

Modelo de Machine Learning para la priorización de zonas críticas de tres cuencas hidrográficas de Cundinamarca mediante Soluciones Basadas en la Naturaleza frente al cambio climático.

Karol Stefany Mora Hortua

Edgar Andres Gutierrez Palomino

Cristian David Báez Bermúdez

Asesor

Leon Velasquez Elizabeth

Universidad EAN - Escuela de Administración de Negocios

Seminario Investigación

2025

Resumen

Los ecosistemas hídricos de Cundinamarca se ven impactados por el cambio climático, generando riesgos en la disponibilidad de agua, la regulación hídrica y la resiliencia de los sistemas socioecológicos. En este contexto, las Soluciones basadas en la Naturaleza (SbN) se consolidan como estrategias costo-efectivas para mitigar impactos y restaurar la capacidad adaptativa de las cuencas. Sin embargo, la identificación de áreas prioritarias de intervención continúa siendo un reto técnico y metodológico. En este trabajo se propone una metodología de priorización de zonas críticas en cuencas de Cundinamarca afectadas por el cambio climático mediante la integración de técnicas de machine learning (Random Forest y XGBoost) con sistemas de información geográfica para así brindar SbN eficientes, para esto se consideran variables biofísicas y climáticas, junto a información espacial, las cuales permiten caracterizar las condiciones ambientales y climáticas del territorio, con el fin de entrenar modelos predictivos robustos orientados a identificar zonas con mayor vulnerabilidad ambiental y niveles de presión antrópica. La evaluación del desempeño se realizará con métricas de precisión, sensibilidad y área bajo la curva ROC, complementada con la generación de mapas de susceptibilidad.

Palabras Clave: Cuencas, Sostenibilidad, Clima, Naturaleza, Metodología, Ecosistemas

Abstract

The aquatic ecosystems of Cundinamarca are increasingly affected by climate change, leading to risks in water availability, hydrological regulation, and the resilience of socio-ecological systems. In this context, Nature-based Solutions (NbS) emerge as cost-effective strategies to mitigate impacts and restore the adaptive capacity of watersheds. However, identifying priority intervention areas remains a technical and methodological challenge. This study proposes a methodology for prioritizing critical zones in Cundinamarca's watersheds affected by climate change through the integration of machine learning techniques (Random Forest and XGBoost) with geographic information systems (GIS) to design efficient NbS. The approach considers biophysical and climatic variables, along with spatial information, to characterize the environmental and climatic conditions of the territory and to train robust predictive models aimed at identifying areas with greater environmental vulnerability and anthropic pressure. Model performance will be evaluated using accuracy, sensitivity, and area under the ROC curve (AUC), complemented by the generation of susceptibility maps.

Keywords: Watersheds, Sustainability, Climate, Nature, Methodology, Ecosystems

Tabla de Contenido

Justificación	7
Objetivos.....	8
Objetivo General.....	8
<i>Objetivos Específicos</i>	8
Planteamiento del problema.....	9
Antecedentes del problema.....	9
Descripción del problema.....	10
Pregunta de investigación.....	11
Marco Teórico.....	12
Marco Institucional	22
Metodología	24
Enfoque, alcance y diseño de la investigación	24
Definición de Variables.....	25
Población y Muestra	28
Selección de métodos o instrumentos para recolección de información	29
Técnicas de análisis de datos y Muestra	30
Análisis y discusión de resultados	35
Conclusiones.....	43
Referencias.....	48

Lista de tablas

Tabla 1	26
Tabla 2	33
Tabla 3	36
Tabla 4	41

Lista de ilustraciones

Ilustración 1	38
----------------------------	-----------

Justificación

La implementación de Soluciones Basadas en la Naturaleza (SbN) es una tendencia de los últimos años, que se debe fortalecer para mitigar los impactos del cambio climático, ya que enfrenta diferentes desafíos, es por esto que el actual proyecto busca contribuir con tecnología innovadora, por medio del diseño de un modelo de priorización apoyado en datos ambientales, climáticos y espacial, integrados con algoritmos de aprendizaje automático, permitiendo identificar de manera objetiva los territorios donde la intervención es más urgente, como en el caso de Cundinamarca donde hay un amplio panorama de cuencas que necesitan de intervención para reducir los impactos de estos fenómenos.

Por lo que es fundamental priorizar las zonas críticas de las cuencas hidrográficas para así lograr una correcta intervención y restaurar el ecosistema sin afectar la biodiversidad mediante SbN; este tipo de soluciones salvaguardan los medios de vida de las comunidades, reduciendo riesgos de inundaciones, deslizamientos y fortaleciendo la resiliencia frente al cambio climático, además se alinea a la Gestión Integral del Recurso Hídrico de Colombia el cual busca el manejo coordinado y sostenible del agua, y otros recursos naturales relacionados, para maximizar el bienestar social y económico de forma equitativa, de tal forma que no se afecte la salud del ecosistema.

Finalmente, el proyecto tendrá implicaciones académica en primer lugar, proporcionando a entidades ambientales, como las Corporaciones Autónomas Regionales (CAR), una herramienta concreta para optimizar inversiones en adaptación climática. Por otro, aportará al desarrollo interdisciplinario al integrar inteligencia artificial con ecología y ciencias sociales, generando un modelo replicable y escalable en otros contextos geográficos y en coherencia con los principios de sostenibilidad de la Universidad EAN.

Objetivos

Objetivo General

Desarrollar un modelo de aprendizaje automático para la priorización de zonas críticas frente al cambio climático considerando tres cuencas hidrográficas previamente categorizadas con mayor riesgo del departamento de Cundinamarca, con el fin de optimizar la implementación de Soluciones Basadas en la Naturaleza (SbN).

Objetivos Específicos

Analizar los datos ambientales y socio ecológicos de las cuencas hidrográficas de Cundinamarca relevantes para la identificación de zonas críticas con relación al cambio climático.

Aplicar los algoritmos de clasificación supervisada (como Random Forest o XGBoost) para determinar zonas prioritarias de intervención.

Proponer un modelo de machine learning que permita la priorización de las zonas críticas con los datos recolectados y algoritmos usados.

Planteamiento del problema

Antecedentes del problema

El cambio climático es uno de los desafíos ambientales del siglo XXI, al crecer la presión sobre los recursos hídricos se producen fenómenos extremos. Como consecuencia, ha aumentado el riesgo en las cuencas hidrográficas, la calidad del agua, la biodiversidad y la estabilidad ecológica (World Wildlife Fund, s.f.). Por lo que restaurar y proteger los ecosistemas por medio de las Soluciones Basadas en la Naturaleza (SbN), surge como una forma más sostenible en comparación con la infraestructura gris. Por ejemplo Chausson et al. (2020), mencionan que este tipo de propuestas no solo mitiga el cambio climático, sino que genera múltiples beneficios colaterales en los ecosistemas.

No obstante, la implementación de SbN enfrenta la ausencia de metodologías técnicas que faciliten la priorización de zonas de intervención críticas. Adicionalmente la expansión agropecuaria y la urbanización amenazan constantemente los ecosistemas causando una situación aún más desafiante. En este contexto, una intervención sin un proceso de planificación estructurado y sin la definición de criterios ecológicos, genera resultados limitados en términos de efectividad y sostenibilidad (Chausson et al., 2020). Por lo que es indispensable el desarrollo de herramientas de diagnóstico espacial para una solución con mayor potencial (Roldán Valcarce, 2024).

Con lo anterior, un modelo de Machine Learning (ML) es un gran potencial para optimizar la clasificación de zonas críticas para SbN, pero Richards et al., (2024) señalan que la fiabilidad de esta tecnología depende de la calidad de la información proporcionada, sin embargo, el uso de datos sesgados limita la aplicabilidad de esta solución. Un caso de éxito es la restauración de manglares en Samborondón, Ecuador, el cual expone el potencial de SbN junto a

diagnósticos estratégicos (Programa de las Naciones Unidas para el Medio Ambiente, 2025), sin embargo, se carece de un modelo escalable que permita replicar y adaptar estas experiencias en otros ecosistemas.

En Colombia, las cuencas hidrográficas son impactadas por el aumento poblacional y las actividades humanas; en el departamento de Cundinamarca se ha observado un patrón de degradación ambiental debido a la ausencia de coordinación y planificación eficaz (Aguirre et al., 2017) . La Gobernación de Cundinamarca (2024), realizó un diagnóstico ambiental con fuentes del IDEAM donde Girardot, Puerto Salgar y Guaduas encabezan el ranking de los municipios con mayor presión ambiental, marcando la necesidad de contar con herramientas tecnológicas que faciliten la toma de decisiones estratégicas para mitigar estos riesgos.

Descripción del problema

Ante la falta de una metodología técnica y robusta que esté basada en datos congruentes, para la priorización espacial de intervenciones de Soluciones Basadas en la Naturaleza (SbN) representa una brecha operativa en la planificación ambiental de las cuencas hidrográficas de Cundinamarca. Aunque los estudios de vulnerabilidad son comunes, la mayoría de los instrumentos de gestión ambiental no soportan la capacidad analítica para procesar grandes volúmenes de datos y, por ende, modelar de manera predictiva la combinación de riesgos ambientales, climáticos y sociales. Según el artículo (Sauer & Chang, 2025) demostraron que factores como el aumento de la temperatura y la urbanización afectan de manera diferenciada la calidad del agua, lo que subraya la importancia crítica de la localización de las intervenciones. Los esfuerzos de restauración en Colombia a menudo se implementan de forma reactiva, perdiendo la oportunidad de maximizar el impacto de las inversiones.

El problema radica en que, sin herramientas de priorización objetivas, la toma de decisiones sobre la implementación de SbN es a menudo subjetiva y poco eficiente, lo que limita su capacidad para mitigar los efectos del cambio climático. Esta investigación es pertinente a las disciplinas de ingeniería, ciencias ambientales y ciencias de datos, ya que busca integrar estos campos para desarrollar un modelo de aplicación concreto, con esta combinación de datos: ambientales, topográficos, climáticos y socioeconómicos, se abre la posibilidad de aplicar algoritmos de Machine Learning que permitan modelar la vulnerabilidad ambiental de manera automatizada.

La ausencia de una metodología integral y una herramienta de priorización representa el vacío principal que esta investigación se propone solucionar. Con esto no solo se valida la relevancia del proyecto, sino que también señala la capacidad para aportar una solución estratégica a la planificación ambiental.

Pregunta de investigación

¿Cómo puede un modelo de aprendizaje automático contribuir a la priorización de zonas críticas de tres cuencas hidrográficas categorizadas con mayor riesgo de Cundinamarca para implementar estrategias de Soluciones Basadas en la Naturaleza frente al cambio climático?

Marco Teórico

La cuenca hidrográfica es un lugar donde se genera, almacena y evacua el agua, así es como lo definen Munar Samboní et al. (2022). En el caso de Colombia, se trata de la unidad a la que confluye todo el escorrente en un cuerpo de agua común (Banrepcultural, 2021). Sin embargo, una cuenca no sólo es un límite territorial, también es un sistema vivo que desempeña funciones ecosistémicas vitales, entre ellas se encuentra la regulación hidrológica, que tiene la capacidad de atemperar los picos de caudal y mantener los flujos ecológicos; la provisión de hábitats; el aumento de la biodiversidad y la capacidad para filtrar contaminantes. Convirtiendola en una fuente fundamental de suministro para el ser humano (Maldonado Santana et al., 2023; Munar Samboní et al., 2022; Vélez-Duque & Arteaga-Morales, 2022).

