



Predicción del precio del agua mediante diferencias finitas y machine learning

Water price prediction using finite differences and machine learning

Christopher Nolan Magaña Barocio¹  , Gerardo Tinoco Guerrero² , José Alberto Guzmán Torres² ,
Francisco Javier Domínguez Mota^{1,2} 

¹ Universidad Michoacana de San Nicolás de Hidalgo, Posgrado en Ciencias en Ingeniería Física.

² Universidad Michoacana de San Nicolás de Hidalgo, Facultad de Ingeniería Civil.

Resumen

El presente trabajo analiza una metodología híbrida para la predicción del precio del agua en México, que integra una simulación de un sistema físico mediante el análisis de datos. Los objetivos principales fueron desarrollar un modelo físico basado en la ecuación de difusión para simular el flujo en una red hídrica en condiciones óptimas, cuyas soluciones se obtienen mediante el método de Diferencias Finitas. Se generó un conjunto de datos sintético a partir de simulaciones que representan las condiciones habituales del fenómeno, el cual se utilizó para entrenar un modelo de Machine Learning para predecir el precio del agua. Se utilizó un modelo de regresión lineal que alcanzó un coeficiente de determinación de 0.9984 y un error cuadrático medio de 0.01002. Los resultados confirman que la metodología propuesta es una herramienta de pronóstico con un excelente grado de certidumbre y precisión para estimar el precio del agua, dadas las condiciones particulares de la red hídrica, ya que el modelo se valida en condiciones ideales y con datos generados de manera sintética.

Palabras clave: precio del agua, diferencias finitas, Machine Learning, modelo de difusión, regresión lineal.

Abstract

This work analyzes a hybrid methodology for water price prediction that integrates physical system simulation with data analysis. The main objective was to develop a physical model based on the diffusion equation to simulate flow in a water network under optimal conditions, which was solved using the Finite Difference method. A synthetic dataset was generated from simulations and used to train a Machine Learning model to predict water prices. A linear regression model was used, which achieved a coefficient of determination (R^2) of 0.9984 and a root mean squared error (RMSE) of 0.01002. The results confirm that the proposed methodology is a reliable and precise forecasting tool for estimating water prices based on the physical conditions of the water network.

Keywords: water price, finite differences, Machine Learning, diffusion model, linear regression.

Historial

Manuscrito recibido: 31 de octubre de 2025

Manuscrito aceptado: 19 de abril de 2026

Manuscrito publicado: agosto de 2026

✉ Autor para correspondencia

Christopher Nolan Magaña Barocio

1339846k@umich.mx

Como citar

Magaña Barocio CN, Tinoco Guerrero G, Guzmán Torres JA, Domínguez Mota FJ (2026). Predicción del precio del agua mediante diferencias finitas y machine learning. *Ciencia Nicolaita* 97:188-194. doi:[10.35830/cn.vi97.952](https://doi.org/10.35830/cn.vi97.952)

Introducción

El costo del agua potable en México no es un valor estandarizado; el panorama se caracteriza por una notable fragmentación y una alta variabilidad de precios entre los municipios y los estados, y carece de una metodología estandarizada para establecer un costo fijo. Los precios dependen fuertemente de factores locales de índole geográfica, política y económica, lo que introduce una considerable incertidumbre en su planeación, gestión y consumo (IMCO, 2023).

Para crear herramientas que simplifiquen esta complejidad, la gestión efectiva de los recursos hídricos requiere herramientas predictivas modernas capaces de integrar el

conocimiento de los procesos físicos con la flexibilidad de los algoritmos de Machine Learning (Kraft *et al.*, 2022). Estos modelos son eficientes y precisos, pero suelen ser computacionalmente costosos y alcanzar una alta precisión depende de una calibración exhaustiva que requiere datos de alta calidad, lo cual muy a menudo es escaso (Olivier *et al.*, 2025). Para abordar la problemática de la

escasez de datos de campo, el modelo físico actúa como generador de conocimiento que encapsula la física del sistema y permite a un modelo de Machine Learning aprender sus relaciones para realizar predicciones casi instantáneas (Dai *et al.*, 2024).

