



Modelado numérico y clasificación del riesgo ecológico por transporte de contaminantes en cuerpos de aguas superficiales

Numerical modeling and classification of ecological risk from pollutant transport in surface water bodies

María Goretti Fraga López¹ , Gerardo Tinoco Guerrero² , José Alberto Guzmán Torres² ,
Francisco Javier Domínguez Mota^{1,2} 

¹ Escuela de Estudios Superiores Interdisciplinarios, Universidad Michoacana de San Nicolás de Hidalgo.

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

Resumen

El presente trabajo analiza la contaminación de los cuerpos de agua y su impacto en la salud. El objetivo principal es identificar zonas de riesgo ecológico mediante una metodología híbrida que combina la solución numérica y el aprendizaje automático. Se implementa la ecuación de advección-difusión en dos dimensiones para modelar la dispersión de contaminantes, generando un conjunto de datos para entrenar un modelo de clasificación supervisada. Los resultados muestran una alta precisión, superior al 95%, en la predicción de las zonas de riesgo (bajo, medio o alto). Se concluye que este enfoque es una herramienta eficaz para la gestión de la calidad del agua, teniendo así el potencial de adaptarse a condiciones reales.

Palabras clave: contaminación, riesgo ecológico, salud, clasificación, ecuación de advección-difusión.

Abstract

This paper examines water pollution and its health risks. The primary goal is to identify ecological risk zones using a hybrid approach that combines numerical solutions and machine learning. The two-dimensional advection-diffusion equation is used to model pollutant dispersion, creating a dataset for training a supervised classification model. The results demonstrate high accuracy, surpassing 95%, in predicting risk zones (low, medium, or high). It is concluded that this method is an effective tool for water quality management thus having the potential to adapt to real conditions.

Keywords: Pollution, ecological risk, health, classification, advection-diffusion equation.

Introducción

La contaminación de los cuerpos de agua superficiales, como ríos, lagos, presas y, en general, cualquier cuerpo de agua (embalses), representa actualmente una de las mayores problemáticas a resolver a nivel mundial, además de ser una constante amenaza ambiental y de salud pública con un alto índice de mortalidad. Se estima que anualmente se descargan entre 300 y 400 millones de toneladas de metales pesados, solventes y otros desechos industriales sobre los recursos hídricos del planeta, con consecuencias devastadoras para los ecosistemas acuáticos y las comunidades humanas que dependen de ellos (IANAS, 2021). La exposición a contaminantes a través del agua no potable es la principal responsable de millones de muertes y enfermedades, que afectan desproporcionadamente a las poblaciones más vulnerables (IANAS, 2021). El origen de estos focos contaminantes puede ir desde vertidos municipales sin tratar hasta efluentes industriales y escorrentía

Historial

Manuscrito recibido: 31 de octubre de 2025

Manuscrito aceptado: 29 de abril de 2026

Manuscrito publicado: agosto de 2026

✉ Autor para correspondencia

María Goretti Fraga López

1702174b@umich.mx

Como citar

Fraga López MG, Tinoco Guerrero G, Guzmán Torres JA, Domínguez Mota FJ (2026). Modelado numérico y clasificación del riesgo ecológico por transporte de contaminantes en cuerpos de aguas superficiales. *Ciencia Nicolaita* 97:62-70. DOI:[10.35830/cn.vi97.927](https://doi.org/10.35830/cn.vi97.927)

agrícola, los cuales, al combinarse en el medio acuático, producen una compleja mezcla de sustancias tóxicas (MINAMBIENTE, 2015).

Ante el constante crecimiento de los problemas relacionados con los recursos hídricos, se ha implementado el desarrollo de herramientas robustas para el monitoreo, la predicción y la correcta gestión de la calidad del agua, que se ha convertido en una necesidad emergente para la toma de decisiones, así como para la implementación de

acciones optimizadas (Pazmiño, 2020).

Naturalmente, el estudio del transporte de contaminantes en el agua se aborda desde dos paradigmas principales: los modelos basados en procesos físicos y los basados en datos (Ramos Herrera *et al.*, 2015). Los modelos numéricos, basados en ecuaciones diferenciales parciales como la ecuación advección-difusión, describen el movimiento y la dispersión de un soluto a partir de los principios de la dinámica de fluidos (Pazmiño, 2020). Estos modelos ofrecen una representación mecanicista del fenómeno, permitiéndonos simular la dinámica de transporte de un contaminante a través de una corriente (advección) y su dispersión debido a la turbulencia y a los gradientes de concentración (difusión) (Parveen *et al.*, 2006; Hamdi, 2007). A pesar de su rigor físico, su principal desventaja radica en el alto costo computacional, lo que nos limita su aplicación a análisis de escenarios en tiempo real o a la evaluación de un gran número de posibles condiciones de descarga (UN-SPIDER, 2020).

Por otra parte, los modelos basados en datos, impulsados por el auge del aprendizaje automático, han demostrado una capacidad extraordinaria para identificar patrones complejos y no lineales en grandes conjuntos de datos (Font, 2019; Espinoza-Zúñiga, 2020). En el ámbito de la calidad del agua, los algoritmos de ML se utilizan cada vez más para el uso de clasificación y predicción, mostrando un alto rendimiento prometedor (Garzón Barrero *et al.*, 2023; Chen *et al.*, 2024). Sin embargo, la eficacia de este modelo depende directa y críticamente de la disponibilidad de datos de entrenamiento extensos, representativos y debidamente etiquetados, un requisito que a menudo resulta difícil de cumplir en las ciencias ambientales debido a su alto costo y a la complejidad logística que conlleva el muestreo y el análisis de campo (UN-SPIDER, 2020).