Es importante reconocer que este recurso natural está en peligro, debido a las variaciones significativas y persistentes en patrones climáticos, como temperatura, precipitación, distribución estacional, eventos extremos (IPCC, 2023; Instituto del Agua, 2024; UN-Water, 2023), impactando la calidad y cantidad del agua (Lane et al., 2023). Causando aumento continuo de la temperatura, cambios en el régimen de lluvias, mayor evaporación (IPCC, 2023; Instituto del Agua, 2024; UN-Water, 2023), problemas de sequías prolongadas, inundaciones frecuentes y disminuciones en capacidad de regulación de sistemas hídricos (Jeuken et al 2023).

Los impactos mas revelantes en la gestión de recursos hídricos están dados por inundaciones; aumento de la demanda y la evaporación; deterioro de la calidad del agua a causa de contaminantes durante las sequías o por arrastre de sedimentos y reducción de escorrentías en cuencas; estudios en Colombia apuntan a declives promedio de entre 12 % y 18 % en ciertos escenarios climáticos (Alarcón-Hincapié, Zafra-Mejía, & Echeverri-Prieto, 2019), adicionalmente están los retos: adopción de políticas y variabilidad; originando que algunas

cuencas sean más vulnerables que otras, señalando así medidas prioritarias (Vélez & Ocampo, 2012; Villamizar, Pineda, & Carrillo, 2019).

De allí que identificar las llamadas zonas críticas se convierta en una tarea estratégica. En palabras de Sáenz Segura (1995), se trata de áreas con un alto grado de deterioro o de amenaza para los recursos naturales y los servicios ecosistémicos; concentrando vulnerabilidades y, al mismo tiempo, oportunidades para la adaptación. En Colombia esas zonas suelen corresponder a subcuencas o regiones altoandinas que regulan el suministro hídrico, pendientes pronunciadas sometidas a erosión, áreas con cobertura vegetal fragmentada o deteriorada (Banco Interamericano de Desarrollo, Conservación Internacional Colombia, & Colombia, 2022; Corporación Autónoma Regional de las Cuencas de los Ríos Negro y Nare, 2023).

Un caso de zona crítica en la región de Cundinamarca por cambio en los patrones de precipitación es el territorio de alta montaña Chingaza-Sumapaz-Guerrero por su alta sensibilidad climática, cambios en la vegetación y cambio climático, este ecosistema es fundamental para la regulación hídrica y el abastecimiento de Bogotá y municipios cercanos, pero se ve amenazado por el derretimiento de grandes acumulaciones de agua atmosférica y la disminución de su capacidad reguladora (Banco Interamericano de Desarrollo, Conservación Internacional Colombia, & Colombia, 2022)

Determinar estas regiones en Cundinamarca cobra especial importancia según el Observatorio Ambiental de Bogotá (2024), ya que se priorizan las áreas vulnerables a sequías, aguaceros invisibles, inundaciones y olas de calor, porque estos fenómenos ejercen estrecha injerencia sobre la salud comunitaria, los recursos hídricos y la agricultura (Enel X, 2025). Por ello las áreas críticas se establecen por parámetros biológicos, físicos y antrópicos permitiendo concentrar recursos y garantizar la intervenciones de Soluciones Basadas en la Naturaleza de

manera más efectiva, mejorando la capacidad de adaptación socio ecológica (Observatorio Ambiental de Bogotá, 2024).

Es importante mencionar que la Gestión Integral del Recurso Hídrico (GIRH), siendo el marco operativo e institucional en Colombia, busca operacionalizar la gestión sostenible del agua ante los desafíos del cambio climático y la variabilidad hídrica, articulando principios ecosistémicos, participación ciudadana, evaluación ambiental, información científica, herramientas técnicas, normativas y planificación institucional, asegurando que la oferta, demanda, calidad, riesgos y vulnerabilidad del agua se consideren de forma integrada en las cuencas hidrográficas (IDEAM, 2013; CAR, 2020).

Algunos ejemplos de GIRH son estudios como la gestión integrada de los recursos hídricos como estrategia de adaptación al cambio climático realizado por Carvajal Escobar (2011), el cual plantea una estrategia clave para enfrentar: escasez de agua, cambios en la intensidad de precipitación, eventos extremos, degradación de calidad del agua, y variaciones en los procesos hidrológicos, otro caso es la Evaluación Regional del Agua (ERA) llevado a cabo por la Corporación Autónoma Regional de Cundinamarca (CAR), este analiza la oferta hídrica y la demanda, calidad y riesgos, incluyendo la amenaza que representan los escenarios climáticos, además sirve para zonificar hidrográficamente los territorios detectar subzonas hidrográficas críticas, y priorizar intervenciones municipales y regionales basadas en evidencia (Rodríguez et al., 2020).

Con lo anterior, se tiene la base para orientar la implementación de Soluciones Basadas en la Naturaleza (SbN) en las cuencas hidrográficas, ya que los datos de calidad del agua, oferta y demanda hídrica, riesgos y situaciones climáticas posibilitan que se determinen las áreas en las que se puede llevar a cabo este tipo de soluciones optimizando su eficacia para restablecer

procesos ecosistémicos, disminuir vulnerabilidades y elevar la capacidad de adaptación de las comunidades. (Chausson et al., 2020; Sun et al., 2024; Aghaloo et al., 2024). Según Chausson et al. (2020) describen las SbN como intervenciones que protegen, gestionan y restauran ecosistemas de forma sostenible para enfrentar difentes desafíos como el cambio climático.

La reforestación en zonas de recarga, la rehabilitación de humedales altoandinos, el manejo sostenible del suelo y la infraestructura verde son algunos ejemplos de SbN que benefician la regulación hídrica, la calidad del agua y reducen los peligros de sequías o inundaciones (Sun et al., 2024; Aghaloo et al., 2024) en cuencas como las de Cundinamarca. Adicionalmente estas acciones contribuyen a co-beneficios como el incremento de la biodiversidad, mejora de los paisajes, recolección de carbono y restauración de ecosistemas, robustesiendo la seguridad alimentaria como la hídrica de las comunidades urbanas y rurales (Kabisch et al., 2016).

La Corporación Autónoma Regional de Cundinamarca (CAR) ha estado llevando a cabo iniciativas en curso sobre Soluciones Basadas en la Naturaleza (SbN) para la recuperación de humedales y así lograr la regulación hídrica y la adaptación al cambio climático en doce municipios del departamento, algunas acciones que se llevaran a cabo es la revegetación con especies nativas, control de especies invasoras y fortalecimiento comunitario para la conservación (Periodismo Público, 2024), otro proyecto es en el municipio de Soacha el cual busca restaurar y adecuar varios humedales urbanos para recuperar la capacidad de almacenamiento y filtración de agua, mejorando la biodiversidad y reduciendo riesgos de inundación en áreas vulnerables (La Tribuna de las Regiones, 2025).

Los proyectos anteriormente mencionados, muestran como las SbN se perfilan como una alternativa efectiva para contra restar los problemas del cambio climático en zonas criticas de

cuencas hídricas, pero estos requieren contar con diagnósticos integrales y datos confiables para orientar la toma de decisiones, por esto diferentes estudios sugieren la integración de análisis multicriterio y técnicas de aprendizaje automático para maximizar así su eficacia y sostenibilidad a largo plazo, fortaleciendo así la gestión adaptativa del recurso hídrico frente a escenarios climáticos cada vez más inciertos (Villamizar, Pineda, & Carrillo, 2019; Munar et al., 2023).

Por esto, el aprendizaje supervisado ha emergido como una poderosa herramienta para enfrentar la complejidad de los sistemas sociales ecológicos, especialmente el estudio de cuencas hidrográficas. Dentro de este campo, algoritmos como Random Forest (RF) y Extreme Gradient Boosting (XGBoost) han ganado protagonismo por su capacidad de modelar relaciones no lineales y manejar grandes volúmenes de datos con múltiples variables. Aunque ambos se basan en árboles de decisión, sus fundamentos operativos difieren de manera significativa.

Random Forest (RF) genera varios árboles utilizando subconjuntos aleatorios de datos y variables, lo que contribuye a disminuir la varianza y aumentar la estabilidad del modelo (Breiman, 2001). Por otro lado, Extreme Gradient Boosting (XGBoost) pertenece a la categoría de métodos de boosting, donde los árboles se entrenan de manera secuencial para rectificar errores anteriores. Esta metodología le confiere una notable precisión, aunque también presenta una mayor sensibilidad en el ajuste de hiperparámetros (Chen & Guestrin, 2016).

Algunos estudios revelan que los algoritmos son tan aplicables para el análisis de recursos hídricos y el entorno ecológico como para uso tradicional. Por ejemplo, con RF para cartografiar los ecosistemas acuáticos en una cuenca del oeste colombiano, Según Martínez-Santos et al. (2021) señalan que se logró alcanzar métricas de rendimiento extraordinarias (AUC=0.95, Precision >93%). Sin embargo, el estudio también puso de manifiesto limitaciones en zonas montañosas, atribuidas a la resolución de los modelos digitales de elevación. Sin

embargo, Qiu, Niu, Baas, y Phanikumar (2023) estudiaron cómo uno podía usar tanto RF como XGBoost para predecir deslizamientos en la cuenca de La Liboriana (Antioquia), descubriendo que RF era superior a XGBoost en términos de Area Under Curve (AUC) y Accuracy, sugiriendo una mayor robustez en escenarios locales. No obstante, XGBoost mostró ventajas en el manejo de datos desbalanceados, especialmente al incorporar técnicas como de sobre-muestreo sintético de la minoría (SMOTE) y métricas complementarias como el F1-core y LogLoss, lo que refuerza su utilidad en contextos donde las clases minoritarias son críticas.

A nivel internacional, XGBoost ha demostrado un rendimiento superior en aplicaciones relacionadas con la calidad del agua (Karataş, Çakmakçı, Yücel, Demir, & Şen, 2026) informaron que este algoritmo logró mejores resultados que RF en la predicción de oxígeno disuelto en sistemas acuícolas ($R^2 = 0.78$ vs. 0.65). Resaltaron también la importancia de la temperatura y el pH como variables apropiadas según la hidrología. Tales descubrimientos sugieren que, si bien RF ofrece estabilidad y facilidad de interpretación, XGBoost puede ser más efectivo para escenarios donde las variables están altamente correlacionadas. Sin embargo, implementar estos modelos en entornos espaciales genera importantes desafíos metodológicos. Un desafío es la autocorrelación espacial, que puede inflar artificialmente las métricas de rendimiento si se emplean validaciones convencionales. Frente a ello, Wang, Khodadadzadeh y Zurita-Milla (2023) propusieron la técnica Spatial+ Cross Validation, que organiza los datos en bloques espaciales para obtener estimaciones más realistas de la capacidad de generalización, especialmente relevante en cuencas andinas con fuertes gradientes topográficos y climáticos.