En este contexto, el presente trabajo tiene como objetivo desarrollar un modelo e implementar una metodología híbrida para predecir el precio del agua en México. La estrategia propuesta combina la simulación numérica del flujo hídrico, basada en la ecuación de difusión y resuelta mediante Diferencias Finitas, con un modelo de regresión lineal simple. El propósito de esto es demostrar que este enfoque puede generar un modelo predictivo fiable y preciso, capaz de predecir el precio del agua dadas las condiciones de una red hídrica bajo condiciones ideales, estableciendo las bases para una herramienta de apoyo en la toma de decisiones estratégicas en la gestión de los recursos hídricos.

Materiales y métodos

El método desarrollado integra un modelo físico para la simulación del flujo del agua, la generación de un conjunto de datos sintético y el entrenamiento de un modelo predictivo de Machine Learning. El comportamiento del nivel del agua en la red se modela mediante la ecuación de difusión unidimensional, que describe cómo una cantidad, en este caso el nivel del agua, se distribuye con el tiempo en un medio. La Ecuación (1) indica dicho modelo:

$$\frac{\partial u}{\partial t} = D \frac{\partial^2 u}{\partial x^2}, \quad (1)$$

donde:

- u : Representa el nivel del agua en un punto de la tubería.
- t : Es el tiempo.
- x : Es la posición a lo largo de la tubería.
- D : Es el coeficiente de difusividad, que indica la rapidez con la que se propaga el agua.

La Ecuación (1) representa una simplificación del flujo en la tubería, donde la variable simboliza el nivel de carga hidráulica en el punto al tiempo. Para este trabajo, se asume que el flujo es unidimensional y que la sección transversal de la tubería es constante; esto permite que el nivel del agua actúe como un análogo del volumen y de la disponibilidad del recurso en ese segmento de la red. Rea-

lizando la simplificación a un modelo de difusión como el mostrado en la Ecuación (1), no captura la totalidad de las dinámicas de un fluido, pero es efectivo para modelar la propagación y el volumen a lo largo del sistema bajo condiciones de flujo lento, sirviendo como base para la generación de datos sintéticos al momento de realizar distintas simulaciones.

Este modelo, aunque es una simplificación, emplea un esquema de Diferencias Finitas que discretiza el problema en el espacio y en el tiempo. Este esquema consiste en discretizar el dominio continuo en una malla de puntos o nodos y aproximar las derivadas parciales mediante Diferencias Finitas en dichos nodos. Esto transforma la ecuación diferencial parcial en un sistema de ecuaciones algebraicas que puede resolverse iterativamente (Burden y Faires, 2011).

Para la Ecuación (1), se utiliza un esquema explícito de Diferencias Finitas adelantadas en el tiempo y centradas en el espacio (FTCS, por sus siglas en inglés). La discretización de las derivadas es la siguiente:

$$\frac{\partial u}{\partial t} \approx \frac{u_i^{k+1} - u_i^k}{\Delta t}, \quad (2)$$

$$\frac{\partial^2 u}{\partial x^2} \approx \frac{u_{i+1}^k - 2u_i^k + u_{i-1}^k}{\Delta x^2}, \quad (3)$$

donde:

$u_i^k = u(x_i, t_k)$ es el valor del nivel del agua en el nodo espacial i y en el paso de tiempo k , mientras que Δx y Δt son los tamaños de paso espacial y temporal respectivamente (LeVeque, 2007). Sustituyendo las Ecuaciones (3) y (4) en la Ecuación (1) y despejando para el nivel del agua en el siguiente paso del tiempo u_i^{k+1} se obtiene la fórmula de iteración explícita:

$$u_i^{k+1} \approx u_i^k + \alpha (u_{i+1}^k - 2u_i^k + u_{i-1}^k), \quad (4)$$

donde:

$$\alpha = \frac{D \Delta t}{\Delta x^2} \text{ es el número de Courant.}$$

Este esquema permite calcular el estado del sistema en el tiempo t_{k+1} basándose únicamente en el estado conocido en el tiempo t_k . Para que la simulación sea numéricamente estable, el esquema debe satisfacer la condición de estabilidad de Courant-Friedrichs-Lewy (CFL). Para la Ecuación (1), la condición está dada por:

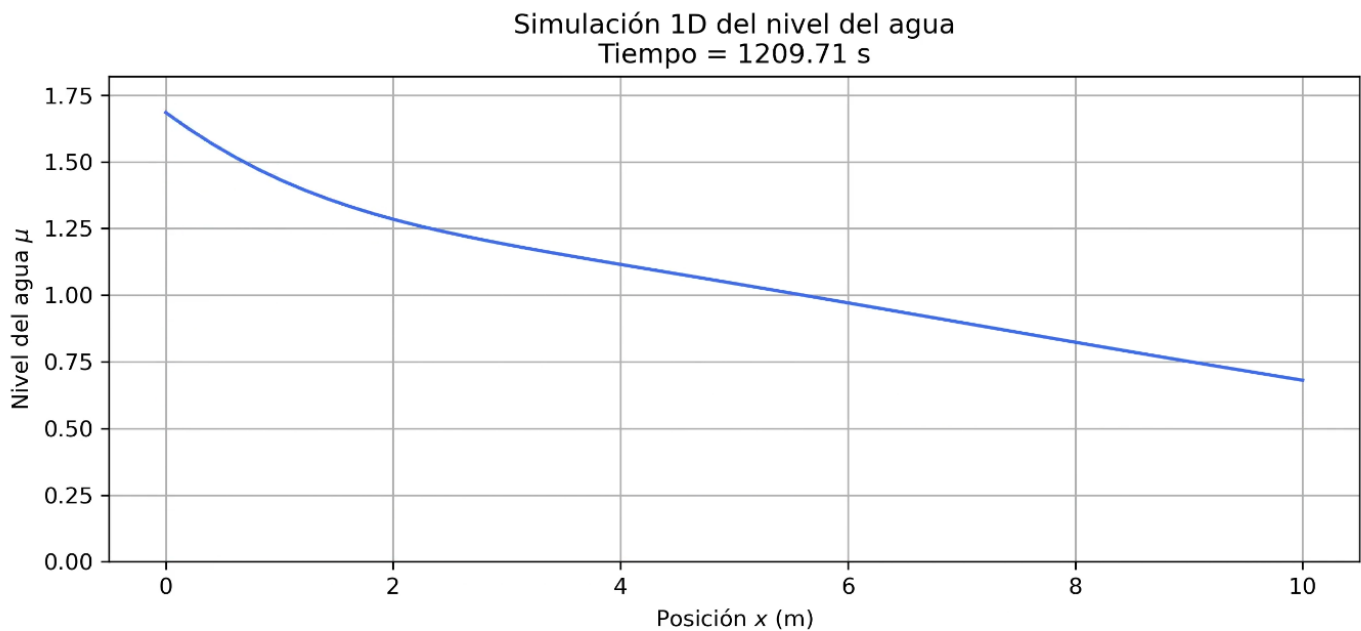


Figura 1. Simulación numérica del esquema de Diferencias Finitas para predecir el precio del agua.

$$\alpha = \frac{D\Delta t}{\Delta x^2} \leq \frac{1}{2}. \quad (5)$$

El no cumplimiento de esta condición conduce a que los errores de redondeo se amplifican exponencialmente en cada paso del tiempo, generando oscilaciones que destruyen la solución numérica (Burden y Faires, 2011).