Finalmente, el objetivo de este trabajo se centra en el desarrollo e implementación de un modelo numérico bidimensional basados en la ecuación de advección-difusión para la simulación de la dispersión de un contaminante en un cuerpo de agua superficial, para con ello, establecer de manera cuantitativa la clasificación del riesgo ecológico debido a las distintas concentraciones de contaminantes simuladas en niveles de riesgo establecidos (bajo, medio, alto). Una vez hecho eso, nuestro objetivo se centra en construir un conjunto de datos etiquetado a partir de datos de simulación numérica para entrenar y validar un modelo de clasificación de aprendizaje automático supervisado que alcance una precisión del 90% en la identificación de las zonas de riesgo en aguas contaminadas. Así como de manera global, interesa la propuesta de esta metodología híbrida, resultante, para su implementación como herra-

mienta eficiente para el monitoreo y la gestión proactiva de la calidad del agua.

Materiales y métodos

El transporte de contaminantes en medios acuáticos se modela mediante la ecuación de advección-difusión bidimensional (LeVeque, 2007):

$$\frac{\partial C}{\partial t} + u \frac{\partial C}{\partial x} + v \frac{\partial C}{\partial y} = D \left(\frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2} \right) + S - kC, \quad (1)$$

donde:

- $C(x,y,t)$ es la concentración del contaminante [mg/L] en la posición al tiempo.
- u, v son las componentes de velocidad del flujo [m/s] en las direcciones x y y respectivamente.
- D es el coeficiente de difusión [m^2/s].
- S representa los términos fuente [$mg/(L \cdot s)$]
- k es la constante de decaimiento [$1/s$].

Esta ecuación describe tres procesos físicos fundamentales:

1. El transporte advectivo debido al movimiento del fluido, representado por los términos $u(\partial C/\partial x) + v(\partial C/\partial y)$;
2. La dispersión difusiva que tiende a homogeneizar la concentración, modelada por $D\nabla^2 C$; y
3. Los procesos de generación y degradación del contaminante, representados por los términos fuente S y decaimiento kC .

Método de diferencias finitas

La resolución numérica de la ecuación (1) puede realizarse mediante el método de diferencias finitas, que aproxima las derivadas parciales mediante expansiones en series de Taylor truncadas. Este método transforma la ecuación diferencial parcial continua en un sistema de ecuaciones algebraicas discretas que pueden resolverse computacionalmente.

El dominio computacional bidimensional se discretiza mediante una malla rectangular estructurada con espaciamiento uniforme Δx y Δy en las direcciones x y y , respecti-

vamente. Los puntos de la malla se representan como (i,j) donde $i = 0,1,\dots,N_x$ y $j = 0,1,\dots,N_y$, correspondiendo a las posiciones espaciales $x_i = i\Delta x$ y $y_j = j\Delta y$. El tiempo se discretiza con un paso temporal Δt , donde $t^n = n\Delta t$ para $n = 0,1,2,\dots,N_t$.

La concentración discreta se denota como $C_{i,j}^n \approx C(x_i, y_j, t^n)$, representando la concentración en el punto de malla (i,j) en el tiempo t^n .

Para la discretización temporal de este problema se emplea un esquema de diferencias finitas hacia adelante (forward Euler) para la derivada temporal:

$$\frac{\partial C}{\partial t} \Big|_{i,j}^n = \frac{C_{i,j}^{n+1} - C_{i,j}^n}{\Delta t} + O(\Delta t), \quad (2)$$

donde $O(\Delta t)$ representa el error de truncamiento de primer orden en el tiempo.

Al respecto de la discretización espacial, los términos advectivos se aproximan mediante diferencias centradas de segundo orden para minimizar la dispersión numérica:

$$u \frac{\partial C}{\partial x} \Big|_{i,j}^n = u \frac{C_{i+1,j}^n - C_{i-1,j}^n}{2\Delta x} + O((\Delta x)^2), \quad (3)$$

$$v \frac{\partial C}{\partial y} \Big|_{i,j}^n = v \frac{C_{i,j+1}^n - C_{i,j-1}^n}{2\Delta y} + O((\Delta y)^2). \quad (4)$$

Por su parte, las segundas derivadas espaciales se aproximan mediante diferencias centradas:

$$\frac{\partial^2 C}{\partial x^2} \Big|_{i,j}^n = \frac{C_{i+1,j}^n - 2C_{i,j}^n + C_{i-1,j}^n}{(\Delta x)^2} + O((\Delta x)^2), \quad (5)$$

$$\frac{\partial^2 C}{\partial y^2} \Big|_{i,j}^n = \frac{C_{i,j+1}^n - 2C_{i,j}^n + C_{i,j-1}^n}{(\Delta y)^2} + O((\Delta y)^2), \quad (6)$$