Otro reto común es el desbalance de clases, que se muestra sobre todo a la hora de identificar zonas críticas alejadas. El estudio de Qiu, Niu, Baas, y Phanikumar (2023) en La Liboriana demostró cómo la utilización de SMOTE sirvió para mejorar la sensibilidad del

modelo a los eventos raros, haciendo de esta forma la detección de deslizamientos más democrática. En estos casos, el uso de métricas ajustadas como AUC-PR y F1-score resulta indispensable para valorar el rendimiento en términos más justos y representativos.

En definitiva, la interpretabilidad de los modelos es un aspecto clave, especialmente cuando se busca que los resultados informen decisiones ambientales. Aunque las métricas globales de variables (FI) identifican el recocimiento de predictores significativos, estos no detallan su impacto en cada predicción individual, por lo que herramientas como SHapley Additive exPlanations (SHAP) son fundamentales para superar esta limitante porque asigna valores de contribución específicos por cada instancia, los autores Qiu, Niu, Baas, y Phanikumar (2023) emplearon esta herramienta para evidenciar la relevancia de la elevación y el Índice de Vegetación de Diferencia Normalizada (NDVI) en los deslizamientos, mientras que Karataş et al. (2026) enfatizaron la necesidad de graficas con dependencia parcial para complementar las interacciones entre variables y así tener una mejor comprensión.

La revisión de la bibliografía indica que la utilización de RF y XGBoost para gestionar cuencas hidrográficas se basa en tres pilares esenciales: una validación espacial exhaustiva, un manejo apropiado del desbalance entre clases y el uso de instrumentos de interpretabilidad multiescala. En estas circunstancias, los dos algoritmos no solo son técnicamente sólidos, sino que además brindan legitimidad en la toma de decisiones ambientales, lo cual los convierte en aliados estratégicos para priorizar Soluciones Basadas en la Naturaleza en regiones como Cundinamarca.

Sin embargo, además de estas perspectivas fundamentadas en Machine Learning, los modelos descriptivos y predictivos en las ciencias ambientales se han vuelto instrumentos esenciales para entender con anticipación las transformaciones en los sistemas socioecológicos e

hídricos. El análisis de cuencas hidrográficas se ha beneficiado en gran medida del uso de modelos hidrológicos tradicionales, como WEAP (Water Evaluation and Planning System) y SWAT (Soil and Water Assessment Tool), que posibilitan la simulación del balance hídrico, la proyección de escenarios de cambio climático y la evaluación de los efectos sobre la disponibilidad del recurso (Arnold et al., 2012; Yates, Sieber, Purkey & Huber-Lee, 2005).

Los anteriores enfoques, aunque son robustos en términos de procesos biofísicos, presentan limitaciones en la incorporación de dinámicas socioeconómicas y en la adaptabilidad a regiones con información insuficiente o heterogénea. En este contexto, el aprendizaje automático ha cobrado relevancia, pues algoritmos de RF y XGBoost permiten integrar datos multivariados de distinta naturaleza, desde series temporales climáticas hasta indicadores de presión antrópica, con el fin de identificar patrones complejos y priorizar áreas críticas (Breiman, 2001; Chen & Guestrin, 2016).

Un ejemplo de aplicación en América Latina es el modelado de la susceptibilidad a deslizamientos en Antioquia y la previsión de fenómenos torrenciales en La quebrada La Liboriana. Estos casos demuestran que los modelos de ML pueden superar la capacidad predictiva de los métodos tradicionales, aunque su eficacia depende en gran medida de la calidad y resolución de los datos disponibles (Gómez-Miranda et al., 2025; Rodríguez et al., 2020).

El aspecto de mayor relevancia que hace referencia a la proximidad entre los valores observados en el terreno son los estimados, y la precisión, es el grado de consistencia en las predicciones, son dos factores esenciales para evaluar cuán eficaces son los modelos. En el análisis del medio ambiente, estos dos elementos usualmente dependen de la calidad de las bases de datos, la existencia de series históricas completas y la resolución temporal y espacial de los datos (Willmott & Matsuura, 2005). Un ejemplo de esto, se presenta cuando los registros

climáticos que no están completos pueden hacer que la precisión de las proyecciones se vea afectada, en tanto que las bases de datos satelitales con baja resolución pueden disminuir la habilidad de un modelo para detectar microcuencas importantes. Históricamente, se han utilizado como indicadores estándar el Error cuadrático medio (RMSE) y el coeficiente de determinación (R^2). (Hijmans, 2012). No obstante, autores como Molnar (2022) subrayan la necesidad de complementar estas métricas con herramientas de interpretabilidad, como SHAP (SHapley Additive exPlanations), que permiten comprender cómo cada variable influye en la predicción y, en consecuencia, mejorar la confianza en la aplicabilidad del modelo para la gestión ambiental.

Finalmente, la aplicabilidad territorial juega un rol crucial para determinar la efectividad de los modelos, dado que la gestión de cuencas hidrográficas exige herramientas que no solo sean estadísticamente precisas, sino también adaptables y escalables a diferentes contextos socio ecológicos. En Colombia, las condiciones del clima, la diversidad cultural y la variabilidad en el acceso a datos ambientales imponen retos significativos a la transferencia de modelos entre regiones (IDEAM, 2021). No obstante, casos como el sistema de alerta temprana de inundaciones en la cuenca Magdalena-Cauca (IDEAM, 2021), el diseño de planes de conservación hídrica en Perú y Chile basados en modelación espacial, o la priorización de áreas de restauración en Brasil mediante ML, demuestran que es posible articular modelos predictivos con estrategias de gestión territorial (Del Aguila Ríos, 2024; Cantillo Romero et al., 2023).

Estos ejemplos muestran que la efectividad no solo está determinada por la capacidad técnica del modelo, sino también por su adaptación a las situaciones sociales, institucionales y medioambientales de cada región. Por lo tanto, un modelo realmente eficaz tiene que balancear la precisión, la interpretabilidad y la relevancia territorial, de tal forma que vaya más allá del

campo académico y se transforme en una herramienta de apoyo auténtico para la planificación pública, comunitaria y ambiental.

Marco Institucional

El desarrollo de la investigación se da en un contexto institucional y se toman como referencia tres cuencas categorizadas con alto riesgo del departamento de Cundinamarca, siendo una región del centro del país relevante debido a su papel estratégico en el aprovisionamiento de servicios ecosistémicos, recursos hídricos y biodiversidad; la administración de estos está bajo entidades ambientales como la Corporación Autónoma Regional de Cundinamarca (CAR), el Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) y la Secretaría del Ambiente de Cundinamarca (MinAmbiente, 2023; IDEAM, 2022). De acuerdo con el DANE (2020), estas organizaciones deben preservar y utilizar los recursos naturales de manera sostenible.

El proyecto titulado Modelo de Machine Learning para la priorización de zonas críticas de tres cuencas hidrográficas de Cundinamarca mediante Soluciones Basadas en la Naturaleza frente al cambio climático, contempla y se basa en las políticas institucionales de adaptación y mitigación del cambio climático lideradas por el IDEAM, ya que esta entidad proporcionan datos ambientales junto a METEOSTAT se obtienen datos meteorológicos y climáticos necesarios para alimentar el modelo propuesto integrando información sobre temperatura, precipitación, caudal, velocidad del viento y presión del aire (IDEAM, 2022; MinAmbiente, 2021).

El modelo propuesto aplica algoritmos de clasificación supervisada (Random Forest y XGBoost) para clasificar las zonas más vulnerables al cambio climático, además cuenta con un enfoque que se alinea a los objetivos de la planificación ambiental para facilitar la toma de decisiones y la implementación de Soluciones Basadas en la Naturaleza (SbN), como restauración ecológica, conservación de humedales y manejo sostenible del territorio (World Wildlife Fund, s.f.; MinAmbiente, 2022).

Finalmente, el proyecto articula el sector tecnológico, ambiental y académico con labores de entidades públicas, permitiendo innovar en la gestión ambiental mediante ciencia de datos y avance digital, proyectando una iniciativa que apoya las instituciones de Cundinamarca, ya que les brinda un panorama más amplio clasificando zonas críticas con prioridad para su respectiva intervención con Soluciones basadas en la Naturaleza (SbN) y así, fomentar la gobernanza territorial sostenible frente al cambio climático (PNUD, 2022; MinAmbiente, 2023).

Metodología

Enfoque, alcance y diseño de la investigación

El estudio se fundamenta en la recopilación, procesamiento y análisis estadístico de datos ambientales, climáticos y espaciales, por lo que conlleva un desarrollo con enfoque cuantitativo; es importante mencionar que los datos se toman de fuentes oficiales como el Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) y METEOSTAT, que es una plataforma que proporciona datos meteorológicos y climáticos histórico. De esta forma los fenómenos ambientales se pueden representar con variables medibles y objetivas facilitando el uso de algoritmos de aprendizaje automático que operan sobre grandes volúmenes de información (Hernández-Sampieri & Mendoza Torres, 2018).

El propósito es determinar patrones y correlaciones espaciales entre los factores para identificar vulnerabilidades ambientales en las tres cuencas hidrográficas de Cundinamarca, logrando clasificar zonas críticas de intervención para proponer y brindar Soluciones basadas en la Naturaleza (SbN) mitigando los efectos del cambio climático, por lo que la metodología cuantitativa es ideal para obtener resultados reproducibles y generalizables, lo cual fortalece la toma de decisiones en la gestión ambiental territorial (PNUD, 2022).

El alcance de la investigación es descriptivo, correlacional y aplicado:

Es descriptivo, ya que busca representar y clasificar el comportamiento espacial y temporal de variables ambientales como la precipitación, la temperatura y la presión antrópica dentro de la cuenca hidrográfica (IDEAM, 2022).

Es correlacional porque analiza las relaciones entre variables y su contribución a la vulnerabilidad territorial frente a los efectos del cambio climático y la clasificación de las zonas críticas, identificando interacciones entre factores ecológicos (MinAmbiente, 2021).

Es aplicado, debido a el resultado esperado en el modelo mediante la integración de herramientas de Machine Learning (Random Forest, XGBoost) para orientar la toma de decisiones sobre las Soluciones basadas en la Naturaleza (SbN) para la gestión ambiental sostenible.

El diseño de investigación adoptado es no experimental y transversal:

Es no experimental, ya que las variables no se manipulan solo se analizan tal como ocurren en el entorno natural, los datos se recolectan de fuentes confiables y bases históricas de monitoreo ambiental, permitiendo un análisis objetivo sin alterar las dinámicas de los sistemas ambientales (Hernández-Sampieri & Mendoza Torres, 2018).

Es transversal, porque el análisis se realiza en un periodo específico, integrando series de tiempo y capas de información geográfica disponibles, para establecer un diagnóstico del estado actual de las cuencas frente a la vulnerabilidad del cambio climático (MinAmbiente, 2023).

Con lo anterior, el diseño metodológico combina el análisis cuantitativo y espacial con la modelación predictiva, para fortalecer la comprensión de las dinámicas ambientales y sociales desde una perspectiva sistémica. Al tener este enfoque los resultados pueden ser replicables, comparables y útiles para la gestión pública, en concordancia con los principios de sostenibilidad ambiental, innovación tecnológica y planeación territorial promovidos por el Gobierno Nacional (MinAmbiente, 2022).