Este esquema de Diferencias Finitas se completa con la definición de condiciones de frontera dinámicas de tipo Neumann, que especifican los flujos en los extremos de la red. Flujo de entrada q_0 se modela como una fuente oscilatoria simulando ciclos de suministro variables; solo se consideran flujos positivos:

$$q_0(t) = \max(A \sin(\omega t), 0). \quad (6)$$

Flujo de salida q_L , se define como una función de la demanda, proporcional al nivel de agua en el penúltimo nodo de la tubería, simulando un consumo que depende de la disponibilidad:

$$q_L(t) = -k_1 u_{i-1}. \quad (7)$$

En la **Figura 1** se muestra la simulación del esquema de Diferencias Finitas utilizado, en el que se observan las condiciones de frontera, el nivel del agua y las fluctuaciones. En este caso se consideró una tubería de 10 metros sin fricción y sin condiciones físicas externas a la tubería, como la temperatura o la presión.

La simulación del esquema tiene como propósito explorar un amplio rango de parámetros para generar un

conjunto de datos sintéticos en el que varían los parámetros físicos que definen el comportamiento del sistema. La combinación de todos los valores resultó en un conjunto de datos con 648,000 registros, un volumen adecuado para entrenar un modelo de regresión lineal múltiple (Goodfellow *et al.*, 2016). Los parámetros y sus rangos se detallan en la **Tabla 1**.

Los parámetros y sus rangos de variación, detallados en la **Tabla 1**, fueron seleccionados para representar un amplio espectro de escenarios operativos plausibles en una red hídrica. Por ejemplo, los valores de la amplitud de flujo de entrada (A) se extienden desde condiciones de bajo suministro (0.5) hasta un flujo intenso (1.5). Dichos parámetros fueron elegidos arbitrariamente, considerando la estabilidad del modelo, ya que cada valor altera la estabilidad del esquema de Diferencias Finitas.

Para cada simulación, se calculó el nivel promedio de agua en la tubería en cada instante del tiempo. A partir de este valor, se generó la variable objetivo del estudio: un precio sintético del agua $P(t)$. Este precio se modeló mediante la Ecuación (8).

Tabla 1. Parámetros de simulación para generar el conjunto de datos sintético.

Parámetro	Símbolo	Valores utilizados
Amplitud del flujo de entrada	A	0.5, 1.0, 1.5
Frecuencia del flujo de entrada	ω	5, 10, 20
Coefficiente de Difusividad	D	0.1, 0.2, 0.5
Constante de demanda de salida de la tubería	k_1	0.05, 0.1, 0.15

$$P(t) = P_{base} + \beta \left[1 - \bar{u}(t) \right] - \epsilon(t), \quad (8)$$

donde:

- P_{base} representa un precio base propuesto constante, fijado en \$1.40.
- $\beta[1 - \bar{u}(t)]$ es la componente de escasez; con se establece una relación lineal inversa entre el nivel promedio del agua y el precio: a menor nivel de agua, mayor es el precio.
- $\epsilon(t)$ es un término de ruido estocástico $\epsilon \sim N(0, \sigma^2)$ que simula la volatilidad e incertidumbre derivadas de variaciones independientes de la red hídrica.

Para establecer un vínculo conceptual entre el estado del sistema y la variable económica mostrada en la Ecuación (8), se generó un precio sintético del agua, $P(t)$, que se fundamenta en el principio económico de escasez: a menor disponibilidad de un recurso, mayor es su valor. En nuestro modelo, el nivel promedio del agua en la tubería $\bar{u}(t)$ funciona como un indicador de la oferta o disponibilidad en la red hídrica. Dicha formulación simplificada ilustra la dinámica fundamental que rige los precios de los recursos hídricos en la realidad.