Sustituyendo las aproximaciones (2) – (6) en la ecuación (1), se obtiene el esquema explícito:

$$C_{i,j}^{n+1} = C_{i,j}^n + \Delta t \left[-\frac{C_{i+1,j}^n - C_{i-1,j}^n}{2\Delta x} - v \frac{C_{i,j+1}^n - C_{i,j-1}^n}{2\Delta y} + D \left(\frac{C_{i+1,j}^n - 2C_{i,j}^n + C_{i-1,j}^n}{(\Delta x)^2} + \frac{C_{i,j+1}^n - 2C_{i,j}^n + C_{i,j-1}^n}{(\Delta y)^2} \right) + S_{i,j}^n - kC_{i,j}^n \right]. \quad (7)$$

La estabilidad del esquema explícito requiere que se satisfagan simultáneamente las condiciones de Courant-Friedrichs-Lewy (CFL) para la advección y la condición de estabilidad para la difusión.

La condición CFL dice que la información numérica no puede propagarse más rápido que la información física:

$$CFL = \max \left(\frac{|u|\Delta t}{\Delta x}, \frac{|v|\Delta t}{\Delta y} \right) \leq C_{CFL}, \quad (8)$$

donde $C_{CFL} = 0.5$ para el esquema centrado utilizado en este trabajo. Esta condición asegura que el dominio de dependencia numérica incluya el dominio de dependencia física.

Para la difusión, la condición de estabilidad (número de Fourier) es:

$$F_o = D\Delta t \left(\frac{1}{(\Delta x)^2} + \frac{1}{(\Delta y)^2} \right) \leq 0.25. \quad (9)$$

Esta condición garantiza que el operador de difusión discreto sea estable y que no se generen oscilaciones espurias en la solución numérica.

Considerando ambas condiciones, es necesario que el paso de tiempo cumpla con:

$$\Delta t = \min \left(\frac{C_{CFL}}{\max \left(\frac{|u|}{\Delta x}, \frac{|v|}{\Delta y} \right)}, \frac{0.25}{D \left(\frac{1}{(\Delta x)^2} + \frac{1}{(\Delta y)^2} \right)} \right). \quad (10)$$

Para que el problema pueda generar una solución única y

estable, también es necesario contar con condiciones de frontera. Para este problema se implementan condiciones de frontera mixtas para representar diferentes escenarios físicos:

Por un lado, se utilizan condiciones de Dirichlet que especifican la concentración en la frontera, las cuales pueden representarse como:

$$C(x, y, t) = C_0(x, y, t) \in \Gamma_D. \quad (11)$$

Además, también se utilizó una condición de frontera de tipo Neumann, la cual especifica el flujo normal en dicha frontera:

$$\frac{\partial C}{\partial \hat{n}} = q_n(x, y, t) \in \Gamma_N, \quad (12)$$

donde $\hat{n}$ es el vector normal unitario hacia el exterior de las fronteras. La implementación numérica de estas condiciones se realiza mediante diferencias hacia atrás o hacia adelante, según la orientación de la frontera. En nuestro caso,

$$\frac{C_{N_x, j}^n - C_{N_x-1}^n}{\Delta x} = q_n. \quad (13)$$

El dominio computacional que se usa para este trabajo representa un tramo de río con:

- Dimensiones de 200 m x 10 m.
- Resolución espacial $\Delta x = 5 \text{ m}$ y $\Delta y = 0.25 \text{ m}$
- Paso temporal $\Delta t = 0.05 \text{ s}$
- Tiempo total de simulación de 300 s.
- Coeficiente de difusión $D = 0.1 \text{ m}^2/\text{s}$.
- Constante de decaimiento $k = 0.001 \text{ s}^{-1}$.

Esta configuración permite una representación detallada del transporte de contaminantes en un canal fluvial típico, con una relación de aspecto realista de 20:1 que refleja la geometría elongada propia de los sistemas fluviales.

Escenarios y características modeladas

Para la simulación, se generaron 33 escenarios variando

sistemáticamente los parámetros hidrodinámicos: 6 configuraciones de velocidad de flujo (muy bajo: 0.1 m/s, bajo: 0.2 m/s, moderado: 0.5 m/s, alto: 0.7 m/s, muy alto: 1.0 m/s, línea base, y extremo: 1.5 m/s) y 5 niveles de tasa de descarga (muy baja: 50 mg/s, baja: 100 mg/s, media: 200 mg/s, alta: 400 mg/s, muy alta: 800 mg/s), cada uno evaluado en 3 ubicaciones estratégicas de la fuente: frontera inferior ($x = 20, y = 1$), centro-izquierda ($x = 1, y = 5$), y frontera superior ($x = 20, y = 9$). El modelo registró los valores de la función en cada nodo de la malla para cada paso de tiempo, por lo que el conjunto de datos generado para entrenamiento, validación y prueba contiene un total de 316,800,000 datos. Estas configuraciones permiten evaluar el comportamiento del transporte en diversas condiciones hidrodinámicas y en posiciones de fuente representativas de distintos escenarios de contaminación accidental o puntual. Es de ahí que se extraen dos conjuntos de características de las simulaciones numéricas para alimentar los algoritmos de machine learning:

Características Fundamentales

1. Concentración máxima.
2. Concentración promedio.
3. Área afectada (concentración $> 0.1 \text{ mg/L}$).
4. Distancia máxima de propagación.
5. Velocidad de flujo.
6. Tasa de descarga.
7. Posición x de la fuente.
8. Posición y de la fuente.