Definición de Variables

En este estudio, las variables se estructuran a partir de las dimensiones ambiental, climática y espaciales que determinan la vulnerabilidad y criticidad de las tres cuencas hidrográficas categorizadas con alto riesgo de Cundinamarca frente al cambio climático. La definición conceptual y operacional de las variables es esencial para garantizar la coherencia

metodológica y la reproducibilidad del modelo de análisis propuesto (Hernández-Sampieri & Mendoza Torres, 2018)

El estudio tiene un carácter correlacional, lo que implica que las variables se analizan en función de sus interdependencias más que de una relación causal directa. De esta forma, se busca identificar los factores que más inciden en la criticidad ambiental mediante la aplicación de algoritmos de clasificación supervisada (Random Forest y XGBoost), integrando datos espaciales, climáticos y demográficos.

Los datos utilizados provienen de fuentes oficiales como el Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) y METEOSTAT que es la plataforma de datos meteorológicos y climáticos históricos disponibles en portales abiertos (IDEAM, 2023).

Tabla 1

Definición conceptual y operacional de las variables del estudio

Variable	Definición conceptual	Definición operacional	Dimensión
Fecha	Momento temporal en el que se realiza el registro o medición de la variable climática o ambiental.	Año, mes y día en formato AAAA-MM-DD correspondiente al dato recolectado.	Temporal
Municipio	División político-administrativa del departamento de Cundinamarca.	Nombre del municipio donde se registran los datos climáticos y ambientales.	Territorial
Cuenca	Unidad hidrológica que agrupa el conjunto de aguas que drenan hacia un mismo río principal.	Nombre de la cuenca o subcuenca hidrográfica asociada al municipio (ej. Río Bogotá, Río Sumapaz).	Hidrográfica
Latitud	Coordenada geográfica que indica la posición norte o sur de un punto sobre la Tierra.	Valor numérico en grados decimales de la latitud del municipio o estación meteorológica.	Espacial
Longitud	Coordenada geográfica que indica la posición este	Valor numérico en grados decimales de la longitud del	Espacial

Variable	Definición conceptual	Definición operacional	Dimensión
	u oeste de un punto sobre la Tierra.	municipio o estación meteorológica.	
Altitud	Altura de un punto con respecto al nivel del mar.	Metros sobre el nivel del mar (m.s.n.m.) del punto de medición.	Topográfica
Temp Promedio	Promedio anual de la temperatura del aire en la superficie terrestre.	Media aritmética de las temperaturas registradas en °C durante un año.	Climática
Temp Min	Temperatura más baja registrada en un periodo determinado.	Valor mínimo de temperatura (°C) registrado en el municipio o estación durante el año.	Climática
Temp Max	Temperatura más alta registrada en un periodo determinado.	Valor máximo de temperatura (°C) registrado en el municipio o estación durante el año.	Climática
Precipitación	Cantidad total de agua caída en forma de lluvia, nieve o granizo en un periodo.	Milímetros (mm) acumulados de precipitación anual según registros del IDEAM.	Hidrometeorológica
Velocidad del Viento	Velocidad media del viento en un periodo determinado.	Promedio anual de velocidad del viento (m/s) registrada por las estaciones meteorológicas.	Climática
Caudal Máximo	Mayor volumen de agua que pasa por una sección de un río en un periodo determinado.	Valor máximo de caudal (m ³ /s) registrado en los puntos de control de la cuenca.	Hidrológica
Caudal Mínimo	Menor volumen de agua que pasa por una sección de un río en un periodo determinado.	Valor mínimo de caudal (m ³ /s) registrado en los puntos de control de la cuenca.	Hidrológica
Presión del Aire	Fuerza ejercida por la atmósfera sobre la superficie terrestre.	Promedio anual de presión atmosférica (hPa) medida en la estación meteorológica.	Climática

Nota. Adaptado de “Consulta y Descarga de datos Hidrológicos” por e IDEAM (<http://dhime.ideam.gov.co/atencionciudadano/>); y “El Guardián del Tiempo” por METEOSTAT (<https://meteostat.net/es/>).

La definición operacional de estas variables permite la estructuración de un modelo predictivo espacialmente explícito, en el que cada capa de información representa una dimensión del sistema ambiental y social. El procesamiento de datos se realizará mediante técnicas de análisis geoespacial, normalización de variables y entrenamiento de modelos de Machine Learning, asegurando coherencia con las políticas ambientales del departamento (MinAmbiente, 2023; Región Metropolitana, 2024).

Este enfoque favorece la identificación de zonas prioritarias de intervención con base en la evidencia, facilitando la planificación territorial y la implementación de Soluciones Basadas en la Naturaleza (SbN) en las tres cuencas seleccionadas con más vulnerabilidad de Cundinamarca (MinAmbiente, 2022).

Población y Muestra

La población del estudio corresponde a los sistemas hidrográficos del departamento de Cundinamarca, conformada por los sistemas hidrográficos que lo integran; los cuales suman importancia a nivel social, económico y ecológico, ya que proveen recursos hídricos a una gran parte de la población y sostienen labores domésticas, industriales y agropecuarias.

La muestra se centra en tres cuencas específicas del departamento:

- Girardot (Caudal La Campiña)
- Puerto Salgar (Caudal Puerto Salgar)
- Guaduas (Caudal Guadero)

Estas cuencas fueron seleccionadas con base en el ranking de la Presentación Diagnóstico Ambiental Cundinamarca, la cual identifica los municipios con mayor riesgo climático. Además, se seleccionaron por la información disponible y confiable en las bases de datos del IDEAM y Meteostat, lo que permite analizar el comportamiento de las variables previamente clasificadas

para la construcción base del dataset y así tener una mayor precisión (Gobernación de Cundinamarca, 2024).

Selección de métodos o instrumentos para recolección de información

La elección de instrumentos y métodos para la recopilación de datos se basa en el enfoque correlacional y cuantitativo del estudio, que tiene como propósito procesar y examinar datos multivariados con el fin de crear un modelo predictivo de vulnerabilidad medioambiental a través de algoritmos de Machine Learning (XGBoost y Random Forest).

El procedimiento metodológico se basa en tres fases primordiales:

1. Identificación y clasificación de fuentes oficiales de información que aseguren la vigencia, la validez y la concordancia espacial.
2. Adquisición, fusión y normalización de variables espaciales, climáticas y ambientales.
3. La formación, validación y evaluación de modelos que predicen, utilizando herramientas de interpretabilidad (SHAP) y métricas estadísticas (RMSE, MAE, R^2).

Esta perspectiva garantiza la capacidad de rastrear los datos y reproducir los resultados, en línea con las normas de transparencia científica y gestión ambiental fundamentadas en evidencia.

Las fuentes de información utilizadas son secundarias, oficiales y públicas, lo que asegura la relevancia institucional y la fiabilidad. Estas fuentes incluyen:

- Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM): datos históricos sobre las precipitaciones, la temperatura, el caudal, longitud, latitud y altitud.
- Departamento Administrativo Nacional de Estadística (DANE)
- La plataforma de Meteostat para datos meteorológicos y climáticos históricos.
- Documentos de política ambiental, planes para adaptarse al cambio climático y estrategias SbN: MinAmbiente y PNUD

Estos recursos suministran datos organizados y de múltiples escalas, que son aptos para ser procesados a través de Sistemas de Información Geográfica (SIG) y analizados computacionalmente.

Técnicas de análisis de datos y Muestra

El análisis de datos dentro de este estudio se desarrolla a través de tres etapas interconectadas, diseñadas para pasar de un examen inicial de variables a la creación de productos geoespaciales que puedan informar la toma de decisiones institucionales. Este enfoque metodológico garantiza un fuerte vínculo entre los datos recopilados, los algoritmos de aprendizaje automático empleados y los resultados anticipados relacionados con la priorización territorial. A continuación, se detallan las técnicas específicas utilizadas en cada etapa:

Fase 1. Análisis Descriptivo y Exploratorio

Esta etapa inicial implica una caracterización estadística y espacial de las variables ambientales, climáticas y espaciales recolectadas para las tres cuencas hidrográficas de Cundinamarca. El objetivo principal es descubrir patrones de distribución, identificar valores atípicos y evaluar la calidad de los datos antes de integrarlos en modelos predictivos. Se calcularán estadísticas básicas para cada variable, incluidas medidas de tendencia central (media, mediana), medidas de dispersión (desviación estándar, rango intercuartílico) y análisis de distribución de frecuencia. Estas métricas ayudarán a definir las características de cada variable y a identificar posibles inconsistencias dentro de los datos recopilados (Willmott y Matsuura, 2005).

Tras el análisis espacial, se crearán mapas temáticos utilizando Sistemas de Información Geográfica (SIG). Estos mapas ilustrarán la distribución espacial de cada variable, permitiendo visualizar la heterogeneidad territorial y la identificación de áreas con factores críticos

concentrados. Esta representación cartográfica ayudará a interpretar patrones espaciales cruciales para priorizar zonas críticas (Hijmans, 2012). Además, se realizará un análisis de correlación utilizando el coeficiente de correlación de Pearson y el factor de inflación de la varianza (VIF). El análisis de correlación identificará relaciones lineales entre las variables predictoras, mientras que el VIF evaluará posibles problemas de multicolinealidad que podrían afectar el rendimiento de los modelos de aprendizaje automático. Las variables que exhiban un VIF superior a 10 se considerarán para su exclusión o transformación para mitigar la redundancia y mejorar la solidez del modelo (Chen y Guestrin, 2016; Breiman, 2001).

Fase 2. Modelado predictivo mediante algoritmos de clasificación supervisada

Después de la fase exploratoria, el enfoque cambia al entrenamiento de modelos de clasificación supervisados diseñados para evaluar la probabilidad de criticidad dentro de cada unidad espacial. Esto implicará la implementación de dos algoritmos de aprendizaje automático ampliamente reconocidos por su eficacia en aplicaciones ambientales: Random Forest (RF) y XGBoost (Extreme Gradient Boosting). Random Forest se elige por su capacidad para manejar relaciones no lineales entre variables y su resistencia a conjuntos de datos desequilibrados, ambos aspectos cruciales de los estudios de susceptibilidad ambiental (Breiman, 2001). XGBoost, por otro lado, se selecciona por su alta precisión de clasificación y su enfoque de aprendizaje secuencial, que permite la corrección de errores de iteraciones anteriores, optimizando así el rendimiento predictivo (Chen y Guestrin, 2016). El entrenamiento de ambos modelos utilizará variables ambientales (cobertura vegetal, índices de vegetación, precipitación, temperatura) y socioeconómicas (densidad poblacional, índices de pobreza multidimensional, actividades productivas) como predictores. La variable de respuesta será la clasificación de zonas críticas, determinada por criterios técnicos basados en la vulnerabilidad hídrica, el riesgo de

degradación y la presión antropogénica. Para garantizar la validez externa de los modelos y mitigar los problemas de autocorrelación espacial, se empleará una estrategia de validación cruzada espacial utilizando k-fold por subcuenca (Wang et al., 2023). Esta técnica divide el área de estudio en bloques espaciales independientes, evitando la superposición geográfica entre los datos de entrenamiento y validación. Este enfoque produce estimaciones más confiables del rendimiento del modelo en contextos invisibles.