Con el conjunto de datos sintético generado para esta investigación, se construye un modelo de regresión lineal capaz de predecir el precio del agua a partir de las condiciones del sistema. Este algoritmo de regresión lineal modela la relación entre una variable dependiente o de salida y una o más variables independientes o de entrada; ajusta una ecuación lineal a los datos observados (Goodfellow *et al.*, 2016). En este caso se trata de un

problema de regresión múltiple, ya que se utilizan varias variables de entrada para predecir un resultado numérico continuo (Hastie *et al.*, 2009).

Las variables para el modelo se dividieron en dos partes:

- Variables de entrada (Input features), X : las variables predictoras incluyen la variable hidrológica principal y los parámetros físicos que definieron la simulación $X = \{\bar{u}(t), A, \omega, D, k^1\}$.
- Variable de salida (Target) y : es la variable que se desea predecir, en este caso es el precio sintético del agua $y = \{P(t)\}$.

De acuerdo con la literatura, para evitar el sobreajuste y evaluar la generalización del modelo, el conjunto de datos sintéticos de 648,000 registros se dividió en dos subconjuntos: el 80% de los registros se destinó al conjunto de entrenamiento y el 20% restante al conjunto de validación (Guzmán-Torres *et al.*, 2022).

Resultados

En la **Figura 2** se observa un análisis estadístico de los datos sintéticos generados por el modelo, que muestra histogramas de distribución del precio y del nivel promedio del agua, y refleja una fuerte correlación entre la oferta y la demanda, lo que condiciona el precio según el nivel promedio del agua en la red hídrica.

Para entender las relaciones entre las variables, se calculó la matriz de correlaciones de Pearson, la cual se muestra en la **Figura 3**. Esta matriz cuantifica la fuerza y

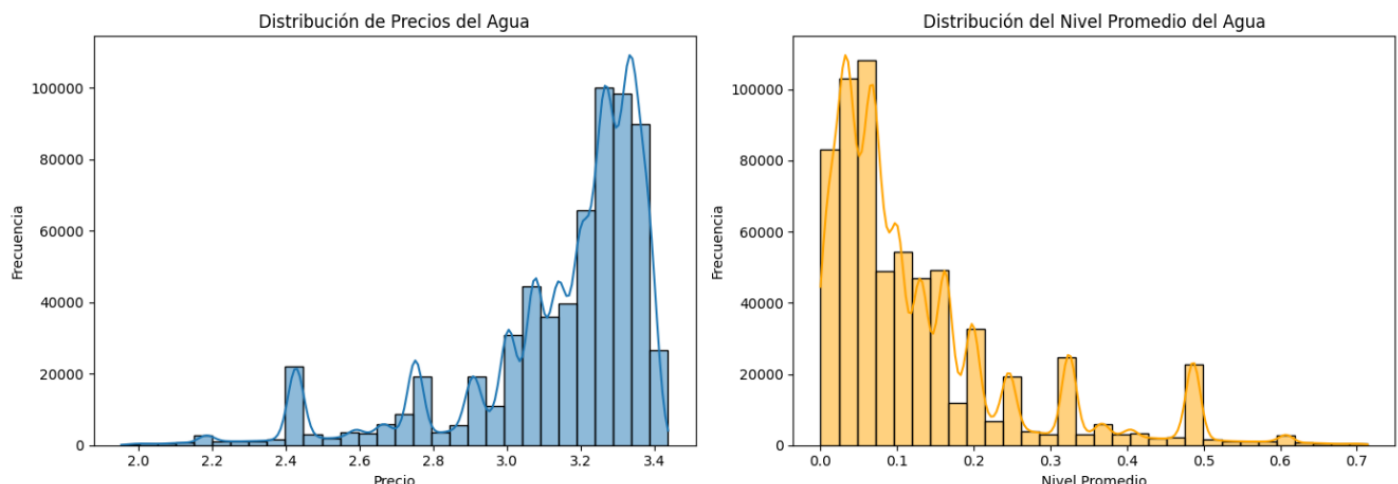


Figura 2. Histogramas de distribución del precio sintético y del nivel promedio de agua en el conjunto de datos.