Características Derivadas Adicionales

1. Gradiente máximo de concentración.
2. Momento de inercia espacial.
3. Índice de dispersión.
4. Tiempo al pico de concentración.
5. Tasa de decaimiento efectiva.
6. Relación del aspecto de la pluma.

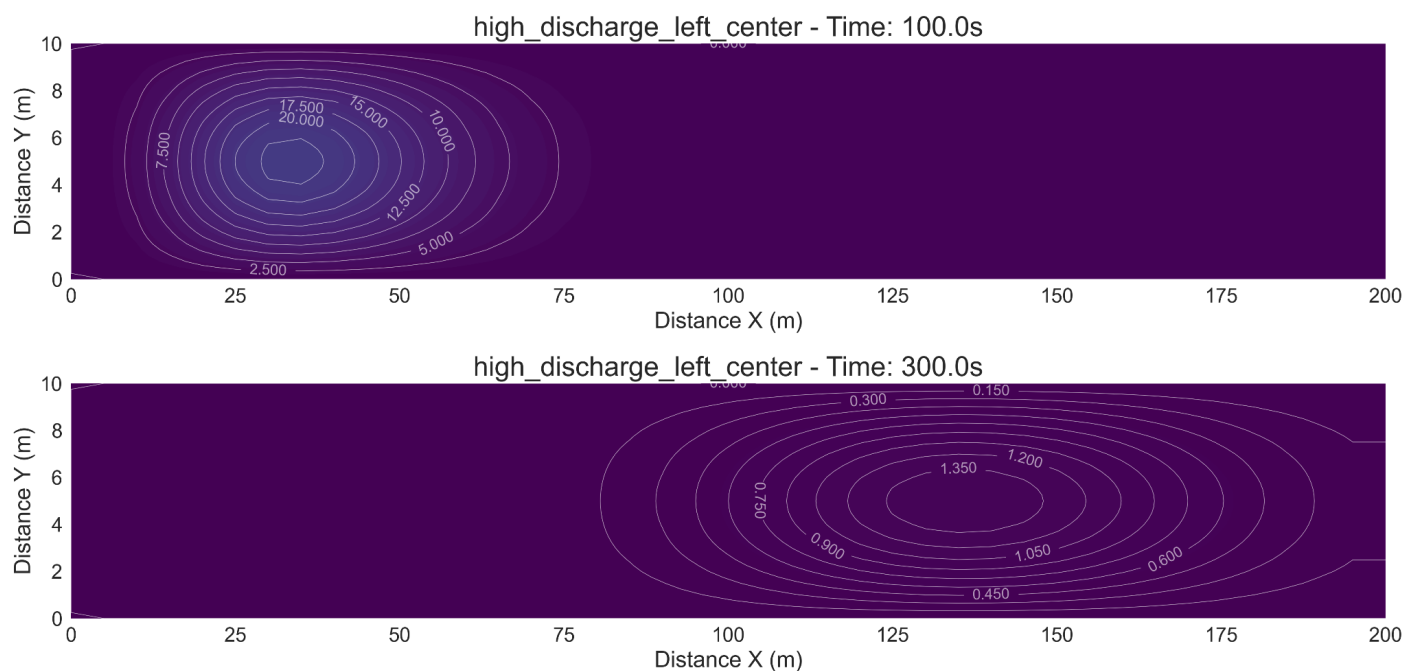


Figura 1. Escenario de simulación numérica del esquema de diferencias finitas, donde el contaminante se coloca del lado izquierdo con una descarga intensa.

7. Concentración en el centroeide.

8. Varianza espacial de concentración.

Se establecen tres niveles de riesgo basados en umbrales de concentración ecológicamente relevantes:

- Riesgo bajo ($C \leq CCC$).
- Riesgo medio ($CCC < C \leq CMC$).
- Riesgo alto ($C > CMC$).

C es la concentración ambiental.

CCC es la Concentración Continua de Criterio.

CMC es la Concentración Máxima de Criterio.

Para usar estos criterios de clasificación, es necesario consultar los valores de CMC y CCC en las tablas de la EPA de EE.UU. correspondientes a cada contaminante de interés (Aquatic Life Criteria Table | US EPA, 2015).

Los umbrales de clasificación de riesgo ecológico para este estudio se basan en los *Criterios de Calidad del Agua para la Vida Acuática* establecidos por la Agencia de Protección Ambiental de Estados Unidos (EPA, 2015). Estos criterios son responsables de la distinción toxicológica entre los efectos agudos y crónicos, los cuales son fundamentales para la correcta protección de

los ecosistemas acuáticos. Adicionalmente, es importante mencionar que, para asegurar la estabilidad numérica del método híbrido, se normalizaron los umbrales de riesgo y de concentración para una mayor robustez en el aprendizaje del modelo.

Estos umbrales reflejan niveles de sensibilidad ecológica típicos de contaminantes en ecosistemas acuáticos, donde concentraciones del orden de microgramos por litro pueden tener efectos significativos en la biota acuática. La **Figura 1** muestra una de las simulaciones donde se aprecia la evolución de la propagación del contaminante desde la parte izquierda, con un caudal intenso.