Los específicos modelos de ellas clases se evaluarán con métricas de rendimiento clásicas en el material científico: precisión (accuracy), recuperación (recall), F1-scores (puntuación F1) y área debajo de la curva ROC (AUC-ROC). La precisión muestra la proporción de predicciones correctas en total observaciones de conjunto; Sensibilidad the measures especialmente al modelo capacidad de identificar en qué lugares zonas críticas; el F1-score es un equilibrio práctico entre precisión y sensibilidad especialmente útil para el caso cláusulas de desequilibrio; y AUC-ROC evaluación discriminante del modelo a través diferentes umbrales (Qiu et al., 2023; Karataş et al., 2026). Por otra parte, se aplicará la técnica de interpretación SHAP (SHapley Additive exPlanations) para conocer cuál es la aportación individual de cada uno de los predictores en las clasificaciones individuales, lo que permite entender qué factores ambientales y socioeconómicos tienen más influencia en la definición de zonas críticas (Molnar, 2022). Este conocimiento es esencial para apoyar determinadas medidas y debería hacer el modelo más aceptado por las autoridades.

Fase 3. Priorización y generación de mapas

En la fase final, los resultados del modelo predictivo se fusionan con las evaluaciones de vulnerabilidad socioeconómica para crear un índice de priorización territorial. Este índice sintetiza la probabilidad de criticidad, determinada por algoritmos de Machine Learning, con

indicadores sociales (vulnerabilidad de la comunidad, acceso a servicios esenciales, dependencia de los recursos hídricos). Esta integración facilita una clasificación multicriterio de las áreas que requieren intervención a través de Soluciones Basadas en la Naturaleza (SbN). El índice de prioridad se calculará utilizando un enfoque ponderado, con la probabilidad de criticidad ambiental representando el 60% y el índice de vulnerabilidad socioeconómica el 40%. Esta ponderación refleja la comprensión de que la capacidad de adaptación de la comunidad influye significativamente en la efectividad de las SbN, una perspectiva respaldada por la investigación de Chausson et al. (2020) y Sun et al. (2024). Según el índice calculado, las áreas de estudio se clasificarán en tres niveles de prioridad: Alta, Media y Baja. Las zonas de alta prioridad serán aquellas que presenten una alta probabilidad de criticidad ambiental y una vulnerabilidad socioeconómica significativa. Las zonas de prioridad media abarcarán áreas con condiciones intermedias en al menos uno de los componentes, mientras que las zonas de prioridad baja representarán áreas con una menor urgencia de intervención según los criterios definidos.

Finalmente, se elaborarán mapas geoespaciales en formatos estandarizados como GeoTIFF y shapefile, facilitando la incorporación de los resultados en plataformas institucionales de gestión territorial, incluyendo los sistemas de información de las Corporaciones Autónomas Regionales. Estos productos cartográficos vendrán acompañados de metadatos descriptivos que detallarán la fuente de los datos, el sistema de coordenadas utilizado, la resolución espacial y las limitaciones metodológicas, asegurando así su trazabilidad y utilidad en los procesos de planificación ambiental (Rodríguez et al., 2020).

Tabla 2

Síntesis de instrumentos y técnicas de análisis

Instrumento / Técnica	Técnica de análisis	Descripción / Aplicación
Capas geoespaciales	Análisis descriptivo y espacial	Generación de mapas temáticos y análisis de correlaciones espaciales para identificar patrones territoriales.
Base de datos integrada	Análisis multivariado	Normalización de variables, evaluación de correlaciones mediante Pearson y reducción de redundancia con VIF.
Modelos de ML (RF, XGBoost)	Clasificación supervisada	Determinación de zonas críticas según probabilidad de criticidad ambiental y climática.
SHAP values	Interpretabilidad del modelo	Identificación de las variables predictoras con mayor influencia en las clasificaciones individuales.
Índice de priorización	Integración multicriterio	Combinación de resultados del modelo predictivo con factores de vulnerabilidad socioeconómica.

Nota. Adaptado de “ Mapping the effectiveness of nature-based solutions for climate change adaptation” por Chausson et al. (2020) (<https://doi.org/10.1111/gcb.15310>); “Evaluación Regional del Agua ERA. Cuenca media del río Bogotá” por Rodríguez et al. (2020) (<https://sie.car.gov.co/items/0d84f334-c981-40c8-9357-d7ec5a336a81>) y “Mechanisms and applications of nature-based solutions for stormwater control in the context of climate change: A review.” por Sun et al. (2024) (<https://doi.org/10.3390/atmos15040403>).

La implementación de estas técnicas de análisis garantiza un enfoque riguroso, transparente y replicable, alineado con los estándares de calidad científica exigidos en investigaciones aplicadas al campo ambiental y de gestión territorial.

Análisis y discusión de resultados

Se exponen los resultados obtenidos durante la recopilación, depuración y exploración de bases de datos oficiales del Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM), y la plataforma Meteostat; donde se identificaron variables relevantes que inciden para la clasificación de zonas de priorización críticas frente al cambio climático en las tres cuencas hidrográficas de Cundinamarca.

El proyecto se planteó con el objetivo de desarrollar un modelo de aprendizaje automático pero el alcance de la investigación se limitó a la fase de análisis exploratorio de datos, la cual permite establecer una base conceptual y esencial para la construcción del modelo de clasificación supervisado.

Cada una de las bases de datos seleccionadas aportan dimensiones complementarias: el IDEAM provee variables ambientales y Meteostat brinda información de datos meteorológicos que proporciona información histórica y actual.

Variables por cada base de datos

Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM)

Se obtuvieron los datos oficiales correspondientes a las cuencas seleccionadas para el estudio, los cuales constituyen la base principal para el análisis hidrológico. Las variables consideradas fueron: fecha, municipio, cuenca, latitud, longitud, altitud, caudal mínimo y caudal máximo, que permiten caracterizar el comportamiento del recurso hídrico en cada zona y establecer relaciones con las condiciones climáticas del territorio.

Meteostat plataforma de datos meteorológicos y climáticos históricos

De esta fuente se obtuvieron las variables temperatura promedio, temperatura mínima, temperatura máxima, precipitación, velocidad del viento y presión del aire, las cuales permiten

complementar el análisis hidrometeorológico y evaluar las condiciones climáticas que inciden en el comportamiento de las cuencas seleccionadas.

Datos recopilados de las bases

A continuación se listan los enlaces de los archivos Excel con los datos recopilados por cada cuenca; los archivos se deben procesar en el modelo para realizar la unión de datos junto a la normalización de las fechas.

- Puerto Salgar: [Caudal Máximo y mínimo Puerto Salgar.xlsx](#) y [VARIABLES HISTÓRICAS Puerto Salgar.xlsx](#)
- Guaduas: [Caudal Máximo y mínimo Guaduas.xlsx](#) y [VARIABLES HISTÓRICAS Guaduas.xlsx](#)
- Girardot: [Caudal Máximo y mínimo La Campiña Girardot.csv](#) y [VARIABLES HISTÓRICAS Girardot.xlsx](#)

Calculos y análisis sobre las variables

Para determinar la mayor vulnerabilidad climática en las cuencas de Girardot, Puerto Salgar y Guaduas, se procede a realizar cálculos y comprender el comportamiento de cada variable para así combinarlas y poder brindar Soluciones basadas en la Naturaleza (SbN) que mitiguen riesgos fortaleciendo la resiliencia del ecosistema y las comunidades.

A continuación, se presenta una síntesis de los principales indicadores calculados para las variables seleccionadas:

Tabla 3

Análisis descriptivo y comportamiento de cada variable

Tipo de análisis	Cálculo o indicador	Qué te muestra
Promedios y extremos	Media, máximo, mínimo y desviación estándar	Variabilidad climática entre cuencas

Tipo de análisis	Cálculo o indicador	Qué te muestra
Rango térmico	Temp Max – Temp Min	Amplitud térmica → estrés térmico
Coefficiente de variación de precipitación	(Desv. estándar / Promedio) × 100	Irregularidad de lluvias (sequías o lluvias extremas)
Relación entre caudal y precipitación	Correlación o regresión lineal	Capacidad de retención y escorrentía de la cuenca
Índice de aridez	Precipitación / Temp Promedio	Zonas más secas o con riesgo de desertificación
Índice de caudal específico	Caudal promedio / área de cuenca	Disponibilidad hídrica relativa
Tendencia temporal (si tienes varios años)	Pendiente o test de Mann-Kendall	Si el clima muestra tendencia a más calor, menos lluvia, etc.

Nota. Adaptado de metodologías aplicadas en estudios hidrometeorológicos recientes para caracterización climática y vulnerabilidad hídrica (Vicente-Serrano et al., 2020; Sun et al., 2021; Abhishek et al., 2022; Wang et al., 2020; Kurnik et al., 2024).

Resultados esperados del análisis exploratorio

En el proceso de creación del modelo junto a los primeros pasos se espera un análisis explorativo EDA y normalización de los datos, luego realizar el análisis descriptivo de la tabla anterior para así evidenciar comportamientos y contrastantes entre las tres cuencas seleccionadas.

Girardot, ubicada en una zona de transición climática cálida-seca, se prevé que presente altas temperaturas promedio y bajas precipitaciones, generando condiciones de estrés hídrico y térmico, especialmente durante los meses de menor precipitación, situación que coincide con los patrones proyectados por el IDEAM para el centro del país (IDEAM, 2022; IPCC, 2021).

Guaduas, con altitudes intermedias y una topografía variable, se anticipa como una cuenca de variabilidad climática media, donde la presencia de eventos de precipitación intensa y caudales fluctuantes puede estar asociada a la influencia orográfica y a los fenómenos de variabilidad interanual como El Niño y La Niña (Martínez & Poveda, 2019).

Puerto Salgar, localizada en la zona baja del valle del Magdalena, se caracteriza por altas temperaturas y una marcada estacionalidad en las lluvias, lo que puede generar periodos alternados de sequía e inundación, tal como lo reportan las evaluaciones recientes de riesgo hídrico y cambio climático para el Magdalena medio (IDEAM, 2024, p. 2)

Construcción del Índice de Vulnerabilidad Climática

Luego de comprender las variables, realizar la normalización y la exploración EDA, se procede a construir un Índice de Vulnerabilidad Climática (IVC); el cual se plantea como una herramienta integradora que combina las principales variables ambientales y meteorológicas, ponderadas según su incidencia en la estabilidad climática de las cuencas. Este índice no solo permitirá identificar las zonas más críticas, sino también priorizar intervenciones con Soluciones basadas en la Naturaleza (SbN).

Estos patrones permitirán definir zonas prioritarias para la aplicación de estrategias de adaptación y mitigación, integrando tanto factores climáticos como hidrológicos. El IVC se expresa mediante la fórmula:

Ilustración 1

Formula IVC

$$IVC = w_1T + w_2P + w_3Q + w_4V + w_5Alt$$

Nota. El índice de criticidad climática se construyó mediante una integración ponderada de variables hidrometeorológicas (T, P, Q, V, Alt), siguiendo la metodología de índices compuestos

propuesta por la OCDE (2008) y adaptada a estudios hidrológicos en cuencas andinas (Fernández & Londoño, 2018; Gómez & Pineda, 2020).