la dirección de la relación lineal entre pares de variables (Goodfellow *et al.*, 2016). Como era de esperarse y dada la Ecuación (8), se observa una correlación negativa entre *nivel_promedio* y *precio*, lo que confirma que el conjunto de datos refleja la premisa de que, a mayor escasez, mayor es el precio. Sin embargo, las correlaciones entre el precio y los parámetros de la simulación descritos en la **Tabla 1** son débiles, lo que sugiere que su influencia sobre el precio es principalmente indirecta, mediada por su efecto sobre el nivel promedio.

La **Tabla 2** resume las métricas de rendimiento globales del modelo en el conjunto de validación. Estos valores proporcionan una medida agregada de la precisión del modelo. Finalmente, la **Figura 4** ofrece una visualización directa de la precisión del modelo mediante un gráfico de dispersión que compara los precios reales (sintéticos) del conjunto de validación con los precios predichos por el modelo.

En la **Figura 4**, cada punto representa una observación del conjunto de validación. El eje horizontal muestra el precio real y el vertical, el predicho por el modelo. La línea roja diagonal representa la predicción ideal, $y = x$. Se observa que los puntos se ajustan a lo largo de esta línea diagonal, lo que confirma visualmente el valor del R^2 y el bajo error del modelo.

Discusión

A pesar de que al modelo se le proporcionó un conjunto reducido de variables de entrada, datos con un componente

Tabla 2. Métricas de Rendimiento del Modelo Predictivo.

Métrica	Valor
Coefficiente de determinación (R^2)	0.998489
Raíz del Error Cuadrático Medio (RMSE)	0.01002
Error absoluto medio (MAE)	0.00797956
Error Cuadrático Medio (MSE)	9.9975e-05
Error Absoluto Mediano (MedAE)	0.00676025
Error Logarítmico Cuadrático (MSLE)	5.9367e-06

de ruido aleatorio y condiciones ideales en la red hídrica, fue capaz de identificar que la relación dominante y directa con el precio provenía del nivel promedio de agua. El modelo aprendió las correlaciones entre las condiciones físicas que se le habían enseñado implícitamente a través de los datos.

Se evidenció que los parámetros simulados descritos en la **Tabla 1** no influyen directamente en el precio, sino que su efecto es indirecto, ya que determinan el comportamiento dinámico del *nivel_promedio*. Esta capacidad del modelo de Machine Learning para distinguir entre correlaciones directas e indirectas y cuantificar con exactitud la relación principal es una validación que permite el uso de una metodología híbrida. Demuestra que un algoritmo de Machine Learning puede extraer y encapsular las relaciones de un sistema físico a partir de sus datos de salida, creando un modelo sustituto que permite un ahorro de tiempo de ejecución de varios órdenes de magnitud, tal como se busca en la investigación sobre modelos en hidrología (Dai *et al.*, 2024).

En este contexto, un rendimiento alto se espera en un entorno controlado, implementando condiciones ideales como las descritas, donde los datos se generan a partir de una ecuación conocida. La relevancia de este resultado