La **Figura 2** ilustra la evolución espaciotemporal del mapa de riesgo ecológico para un escenario de descarga intensa desde la frontera izquierda. La imagen superior es una instantánea en el tiempo $t = 100\text{ s}$, donde la zona de alto riesgo (rojo) se concentra cerca del punto de vertido. La imagen inferior, en $t = 300\text{ s}$, muestra cómo el riesgo se ha transportado aguas abajo y se ha dispersado, lo que ha expandido significativamente el área afectada por niveles de riesgo medio y alto.

ALGORITMOS DE MACHINE LEARNING

Se implementaron cuatro algoritmos de clasificación supervisada, seleccionados por su complementariedad en el manejo de distintos tipos de relaciones en los datos y por su robustez frente a problemas de clasificación multiclase. La selección abarca desde métodos de ensemble hasta modelos lineales, proporcionando una evaluación comple-

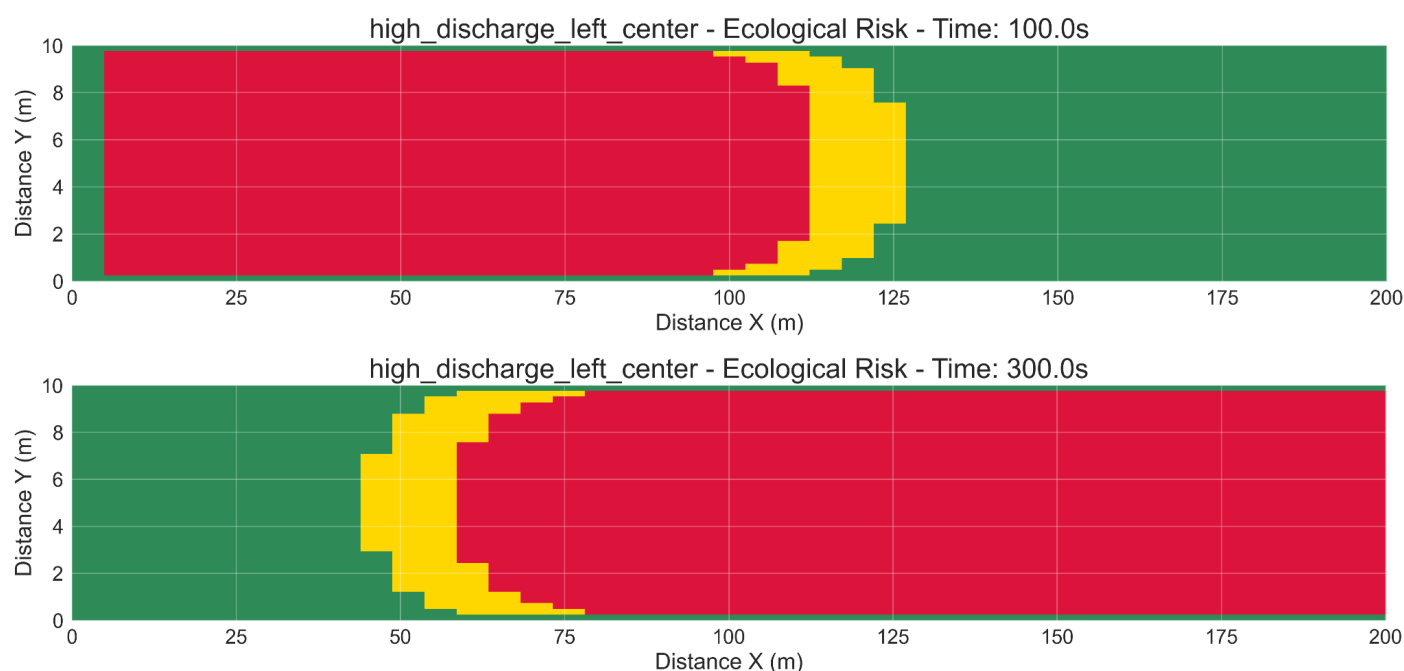


Figura 2. Visualización del proceso de advección y difusión en la simulación numérica.

ta de enfoques algorítmicos distintos (Goodfellow *et al.*, 2016).

Support Vector Machine (SVM)

Algoritmo enfocado en encontrar el hiperplano óptimo para separar dos clases. Cuando los datos no son separables, utiliza el “truco del Kernel” para transformarlos a una dimensión superior hasta encontrar un separador lineal.

Random Forest

Conjunto de árboles de decisión que utiliza submuestras aleatorias y la votación mayoritaria para lograr predicciones precisas y robustas.

Gradient Boosting

Es un método de ensamblaje que construye modelos de forma secuencial y aditiva.

Logistic Regression

Modelo estadístico de clasificación que estima la probabilidad de un resultado binario aplicando la función logística (sigmoide) a una combinación lineal de variables para modelar el logaritmo de las probabilidades (log-odds).

La evaluación comparativa se realiza mediante validación cruzada estratificada con $k = 5$ pliegues, asegurando la re-

presentación proporcional de cada clase de riesgo en cada pliegue. Se implementa la búsqueda en grilla (GridSearch-CV) para optimizar hiperparámetros en los algoritmos más sensibles (Random Forest y SVM), utilizando el conjunto de validación para evitar el sobreajuste en la selección de modelos.