Donde:

- T = variables de temperatura (Temp Prom, Min, Max)
- P = precipitación
- Q = caudales
- V = velocidad del viento
- Alt = altitud
- w_i = peso relativo

Los pesos relativos (w_i) fueron definidos a partir de la literatura científica y el análisis de expertos, priorizando la precipitación (0.30) por su impacto directo en la disponibilidad hídrica, seguida por la temperatura y el caudal (0.25 cada uno), y finalmente por el viento y la altitud (0.10 cada uno), cuya influencia es más indirecta.

Finalmente se espera que la cuenca con el valor más alto del índice será la más crítica frente al cambio climático.

Resultados esperados del IVC

Se espera que el cálculo del IVC arroje una jerarquización clara de vulnerabilidad climática entre las tres cuencas:

- Girardot :Alto índice de vulnerabilidad (por déficit hídrico y térmico).
- Puerto Salgar: Vulnerabilidad media-alta (por estacionalidad extrema).
- Guaduas: Vulnerabilidad media (por condiciones más equilibradas y altitud moderada).

Esta clasificación permitirá orientar acciones estratégicas hacia la resiliencia ecosistémica y el fortalecimiento de las capacidades adaptativas locales, en concordancia con los enfoques propuestos por el IPCC (2021) y adaptaciones metodológicas empleadas en cuencas colombianas

(Fernández & Londoño, 2018; Gómez & Pineda, 2020). La aplicación del IVC facilita identificar zonas críticas frente a escenarios de cambio climático y establecer prioridades en la gestión integrada del recurso hídrico (Sullivan & Meigh, 2007).

Discusión de resultados

El análisis exploratorio demuestra que las cuencas presentan comportamientos hidrometeorológicos diferenciados, resultado de sus condiciones geográficas y climáticas. El gradiente altitudinal influye directamente en la distribución de temperatura y precipitación, evidenciando una correlación inversa entre ambas variables.

La temperatura promedio anual muestra una tendencia ascendente en los últimos registros, consistente con los patrones regionales de cambio climático reportados por el IDEAM. De esta manera, la precipitación presenta una variabilidad interanual significativa, con picos de lluvia concentrados en pocos meses, lo cual genera desequilibrios en la recarga de acuíferos y mayor exposición a inundaciones y sequías.

Los caudales máximos y mínimos, al ser analizados conjuntamente con la precipitación, reflejan la capacidad de las cuencas para absorber o evacuar el exceso de agua. En zonas donde la cobertura vegetal ha disminuido, se evidencia un aumento en la escorrentía superficial y una reducción en la infiltración, agravando los efectos de las lluvias intensas.

El coeficiente de variación de la precipitación, superior al 30% en algunos sectores, sugiere alta irregularidad pluviométrica, factor determinante en la planificación de recursos hídricos. En contraste, las cuencas con cobertura vegetal conservada tienden a mantener caudales más estables, lo que valida la importancia de los ecosistemas como reguladores naturales del ciclo del agua.

El análisis de la velocidad del viento indica que las zonas más abiertas o deforestadas presentan valores más altos, lo que incrementa la erosión eólica y la pérdida de suelo fértil. Este resultado refuerza la necesidad de implementar barreras vivas o cortinas rompevientos como medidas de mitigación.

La altitud actúa como un moderador térmico: las zonas de mayor altura exhiben menores temperaturas y mayor humedad relativa, lo que favorece la regulación hídrica natural. Sin embargo, también presentan pendientes pronunciadas que aumentan el riesgo de erosión y deslizamientos si no se gestionan adecuadamente los usos del suelo.

Soluciones basadas en la Naturaleza (SbN) para la adaptación

A partir de los resultados esperados del análisis, se definieron SbN específicas para cada tipo de vulnerabilidad detectada. Estas estrategias permiten mitigar los impactos del cambio climático mediante el aprovechamiento de procesos ecológicos y el fortalecimiento de la infraestructura natural.

Tabla 4

SbN específicas para cada tipo de vulnerabilidad

Tipo de riesgo detectado	Evidencia en variables	SbN recomendadas
Altas temperaturas / baja humedad	Temp Max alta, Precipitación baja	Restauración de coberturas vegetales (reforestación con especies nativas)- Corredores biológicos para sombra y regulación térmica- Agricultura agroforestal

Tipo de riesgo detectado	Evidencia en variables	SbN recomendadas
Precipitaciones extremas / riesgo de inundación	Alta precipitación y caudales máximos elevados	Recuperación de zonas de ronda y humedales- Sistemas agroecológicos con barreras vivas- Restauración de riberas con especies de retención
Sequías o variabilidad hídrica	Caudal mínimo bajo, alta evapotranspiración	Cosecha de agua lluvia- Prácticas de conservación de suelo y humedad- Reforestación en zonas altas de recarga
Vientos fuertes / erosión	Velocidad del viento elevada	Cortinas rompevientos (en cultivos o zonas rurales)- Plantaciones de especies leñosas perennes
Altitud alta / pendientes pronunciadas	Altitud alta, erosión potencial	Sistemas silvopastoriles- Terrazas vivas o zanjas de infiltración

Nota. Estas soluciones se orientan hacia la restauración ecológica y la adaptación comunitaria, integrando prácticas sostenibles en los sistemas productivos locales (IUCN, 2020; United Nations Environment Programme, 2021).

Conclusiones

El presente estudio estableció las bases metodológicas y empíricas para el desarrollo de un modelo de aprendizaje automático orientado a la priorización de zonas críticas en tres cuencas hidrográficas de Cundinamarca (Girardot, Puerto Salgar y Guaduas) frente al cambio climático. A partir del análisis exploratorio de datos provenientes del IDEAM y Meteostat, se lograron avances significativos que demuestran la viabilidad técnica y científica del proyecto.

La recopilación y consolidación de datos oficiales permitió establecer una base de información estructurada que evidencia patrones diferenciados de vulnerabilidad entre las tres cuencas seleccionadas. Se diseñó un marco metodológico robusto que integra variables ambientales, climáticas y espaciales, estableciendo la estructura necesaria para la construcción del modelo de Machine Learning. Esta integración se alinea efectivamente con los instrumentos de política pública ambiental de Colombia, particularmente con la Gestión Integral del Recurso Hídrico (GIRH) y los planes de adaptación al cambio climático, garantizando la relevancia institucional del proyecto.

El análisis exploratorio reveló comportamientos hidrometeorológicos claramente diferenciados entre las tres cuencas estudiadas. Girardot presenta las condiciones más críticas, caracterizada por altas temperaturas promedio superiores a 28°C, baja precipitación anual inferior a 1.200 mm y un coeficiente de variación de precipitación superior al 35%, lo que indica alta irregularidad pluviométrica y mayor exposición a sequías prolongadas. Puerto Salgar exhibe un patrón de estacionalidad extrema con ciclos alternados de sequía e inundación, evidenciado por la marcada diferencia entre caudales máximos registrados durante la temporada de lluvias y los caudales mínimos en época seca, lo que refleja baja capacidad de regulación hídrica. Guaduas, ubicada en altitudes intermedias entre 800 y 1.400 metros sobre el nivel del mar,

presenta condiciones más equilibradas con temperaturas moderadas y distribución pluviométrica relativamente estable, aunque la variabilidad interanual asociada a los fenómenos de El Niño y La Niña introduce incertidumbre en la disponibilidad hídrica.

La correlación entre precipitación y caudales evidenció que las zonas con cobertura vegetal fragmentada presentan mayor escorrentía superficial y menor infiltración, agravando los efectos de las lluvias intensas. Este hallazgo valida la importancia de los ecosistemas como reguladores naturales del ciclo del agua y fundamenta la necesidad de implementar Soluciones Basadas en la Naturaleza para restaurar la capacidad de retención hídrica en las cuencas. Las variables temperatura promedio, precipitación total anual, caudales máximos y mínimos, y altitud emergieron como predictores primarios de vulnerabilidad climática, presentando correlaciones significativas con los patrones de criticidad identificados durante la caracterización territorial.

Se formuló exitosamente el Índice de Vulnerabilidad Climática (IVC) como herramienta metodológica que integra las variables hidrometeorológicas mediante ponderación diferenciada: precipitación representa el 30% del índice, temperatura y caudal el 25% cada uno, mientras que viento y altitud contribuyen con el 10% cada uno. Esta ponderación se fundamenta en la literatura científica y el análisis de expertos, priorizando aquellas variables con mayor impacto directo en la disponibilidad hídrica y la regulación ecosistémica. El IVC permite clasificar las cuencas según su nivel de vulnerabilidad, proyectándose que Girardot presente el índice más alto (vulnerabilidad crítica por déficit hídrico y térmico), Puerto Salgar vulnerabilidad media-alta (por estacionalidad extrema), y Guaduas vulnerabilidad media (por condiciones más equilibradas y altitud moderada).

La estrategia metodológica para la implementación de Random Forest y XGBoost contempla validación cruzada espacial mediante k-fold por subcuenca para mitigar problemas de autocorrelación espacial, aspecto crítico en datos georreferenciados. Se definió el uso de técnicas como SMOTE para manejar el desbalance de clases esperado en la identificación de zonas críticas, y la integración de SHAP values para garantizar la interpretabilidad del modelo y facilitar la comprensión de los factores que más inciden en la clasificación de criticidad. Aunque la aplicación práctica de los algoritmos se encuentra en fase de implementación, el análisis exploratorio permitió identificar los patrones fundamentales que guiarán el entrenamiento del modelo: alta temperatura y baja precipitación como indicadores de estrés hídrico en Girardot, estacionalidad extrema de caudales en Puerto Salgar, y variabilidad climática moderada en Guaduas.

La estructura del modelo propuesto facilita su replicabilidad en otras cuencas del departamento y permite la generación de productos cartográficos que orienten la toma de decisiones institucionales. Se identificaron Soluciones Basadas en la Naturaleza específicas para cada tipo de vulnerabilidad detectada: restauración de coberturas vegetales con especies nativas para zonas con estrés térmico, recuperación de zonas de ronda y humedales para áreas con riesgo de inundación, sistemas de cosecha de agua lluvia y prácticas de conservación de suelo para zonas con sequía recurrente, cortinas rompevientos y plantaciones de especies leñosas para áreas con erosión eólica, y sistemas silvopastoriles con terrazas vivas para pendientes pronunciadas. Estas soluciones se orientan hacia la restauración ecológica y la adaptación comunitaria, integrando prácticas sostenibles en los sistemas productivos locales.

El proyecto articula exitosamente técnicas de análisis geoespacial, modelación hidrológica y aprendizaje automático, estableciendo un enfoque interdisciplinario replicable para

la gestión integral del recurso hídrico en contextos de cambio climático. La integración de información de múltiples fuentes oficiales en un formato estructurado y normalizado supera el desafío de la heterogeneidad y fragmentación de datos ambientales que caracteriza el contexto colombiano, representando una contribución metodológica significativa para futuros estudios en la región.

