C., Hrachowitz, M., Tetzlaff, D., Soulsby, C., & Savenije, H. H. (2019). A simple topography-driven and calibration-free runoff generation module. *Hydrology and Earth System Sciences*, 23(2), 787--809.
- Gauch, M., Kratzert, F., Klotz, D., Nearing, G., Lin, J., & Hochreiter, S. (2021). Rainfall--runoff prediction at multiple timescales with a single Long Short-Term Memory network. *Hydrology and Earth System Sciences*, 25(4), 2045--2062.

- Gutierrez, J. C., Caballero, C. B., Vasconcellos, S. M., Vanelli, F. M., & Bravo, J. M. (2022). Multi-objective calibration of Tank model using multiple genetic algorithms and stopping criteria. *RBRH*, e31.
- Hadi, S. J., & Tombul, M. (2018). Monthly streamflow forecasting using continuous wavelet and multi-gene genetic programming combination. *Journal of Hydrology*, 674--687.
- Herath, H. M., Chadalawada, J., & Babovic, V. (2021). Hydrologically informed machine learning for rainfall--runoff modelling: towards distributed modelling. *Hydrology and Earth System Sciences*, 4373--4401.
- Hsu, P.-Y., & Yeh, Y.-L. (2015). Study on Flood Para-Tank Model Parameters with Particle Swarm Optimization. *J. Inf. Hiding Multim. Signal Process.*, 6(5), 911--923.
- Humphrey, G. B., Gibbs, M. S., Dandy, G. C., & Maier, H. R. (2016). A hybrid approach to monthly streamflow forecasting: integrating hydrological model outputs into a Bayesian artificial neural network. *Journal of Hydrology*, 623--640.
- Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall--runoff modelling using long short-term memory (LSTM) networks. *Hydrology and Earth System Sciences*, 22(11), 6005--6022.
- Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., & Nearing, G. S. (2019). Toward improved predictions in ungauged basins: Exploiting the power of machine learning. *Water Resources Research*, 55(12), 11344--11354.

- Kratzert, F., Klotz, D., Hochreiter, S., & Nearing, G. S. (2021). A note on leveraging synergy in multiple meteorological data sets with deep learning for rainfall--runoff modeling. *Hydrology and Earth System Sciences*, 25(5), 2685--2703.
- Mohammadi, B. (2021). A review on the applications of machine learning for runoff modeling. *Sustainable Water Resources Management*, 98.
- Mohammadi, B. A. (2021). A review on the applications of machine learning for runoff modeling. *Sustainable Water Resources Management*, 98.
- Mohammadi, B., & Mehdizadeh, S. (2020). Modeling daily reference evapotranspiration via a novel approach based on support vector regression coupled with whale optimization algorithm. *Agricultural Water Management*, 237, 106145.
- Mohammadi, B., & Mehdizadeh, S. (2020). Modeling daily reference evapotranspiration via a novel approach based on support vector regression coupled with whale optimization algorithm. *Agricultural Water Management*, 237, 106145.
- Mohammadi, B., Guan, Y., Aghelpour, P., Emamgholizadeh, S., Pillco Zolá, R., & Zhang, D. (2020). Simulation of Titicaca lake water level fluctuations using hybrid machine learning technique integrated with grey wolf optimizer algorithm. *Water*, 3015.
- Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., . . . Gupta, H. V. (2021). What role does hydrological science play in the age of machine learning? *Water Resources Research*, 57(3), e2020WR028091.