La selección del modelo final se basa en múltiples métricas: precisión global, precisión por clase, recall, F1-score y análisis de la matriz de confusión, priorizando el equilibrio entre todas las clases de riesgo en aplicaciones de gestión ambiental, donde tanto los falsos positivos como los falsos negativos tienen implicaciones críticas.

La evaluación se realizó mediante validación cruzada estratificada ($k = 5$) utilizando las métricas: precisión (Accuracy) como fracción de predicciones correctas:

- Precisión por clase: $TP / (TP + FP)$,
- Recall: $TP / (TP + FN)$,
- F1-Score: $2 \times (Precisión \times Recall) / (Precisión + Recall)$.

Este conjunto de datos simulados, aunque sintético, está fundamentado en las leyes físicas del transporte de masa, lo que nos proporciona una base completa para entrenar un modelo de aprendizaje automático. Este es un enfoque muy importante, ya que no solo soluciona el problema de la escasez de datos, sino que también nos asegura de que

Tabla 1. Resultados de los clasificadores

Modelo	Accuracy	Precisión (Macro)	Recall (Macro)	F1-Score (Macro)
Random Forest	0.9968	0.9951	0.9893	0.9922
SVM	0.9625	0.9358	0.9167	0.9256
Gradient Boosting	0.9997	0.9992	0.9995	0.9994
Logistic Regression	0.8662	0.5779	0.6301	0.6029

nuestro modelo de Machine Learning aprenda patrones específicos de contaminación que sean físicamente realistas.

Resultados

Los resultados de la evaluación de los cuatro algoritmos de machine learning se presentan en la **Tabla 1**. El análisis se realizó tanto con características fundamentales como con el conjunto completo de características.

El análisis de importancia de características, mostrado en la **Figura 3**, revela que las variables más influyentes

para la clasificación de riesgo son, en este orden: Intensidad de la fuente, distancia a la fuente, posiciones espaciales y Parámetros hidrodinámicos de velocidad. Las características derivadas, como los números de Péclet y los tiempos de viaje, proporcionan información complementaria sobre el régimen de transporte dominante.

El análisis detallado por clase de riesgo para el mejor modelo (Gradient Boosting) muestra un rendimiento muy prometedor:

- Riesgo Bajo:
 - » Precisión: 1.0000.
 - » Recall: 0.9996.
 - » F1-Score: 0.9998.
- Riesgo Medio:
 - » Precisión: 0.9978.
 - » Recall: 0.9989.
 - » F1-Score: 0.9984.
- Riesgo Alto:
 - » Precisión: 0.9998.
 - » Recall: 1.0000.
 - » F1-Score: 0.9999.