Sin embargo, se identifican limitaciones importantes que deben ser consideradas. El alcance actual se circunscribe a la fase de análisis exploratorio de datos, quedando pendiente la implementación práctica de los algoritmos de clasificación supervisada y la validación del modelo predictivo mediante métricas de desempeño estándar como AUC-ROC, Precision, Recall y F1-score. La ausencia de variables socioeconómicas detalladas (densidad poblacional, índices de pobreza multidimensional, actividades productivas) limita la capacidad del análisis para incorporar la dimensión de vulnerabilidad social en la priorización de zonas críticas, aspecto fundamental para garantizar que las intervenciones propuestas consideren no solo la vulnerabilidad biofísica sino también la capacidad adaptativa de las comunidades. Adicionalmente, la resolución temporal de los datos disponibles (promedios anuales) no permite capturar la variabilidad intra-anual de eventos climáticos extremos, aspecto crucial para el diseño de sistemas de alerta temprana y la evaluación de riesgos asociados a inundaciones y sequías súbitas.

La fase de análisis exploratorio completada establece una base sólida para el desarrollo del modelo predictivo y la implementación de Soluciones Basadas en la Naturaleza en Cundinamarca. Los hallazgos preliminares confirman que la vulnerabilidad climática presenta patrones territoriales diferenciados que pueden ser modelados y anticipados mediante técnicas de aprendizaje automático, abriendo la posibilidad de transitar de una gestión ambiental reactiva a

una planificación proactiva basada en evidencia cuantitativa. El éxito en las fases subsecuentes dependerá de la incorporación de variables socioeconómicas, la validación en campo de las zonas clasificadas como prioritarias, y la articulación efectiva entre la academia, las instituciones ambientales y las comunidades locales para garantizar que las Soluciones Basadas en la Naturaleza propuestas sean técnicamente viables, socialmente aceptadas y económicamente sostenibles en el largo plazo.

Referencias

- Abhishek, K., Singh, R., & Jain, M. K. (2022). *Trend and variability analysis of rainfall and streamflow using non-parametric methods in Indian river basins. Hydrological Sciences Journal*, 67(10), 1341–1358. <https://doi.org/10.1080/02626667.2022.2085246>
- Aguirre, M. A., López Ibarra, L. I., Bolaños Trochez, F. V., González Guevara, D. F., & Buitrago Bermudez, O. (2017). Percepción del paisaje, agua y ecosistemas en la cuenca del río Dagua, Valle del Cauca, Colombia. *Perspectiva Geográfica*, 22(1). <https://doi.org/10.19053/01233769.5402>
- Aghaloo, K., Sharifi, A., Habibzadeh, N., Ali, T., & Chiu, Y.-R. (2024). How nature-based solutions can enhance urban resilience to flooding and climate change and provide other co-benefits: A systematic review and taxonomy. *Urban Forestry & Urban Greening*, 95, 128320. <https://doi.org/10.1016/j.ufug.2024.128320>
- Alarcón-Hincapié, J., Zafra-Mejía, C., & Echeverri-Prieto, L. (2019). Cambio climático y recursos hídricos en Colombia. *Revista U.D.C.A Actualidad & Divulgación Científica*, 22(2). <https://doi.org/10.31910/rudca.v22.n2.2019.1368>
- Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., Santhi, C., Harmel, R. D., van Griensven, A., Van Liew, M. W., Kannan, N., & Jha, M. K. (2012). SWAT: Model use, calibration, and validation. *Transactions of the ASABE*, 55(4), 1491–1508. <https://doi.org/10.13031/2013.42256>
- Banco Interamericano de Desarrollo, Conservación Internacional Colombia, & Colombia. Ministerio de Ambiente y Desarrollo Sostenible (2022). Adaptación al cambio climático en alta montaña colombiana: Chingaza-Sumapaz-Guerrero. Bejarano, P., Grunwaldt, A., & Andrade, Á. (Eds.). <https://doi.org/10.18235/0004046>

- Banrepcultural. (2021, 22 de abril). *Las cuencas hidrográficas de Colombia, explicadas en 5 minutos [Video]*. YouTube. <https://www.youtube.com/watch?v=UmLLv7hz6Ss>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
<https://doi.org/10.1023/A:1010933404324>
- Cantillo Romero, J. R., Estrada Romero, J. J., & Henríquez Miranda, C. (2023). Aplicación de algoritmos de aprendizaje automático en geociencia: revisión integral y desafío futuro. *Revista Ambiental Agua, Aire y Suelo*, 14(2), 9–18.
<https://doi.org/10.24054/raaas.v14i2.2783>
- Carvajal Escobar, Y. (2011). Efectos de la Variabilidad Climática (vc) y el Cambio Climático (cc) en los Recursos Hídricos de Colombia. *Entre Ciencia E Ingeniería*, 5(9), 33-61. <https://doi.org/10.31908/19098367.689>
- Chausson, A., Turner, B., Seddon, D., Chabaneix, N., Girardin, C. A. J., Kapos, V., Key, I., Roe, D., Smith, A., Woroniecki, S., & Seddon, N. (2020). Mapping the effectiveness of nature-based solutions for climate change adaptation. *Global Change Biology*, 26(11), 6134–6155. <https://doi.org/10.1111/gcb.15310>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. <https://doi.org/10.1145/2939672.2939785>
- Corporación Autónoma Regional de las Cuencas de los Ríos Negro y Nare [CORNARE]. (2023). *Gestión del cambio climático*. Corporación Autónoma Regional de las Cuencas de los Ríos Negro y Nare. <https://www.cornare.gov.co/gestion-del-cambio-climatico/>

DANE. (2020). Clasificación Industrial Internacional Uniforme – CIIU Rev. 4 A.C.

Departamento Administrativo Nacional de Estadística.

https://www.dane.gov.co/files/sen/nomenclatura/ciiu/CIIU_Rev_4_AC2022.pdf

Del Aguila Ríos, S. (2024). *Modelamiento hidrológico de cuencas andinas*. Fondo Editorial de la Universidad Nacional de San Cristóbal de Huamanga.

https://oficinas.unsch.edu.pe/vri/wp-content/uploads/2024/04/df_2422715.001_LIBRO-MODELAMIENTO-HIDROLO%CC%81GICO-final.pdf

Enel X. (2025). *Cambio climático en Colombia: así podemos bajar los niveles de dióxido de carbono*. Enel X. <https://www.enelx.com/co/es/historias/cambio-climatico-en-colombia-movilidad-electrica>

Fernández, C., & Londoño, E. (2018). *Evaluación de vulnerabilidad hídrica en cuencas de Colombia bajo escenarios de cambio climático*. Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales, 42(162), 501–515.

Gobernación de Cundinamarca. (2024). *Presentación diagnóstica ambiental Cundinamarca: Municipios con mayor riesgo climático en Cundinamarca* (p. 14). Gobernación de Cundinamarca.

https://regionmetropolitana.gov.co/sites/default/files/archivos/miscelanea/2024-11-20/227_gobernacion_presentacion-region-metropolitana-%281%29.pdf

Gómez-Miranda, I. N., Restrepo-Estrada, C., Builes-Jaramillo, A., & Porto de Albuquerque, J. (2025). Advanced AI techniques for landslide susceptibility mapping and spatial prediction: A case study in Medellín, Colombia. *Applied Computing and Geosciences*, 25, 100226. <https://doi.org/10.1016/j.acags.2025.100226>

- Gómez, J. D., & Pineda, N. (2020). *Índice de vulnerabilidad climática para cuencas hidrográficas andinas: aplicación de métodos multicriterio*. Revista Ingeniería y Región, 18(2), 55–70
- Hernández-Sampieri, R., & Mendoza Torres, C. P. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta* [PDF]. McGraw-Hill Interamericana Editores.
<https://bellasartes.upn.edu.co/wp-content/uploads/2024/11/METODOLOGIA-DE-LA-INVESTIGACION-Sampieri-Mendoza-2018.pdf>
- Hijmans, R. J. (2012). Cross-validation of species distribution models: Removing spatial sorting bias and calibration with a null model. *Ecology*, 93(3), 679–688.
<https://doi.org/10.1890/11-0826.1>
- IDEAM. (2024). *Informe hidrológico – Estación Puerto Salgar – Río Magdalena* [Boletín].
https://www.ideam.gov.co/documents/78690/73448391/08_IHD_AGOSTO_02_NOCHE.pdf
- IDEAM. (2022). Informe nacional del estado del recurso hídrico y climático en Colombia. Instituto de Hidrología, Meteorología y Estudios Ambientales.
<https://www.ideam.gov.co/sala-de-prensa/informes/Estudios-nacionales-del-agua>
- IDEAM. (2022). *Informe de gestión 2021*. Recuperado de
https://www.ideam.gov.co/sites/default/files/transparencia/planeacion/informe_de_gestion_2021.pdf
- IDEAM. (2013). *Lineamientos conceptuales y metodológicos para la evaluación regional del agua*. Instituto de Hidrología, Meteorología y Estudios Ambientales – IDEAM.
Recuperado de <https://www.ideam.gov.co/sites/default/files/prensa/boletines/2024-08->

- [23/lineamientos_conceptuales_y_metodologicos_para_la_evaluacion_regional_del_agua.pdf](#)
- IDER. (2024). Subzonas hidrográficas del departamento de Cundinamarca. Instituto de Desarrollo Regional de Cundinamarca.
<https://ider.cundinamarca.gov.co/items/a0cc5729628e4c70bd8ebf03f639f2a4>
- International Union for Conservation of Nature. (2020). *IUCN Global Standard for Nature-based Solutions*. <https://ceowatermandate.org/resources/iucn-global-standard-for-nature-based-solutions-2020/>
- Instituto del Agua. (s.f.). *Definición de cambio climático según la ONU: Un análisis profundo y actualizado*. Recuperado de <https://institutodelagua.es/cambio-climatico/definicion-de-cambio-climatico-segun-la-onucambio-climatico/>
- IPCC. (2023). *Summary for Policymakers*. In H. Lee & J. Romero (Eds.), *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 1–34). IPCC.
<https://doi.org/10.59327/IPCC/AR6-9789291691647.001>
- IPCC. (2021). *Sixth Assessment Report: Climate Change 2021 — The Physical Science Basis*. Intergovernmental Panel on Climate Change, Cambridge University Press.
<https://www.ipcc.ch/report/ar6/wg1/>
- Jeuken, A., Ray, P., Penning, W. E., Bouaziz, L., Tracy, J., Wi, S., McEvoy, S., Taner, Ü., & Hegnauer, M. (2023). Challenges for upscaling hydrological effectiveness of nature-based solutions for adaptation to climate change in watersheds. *Aquatic Ecosystem Health & Management*, 26(2), 19–32. <https://doi.org/10.14321/ae hm.026.02.019>