- Pascual Aguilar, J. A., & Díaz Martín, M. (2016). Guía práctica sobre la modelización hidrológica y el modelo. *IMDEA AGUA, Centro para el Conocimiento del Paisaje*.
- Rafael-Miñope, W. F., Vilcherres-Lizárraga, P. V., Muñoz-Pérez, S. P., Tuesta-Montez, V. A., & Mejía-Cabrera, H. I. (2022). Modelamiento de procesos hidrológicos aplicando técnicas de inteligencia artificial: una revisión sistemática de la literatura. *Iteckne*, 46--60.
- Rakovec, O., Kumar, R., Attinger, S., & Samaniego, L. (2016). Improving the realism of hydrologic model functioning through multivariate parameter estimation. *Water Resources Research*, 52(10), 7779--7792.
- Santos, C. A., Freire, P., K. M. M., M. S., & Soares Júnior, A. (2011). Application of a particle swarm optimization to the tank model. *IAHS Publ*, 114--120.
- Sarmiento, I. (28 de Abril de 2020). *Movilidadprofe*. Recuperado el Octubre de 2020, de [https://www.youtube.com/watch?v=OQ7CKqNkQ0U&ab\\_channel=Movilidadprofe](https://www.youtube.com/watch?v=OQ7CKqNkQ0U&ab_channel=Movilidadprofe)
- Shin, M.-J., Guillaume, J. H., Croke, B. F., & Jakeman, A. J. (2015). A review of foundational methods for checking the structural identifiability of models: Results for rainfall-runoff. *Journal of Hydrology*, 520, 1--16.
- Tikhamarine, Y., Souag-Gamane, D., Ahmed, A. N., Kisi, O., & El-Shafie, A. (2020). Improving artificial intelligence models accuracy for monthly streamflow

- forecasting using grey Wolf optimization (GWO) algorithm. *Journal of Hydrology*, 582, 124435.
- Tikhamarine, Y., Souag-Gamane, D., Ahmed, A. N., Sammen, S. S., Kisi, O., Huang, Y. F., & El-Shafie, A. (2020). Rainfall-runoff modelling using improved machine learning methods: Harris hawks optimizer vs. particle swarm optimization. *Journal of Hydrology*, 589, 125133.
- Tikhamarine, Y., Souag-Gamane, D., Ahmed, A. N., Sammen, S. S., Kisi, O., Huang, Y. F., & El-Shafie, A. (2020). Rainfall-runoff modelling using improved machine learning methods: Harris hawks optimizer vs. particle swarm optimization. *Journal of Hydrology*, 589, 125133.
- Tongal, H., & Booij, M. J. (2018). Simulation and forecasting of streamflows using machine learning models coupled with base flow separation. *Journal of hydrology*, 266--282.
- Tyralis, H., Papacharalampous, G., & Langousis, A. (2021). Super ensemble learning for daily streamflow forecasting: Large-scale demonstration and comparison with multiple machine learning algorithms. *Neural Computing and Applications*, 33(8), 3053--3068.
- Valipour, M. a. (2012). Optimize of all effective infiltration parameters in furrow irrigation using visual basic and genetic algorithm programming. *Aust. J. Basic Appl. Sci*, 132--137.

- Valipour, M., & Montazar, A. (2012). Optimize of all effective infiltration parameters in furrow irrigation using visual basic and genetic algorithm programming. *Aust. J. Basic Appl. Sci*, 6(6), 132-137.
- Wang, J., Shi, P., Jiang, P., Hu, J., Qu, S., Chen, X., . . . Xiao, Z. (2017). Application of BP neural network algorithm in traditional hydrological model for flood forecasting. *Water*, 9(1), 48.
- Wang, R., Yuan, X., Tian, F., Liu, M., Wang, X., Li, X., & Wu, M. (2024). Flood Classification and Improved Loss Function by Combining Deep Learning Models to Improve Water Level Prediction in a Small Mountain Watershe. *Available at SSRN 4779010*.
- Zhang, L., Zhao, Y., Ma, Q., Wang, P., Ge, Y., & Yu, W. (2021). A parallel computing-based and spatially stepwise strategy for constraining a semi-distributed hydrological model with streamflow observations and satellite-based evapotranspiration. *Journal of Hydrology*, 599, 126359.
- Zhu, S., Luo, X., Yuan, X., & Xu, Z. (2020). An improved long short-term memory network for streamflow forecasting in the upper Yangtze River. *Stochastic Environmental Research and Risk Assessment*, 34.  
<https://doi.org/10.1007/s00477-020-01766-4>