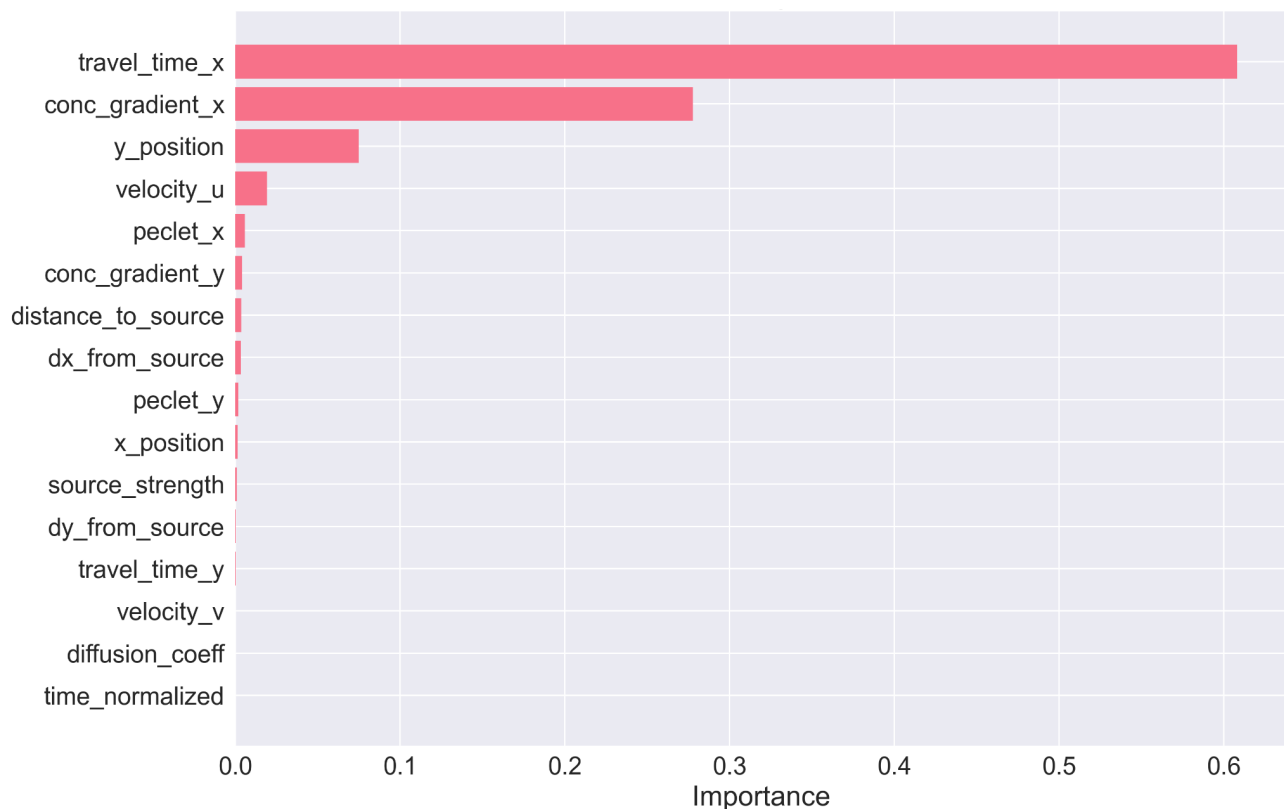


Figura 3. Gráfico de la importancia relativa de las características para la predicción del riesgo ecológico.

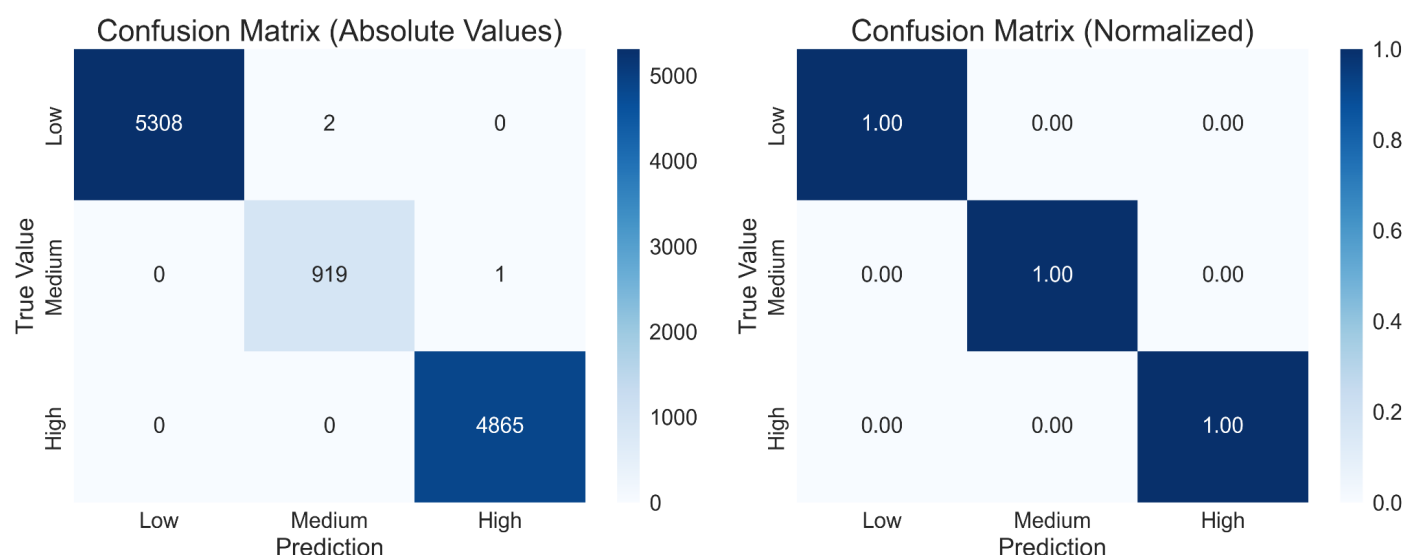


Figura 4. Matriz de confusión.

Discusión

Los resultados muestran la eficacia de la metodología híbrida propuesta para la evaluación automatizada del riesgo ecológico. Gradient Boosting resulta ser el algoritmo más efectivo, alcanzando una precisión del 99.97% con métricas macro de precisión (99.92%), recall (99.95%) y F1-score (99.94%). Random Forest también muestra un rendimiento prometedor con 99.68% de precisión, mientras que SVM alcanza 96.25%, y Logistic Regression presenta el rendimiento más bajo, con 86.62%, particularmente afectado por dificultades en la clasificación de la clase de riesgo medio.

El análisis de importancia de características indica que la intensidad de la fuente y la distancia a la fuente son los predictores más significativos del riesgo ecológico, lo cual es consistente con los principios del transporte de contaminantes. La relevancia de las posiciones espaciales y de los parámetros hidrodinámicos refleja la importancia de la configuración geométrica y de las condiciones de flujo en la determinación del riesgo. Los números de Péclet proporcionan información valiosa sobre el régimen de transporte dominante (advección vs. difusión).

La metodología presenta ventajas significativas, como la automatización, que elimina la evaluación manual experta; rapidez mediante evaluación instantánea una vez entrenados los modelos; consistencia mediante criterios uniformes y reproducibles; escalabilidad para procesar múltiples escenarios simultáneamente; e interpretabilidad mediante la identificación de factores críticos.

Sin embargo, existen limitaciones: dependencia de la representatividad de los datos de entrenamiento, simplificaciones del modelo físico 2D que no captan completa-

mente la complejidad tridimensional, umbrales de riesgo fijos que pueden requerir ajustes contextuales y necesidad de validación experimental con datos de campo.