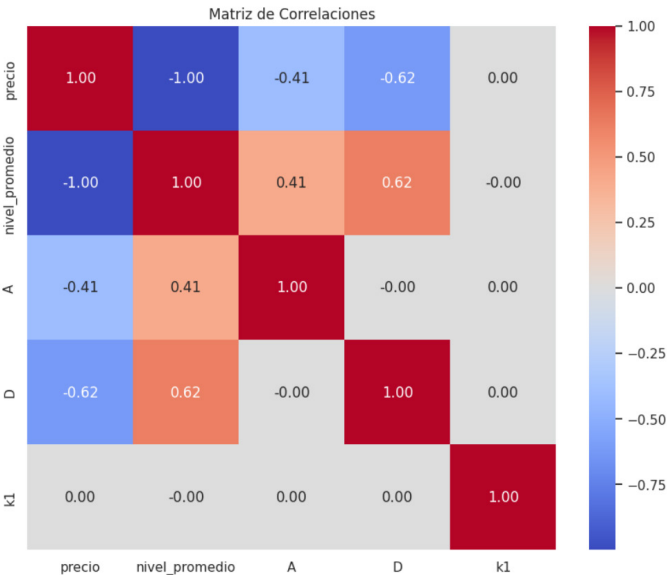


Figura 3. Matriz de correlaciones entre las variables del conjunto de datos sintético.

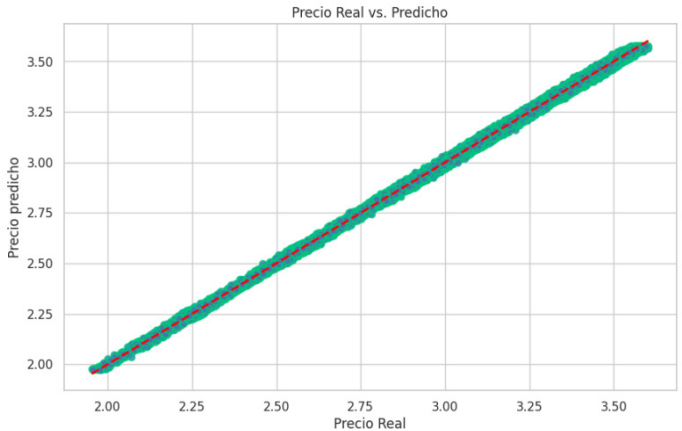


Figura 4. Comparación entre precios reales y predichos en el conjunto de validación.

no reside en una predicción precisa de un fenómeno desconocido, sino en la confirmación de la integridad de la metodología propuesta. Demuestra que la simulación de los parámetros físicos genera datos consistentes y que el modelo de Machine Learning es capaz de aprender de ellos de manera efectiva.

Muchos estudios en este campo utilizan el Machine Learning para corregir sesgos sistemáticos en los modelos físicos o para modelar subprocesos complejos que resultan difíciles de formular matemáticamente (Ampas *et al.*, 2025). El enfoque presentado aquí, sin embargo, representa una aplicación particular con condiciones ideales, lo que permite el uso del modelo físico como generador de datos de entrenamiento basado en la física, lo que facilita la generación de datos y evita inconvenientes asociados a la escasez de datos de campo (Xu *et al.*, 2022).

Conclusiones

Este trabajo ha desarrollado y validado una metodología híbrida que integra la simulación numérica y el Machine Learning para la predicción del precio del agua en México. Se ha demostrado que el enfoque puede producir un modelo predictivo con un coeficiente de determinación R^2 superior al 99% y errores de predicción mínimos en un entorno de datos sintéticos.

La principal contribución del estudio no es solo la alta precisión alcanzada, sino también la validación de una metodología. El modelo de regresión lineal no solo aprendió a predecir el precio, sino que también logró inferir con precisión las leyes físicas implícitas en los datos de entrenamiento. Este resultado confirma que el enfoque de utilizar un modelo físico como generador de datos para entrenar un modelo de Machine Learning es una estrategia viable, especialmente en dominios donde los datos reales son limitados.

La metodología propuesta da como resultado un modelo computacionalmente eficiente y preciso para condiciones ideales en la red hídrica. El modelo propuesto tiene un alto potencial como herramienta de apoyo para la toma de decisiones estratégicas en la gestión de recursos hídricos. Permite a los planificadores y operadores explorar el impacto de múltiples escenarios de oferta (p. ej., cambios en los regímenes de bombeo) y de demanda sobre la dinámica de la red y, consecuentemente, sobre la estructura de precios. Al proporcionar pronósticos rápidos y fiables, esta herramienta puede contribuir a una gestión más proactiva, eficiente y resiliente de los recursos hídricos.