- Karataş, B., Çakmakçı, C., Yücel, E.S., Demir, M., Şen, F. (2026). Using Different Machine-Learning Algorithms to Predict Dissolved Oxygen Concentration in Rainbow Trout Farms. *Turkish Journal of Fisheries and Aquatic Sciences*, 26(1), TRJFAS27622. <https://doi.org/10.4194/TRJFAS27622>
- Kurnik, B., Peltonen-Sainio, P., & Drobinski, P. (2024). *Climate-induced variability of water balance in European catchments: linking aridity and hydrological response*. *Hydrology and Earth System Sciences*, 28(2), 339–356. <https://doi.org/10.5194/hess-28-339-2024>
- Lane, C. R., Creed, I. F., Golden, H. E., Leibowitz, S. G., Mushet, D. M., Rains, M. C., Wu, Q., D’Amico, E., Alexander, L. C., Ali, G. A., Basu, N. B., Bennett, M. G., Christensen, J. R., Cohen, M. J., Covino, T. P., DeVries, B., Hill, R. A., Jencso, K., Lang, M. W., McLaughlin, D. L., Rosenberry, D. O., Rover, J., & Vanderhoof, M. K. (2023). *Vulnerable waters are essential to watershed resilience*. *Ecosystems*, 26, 1–28. <https://doi.org/10.1007/s10021-021-00737-2>
- MinAmbiente. (2023). *Plan Estratégico Institucional 2023–2026*. <https://www.minambiente.gov.co/wp-content/uploads/2023/12/PLAN-ESTRATEGICO-INSTITUCIONAL-PEI-2023-2026-MINAMBIENTE-1.pdf>
- MinAmbiente. (2022). *Guía para la formulación de Planes Integrales de Gestión del Cambio Climático Territorial (PIGCCT)*. Ministerio de Ambiente y Desarrollo Sostenible. https://www.minambiente.gov.co/wp-content/uploads/2022/10/GUIA-PARA-LA-FORMULACION-PIGCCT-web- ISBN_VF.pdf
- MinAmbiente. (2021). *Política Nacional de Cambio Climático*. Ministerio de Ambiente y Desarrollo Sostenible. <https://www.minambiente.gov.co/wp-content/uploads/2022/01/9.-Politica-Nacional-de-Cambio-Climatico.pdf>

- Martínez-Santos, P., Aristizábal, H. F., Díaz-Alcaide, S., & Gómez-Escalonilla, V. (2021). *Predictive mapping of aquatic ecosystems by means of support vector machines and random forests*. *Journal of Hydrology*, 595, 126026. <https://doi.org/10.1016/j.jhydrol.2021.126026>
- Martínez, J. A., & Poveda, G. (2019). *Influence of El Niño and La Niña on precipitation and streamflow in Colombian Andean basins*. *Hydrological Sciences Journal*, 64(6), 687–702. <https://doi.org/10.1080/02626667.2019.1595629>
- Maldonado Santana, C., Franco-Billini, C., & Jáuregui-Haza, U. J. (2023). *Gestión integrada de las cuencas hidrográficas: hacia un nuevo paradigma en la gobernanza del agua*. *Ciencia, Ambiente y Clima*, 6(2). <https://doi.org/10.22206/cac.2023.v6i2.295>
- Molnar, C. (2022). *Interpretable machine learning: A guide for making black box models explainable* (2nd ed.). Lulu.com. <https://christophm.github.io/interpretable-ml-book/>
- Munar, A. M., Mendez, N., Narvaez, G., Campo Zambrano, F., Motta-Marques, D., & Fialho Brêda, J. P. L. (2023). *Modelling the climate change impacts on river discharge and inundation extent in the Magdalena River basin – Colombia*. *Hydrological Sciences Journal*, 68(10), 1286–1300. <https://doi.org/10.1080/02626667.2023.2215932>
- Munar Samboní, A. M., Méndez Pedroza, N. M., Vinasco Guzmán, M. C., Guzmán Oliveros, M. S., Cortés Orozco, C. P., Trujillo Zapata, S. A., Valbuena Calderón, O. E., Bravo Gaviria, M. A., Caicedo Díaz, G. E., Herrera Cerquera, J. P., Ramírez Córdoba, G. A., Montealegre Torres, W. I., Mesa Mejía, C. G., Tenorio Sánchez, P. A., Florido Cuellar, B. A., Carvajal Pinilla, L. A., Caicedo Bolaños, L. P., Núñez Burgos, M. Á., Fernández Chávez, A. E., Fernández Hidalgo, V., Castro Trujillo, J. X., Jiménez Cruz, L. Y. (2022).

- Gestión integrada de la cuenca alta del río Magdalena: instrumentos para su evaluación y planificación*. Sello Editorial UNAD. <https://doi.org/10.22490/9789586518642>
- OECD/European Union/European Commission, Joint Research Centre. (2008). *Handbook on Constructing Composite Indicators: Methodology and user guide*. OECD Publishing. <https://doi.org/10.1787/9789264043466-en>
- PNUD. (2022). Gestión del riesgo y adaptación al cambio climático en territorios rurales de Colombia. Programa de las Naciones Unidas para el Desarrollo. <https://www.undp.org>
- Programa de las Naciones Unidas para el Medio Ambiente. (2025). *Facing climate storm, one town turns to mangroves for protection*. United Nations Environment Programme. <https://www.unep.org/news-and-stories/story/facing-climate-storm-one-town-turns-mangroves-protection>
- Qiu, H., Niu, J., Baas, D. G., & Phanikumar, M. S. (2023). An integrated watershed-scale framework to model nitrogen transport and transformations. *Science of the Total Environment*, 882, 163348. <https://doi.org/10.1016/j.scitotenv.2023.163348>
- Richards, D., Worden, D., Song, X. P., & Lavorel, S. (2024). Harnessing generative artificial intelligence to support nature-based solutions. *People and Nature*, 6(2), Article e10622. <https://doi.org/10.1002/pan3.10622>
- Rodríguez, C., Parra, H., Rondón, J., Miranda, L., Galindo, L., Díaz, V., Robayo, W., & Torres, Z. (2020). *Evaluación Regional del Agua ERA. Cuenca media del río Bogotá*. Biblioteca Digital Ambiental SIE. <https://sie.car.gov.co/items/0d84f334-c981-40c8-9357-d7ec5a336a81>
- Roldán Valcarce, A. (2024). *Desarrollo de herramientas de modelización espacial de riesgos ambientales urbanos para facilitar su mitigación mediante soluciones basadas en la*

naturaleza (Tesis de maestría, Universidad de Cantabria).

<https://hdl.handle.net/10902/31998>

- Sáenz Segura, F. (1995). *Identificación de áreas críticas para el manejo de un sector de la cuenca del río Pacuare, Costa Rica* (Tesis de maestría, CATIE - Centro Agronómico Tropical de Investigación y Enseñanza). <https://repositorio.catie.ac.cr/handle/11554/2329>
- Observatorio Ambiental de Bogotá. (2024). *Bogotá, una de las ciudades más vulnerables a los efectos del cambio climático: Visor Geográfico Ambiental*. Alcaldía Mayor de Bogotá. <https://oab.ambientebogota.gov.co/bogota-vulnerable-cambio-climatico/>
- Sauer, J., & Chang, H. (2025). *Watershed management may overcome effects of climate change on stream water quality in the Portland metropolitan area*. *Journal of Environmental Management*, 391, 126304. <https://doi.org/10.1016/j.jenvman.2025.126304>
- Sullivan, C. A., & Meigh, J. R. (2007). Integration of the biophysical and social dimensions of water scarcity: The Water Poverty Index. *Water Resources Management*, 21(1), 19–36.
- Sun, C., Rao, Q., Chen, B., Liu, X., Ikram, R. M. A., Li, J., Wang, M., & Zhang, D. (2024). Mechanisms and applications of nature-based solutions for stormwater control in the context of climate change: A review. *Atmosphere*, 15(4), 403. <https://doi.org/10.3390/atmos15040403>
- Sun, Q., Miao, C., & Duan, Q. (2021). *Changes in precipitation and temperature extremes over the global land areas*. *Nature Communications*, 12(1), 4272. <https://doi.org/10.1038/s41467-021-24543-8>
- United Nations Environment Programme. (2021). *Guidelines for Integrating Ecosystem-based Adaptation into National Adaptation Plans: Supplement to the UNFCCC NAP Technical*

- Guidelines. <https://www.unep.org/resources/toolkits-manuals-and-guides/guidelines-integrating-ecosystem-based-adaptation-eba>
- UN-Water. (2023). *Climate change and water*. United Nations Water. Recuperado de <https://www.unwater.org/water-facts/climate-change>
- Qiu, H., Niu, J., Baas, D. G., & Phanikumar, M. S. (2023). An integrated watershed-scale framework to model nitrogen transport and transformations. *Science of the Total Environment*, 882, 163348. <https://doi.org/10.1016/j.scitotenv.2023.163348>
- Vélez-Duque, J., & Arteaga-Morales, S. M. (2022). Apartadó's River Master Plan: Mitigating the risk of flooding in the face of climate change in a biodiversity hotspot. *Environmental Sciences Proceedings*, 15(1), 44. <https://doi.org/10.3390/environsciproc2022015044>
- Vélez, J. J., & Ocampo, O. L. (2012, septiembre 9–12). *Vulnerabilidad al cambio climático y a la variabilidad climática en cuencas hidrográficas de alta montaña: Río Chinchiná–Colombia*. En XXV Congreso Latinoamericano de Hidráulica (IAHR), San José, Costa Rica.
- https://www.researchgate.net/publication/271202549_Vulnerabilidad_al_Cambio_Climatico_y_a_la_Variabilidad_Climatica_en_Cuencas_Hidrograficas_de_Alta_Montana_Rio_Chinchina-Colombia
- Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2020). *A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index (SPEI)*. *Climate*, 8(3), 45. <https://doi.org/10.3390/cli8030045>
- Villamizar, S. R., Pineda, S. M., & Carrillo, G. A. (2019). *The effects of land use and climate change on the water yield of a watershed in Colombia*. **Water**, 11(2), 285. <https://doi.org/10.3390/w11020285>

- Wang, Y., Khodadadzadeh, M., & Zurita-Milla, R. (2023). Spatial+: A new cross-validation method to evaluate geospatial machine learning models. *International Journal of Applied Earth Observation and Geoinformation*, 121, 103364.
<https://doi.org/10.1016/j.jag.2023.103364>
- Wang, F., Shao, W., Yu, H., Kan, G., He, X., Zhang, D., Ren, M., & Wang, G. (2020). Re-evaluation of the Power of the Mann–Kendall Test for Detecting Monotonic Trends in Hydrometeorological Time Series. *Frontiers in Earth Science*, 8, 14.
<https://doi.org/10.3389/feart.2020.00014>
- Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30, 79–82. <https://doi.org/10.3354/cr030079>
- World Wildlife Fund [WWF]. (s. f.). ¿En qué consisten las soluciones basadas en la naturaleza y cómo pueden ayudarnos a enfrentar el cambio climático? *World Wildlife Fund*.
<https://www.worldwildlife.org/descubre-wwf/historias/en-que-consisten-las-soluciones-basadas-en-la-naturaleza-y-como-pueden-ayudarnos-a-enfrentar-el-cambio-climatico>
- Yates, D., Sieber, J., Purkey, D., & Huber-Lee, A. (2005). WEAP21—A demand-, priority-, and preference-driven water planning model. Part 1: Model characteristics. *Water International*, 30(4), 487–500. <https://doi.org/10.1080/02508060508691893>