Las aplicaciones prácticas de este trabajo pueden incluir el monitoreo ambiental en tiempo real, la evaluación rápida del impacto ambiental, los sistemas de alerta temprana para la contaminación, la optimización de estrategias de remediación y la planificación de la ubicación de industrias.

Conclusiones

Este trabajo presenta una metodología híbrida exitosa que integra la simulación numérica del transporte de contaminantes con técnicas de machine learning para la evaluación automatizada del riesgo ecológico. Los principales logros incluyen: desarrollo de un framework computacional robusto que combina modelado físico riguroso con algoritmos de clasificación avanzados, una muestra de precisión predictiva prometedora con Gradient Boosting alcanzando 99.97% y Random Forest 99.68%, identificación de características críticas donde la intensidad de la fuente y la distancia a la fuente resultan ser los predictores más importantes, junto con parámetros hidrodinámicos y números de Péclet que caracterizan el régimen de transporte, validación de la superioridad de métodos ensemble sobre enfoques lineales, y establecimiento de una base metodológica para la evaluación automatizada de riesgo ambiental con potencial de extensión a otros sistemas.

La metodología propuesta representa un avance significativo en la automatización de la evaluación de riesgo ecológico, ofreciendo una herramienta valiosa para la ges-

tión ambiental proactiva. Los resultados muestran que la integración del modelado físico con el machine learning puede proporcionar evaluaciones rápidas, precisas y consistentes del riesgo de contaminación.

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 (SECIH-TI) 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

Espinosa-Zúñiga JJ (2020). Aplicación de algoritmos Random Forest y XGBoost en una base de solicitudes de tarjetas de crédito. *Ingeniería, Investigación y Tecnología* 21(3):1–13. doi:[10.22201/fi.25940732e.2020.21.3.022](https://doi.org/10.22201/fi.25940732e.2020.21.3.022).

Font X (2019). Técnicas de clasificación supervised learning: *Editorial UOC*, Barcelona, p. 62. PID_00264729.

Garzón Barrero J, Sánchez Pineda NE, Londoño Pinilla DF (2023). Evaluación comparativa de los algoritmos de aprendizaje automático Support Vector Machine y Random Forest: efectos del tamaño del conjunto de entrenamiento. *Ciencia e Ingeniería Neogranadina* 33(2):131–148. doi:[10.18359/rcin.6996](https://doi.org/10.18359/rcin.6996).

Goodfellow I, Bengio Y, Courville A (2016). *Deep Learning*: Cambridge, MA, MIT Press. P.800. ISBN: 978-0262035613

Hamdi A (2007). Identification of point sources in two-dimensional advection-diffusion-reaction equation: Application to pollution sources in a river. Stationary case. *Inverse Problems in Science and Engineering* 15(8):885–870. doi:[10.1080/17415970601162198](https://doi.org/10.1080/17415970601162198).

Chen W, Xu D, Pan B, Zhao Y, Song Y (2024). Machine Learning-Based Water Quality Classification Assessment. *Water* 16(20):2951. doi:[10.3390/w16202951](https://doi.org/10.3390/w16202951).

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

Ministerio de Ambiente y Desarrollo Sostenible de Colombia (2015). “Resolución 0631 de 2015”, *Minambiente* <https://www.minambiente.gov.co/wp-content/uploads/2021/11/resolucion-631-de-2015.pdf>, [consultado el 25 de julio de 2025].

Parveen N, Abbasi T, Arya A (2018). A Modelling Framework For Simulating River y Stream Water Quality. *International Journal of Advance Research in Science and Engineering* 7(1):382–388.

Pazmiño JC (2020). Fundamentos de Modelización de la Calidad del Agua: Editorial Grupo Compás, Guayaquil, Ecuador. P. 165 ISBN: 978-9942-33-277-6.

Ramos Herrera S, Magaña Villegas E, Carrera Velueta JM (2015). Introducción a la modelación del agua, del aire y del transporte de contaminantes en el suelo: *Universidad Juárez Autónoma de Tabasco, Villahermosa*, México, p. 126. ISBN: 978-607-606-208-1.

Red Interamericana de Academias de Ciencias (IANAS) (2021). Calidad del agua: Salud de los ecosistemas y salud humana. *InterAmerican Network of Academies of Sciences*. https://www.interacademies.org/sites/default/files/2021-09/Libro_Calidad_del_Agua-1.pdf, [consultado el 25 de julio de 2025]

UN-SPIDER (s.f.). Aplicación de Datos del Mes: Aprendizaje Automático Para La Detección de Inundaciones, *Plataforma de las Naciones Unidas de Información Obtenida desde el Espacio para la Gestión de Desastres y la Respuesta a Emergencias*, <https://www.un-spider.org/es/links-and-resources/data-sources/daotm-floods-ml>, [consultado el 25 de julio de 2025].

US EPA (s.f.). Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses, *US EPA*. <https://www.epa.gov/wqc/guidelines-deriving-numerical-national-water-quality-criteria-protection-aquatic-organisms-and>, [consultado el 7 de septiembre de 2025].

US EPA (s.f.). National Recommended Water Quality Criteria - Aquatic Life Criteria Table, *US EPA*. <https://www.epa.gov/wqc/national-recommended-water-quality-criteria-aquatic-life-criteria-table>, [consultado el 7 de septiembre de 2025].