Los resultados obtenidos demuestran la alta precisión del modelo en un entorno controlado con datos sintéticos.

Esto valida la integridad de la metodología híbrida: el modelo físico genera datos consistentes y el algoritmo de Machine Learning aprende las relaciones subyacentes con alta efectividad. Sin embargo, para consolidar la aplicabilidad práctica del presente trabajo, es necesaria una validación con datos empíricos. Como trabajo a futuro, se trabajará en la recolección de datos históricos de una red hídrica real, incluyendo mediciones de presión o caudal y las tarifas o precios del agua correspondientes a ese mismo periodo. Esto permitirá evaluar el rendimiento del modelo frente a un sistema real.

Agradecimientos

Nos gustaría agradecer a AULA CIMNE-Morelia el apoyo tecnológico para este trabajo, así como a la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) por la provisión de recursos financieros.

Expresamos nuestra gratitud a SIIIA MATH y a la Universidad Michoacana de San Nicolás de Hidalgo por la infraestructura computacional.

Referencias

- Burden RL, Faires JD (2011). Análisis Numérico (9a ed.), *Cengage Learning Editores, S.A. de C.V.* ISBN: 9786074816631.
- Dai T, Maher K, Perzan Z (2024). Machine Learning surrogates for efficient hydrologic modeling: Insights from stochastic simulations of managed aquifer recharge. *arXiv* <https://arxiv.org/abs/2407.20902>
- Goodfellow I, Bengio Y, Courville A (2016). Deep Learning: Cambridge, MA, *MIT Press*. p. 800. ISBN: 978-0262035613
- Guzmán-Torres JA, Naser MZ, Domínguez-Mota, FJ (2022). Effective medium crack classification on laboratory concrete specimens via competitive Machine Learning. *Structures* 37:858–870. doi:[10.1016/j.istruc.2022.01.061](https://doi.org/10.1016/j.istruc.2022.01.061)
- Hastie T, Tibshirani R, Friedman J (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.), *Springer*, pp. 43–48. ISBN: 978–0387848570.
- IMCO, (2023) “Costo Real del Agua en México”, Instituto Mexicano para la Competitividad, https://imco.org.mx/wp-content/uploads/2023/08/Investigacion_Costo-real-del-agua-en-Mexico_31082023-1.pdf, [consultado el 23 de junio de 2025].
- Kraft B, Jung M, Körner M, Koirala S, Reichstein M (2022). Towards hybrid modeling of the global hydrological cycle. *Hydrology and Earth System Sciences* 26:1579–1614. doi:[10.5194/hess-26-1579-2022](https://doi.org/10.5194/hess-26-1579-2022).
- LeVeque RJ (2007). Finite Difference Methods for Ordinary and Partial Differential Equations: Steady-State and Time-Dependent Problems:

Philadelphia. *Society for Industrial and Applied Mathematics (SIAM)*, p. 356. ISBN: 978-0-89871-629-0

Olivier A, Zare A, Jafarzadegan K (2025). Integration of hydrological models and Machine Learning techniques for water resources management. *Journal of Civil Engineering Science and Technology* 16(1):1–5. doi:[10.33736/jcest.9191.2025](https://doi.org/10.33736/jcest.9191.2025)

Ampas H, Refanidis I, Ampas V (2025). Hybrid hydrological forecasting through a physical model and a weather-informed transformer model: a case study in Greek Watershed. *Applied Sciences* 15(12):6679. doi:[10.3390/app15126679](https://doi.org/10.3390/app15126679)

Xu T, Longyang Q, Tyson C, Zeng R, Neilson BT (2022). Hybrid physically based and deep learning modeling of a snow-dominated, mountainous, karst watershed. *Water Resources Research* 58:e2021WR030993. doi:[10.1029/2021WR030993](https://doi.org/10.1029/2021WR030993).