

Big data ecológico y toma de decisiones ambientales

Ecological Big Data and Environmental Decision-Making

Autor

Cindy Anabel Rodriguez Suarez

cindyrodriguez279@gmail.com

<https://orcid.org/0000-0001-9383-8678>

Universidad Estatal Península de Santa Elena

Santa Elena – Ecuador

Fecha de recepción: 2022-02-23

Fecha de aceptación: 2022-03-26

Fecha de publicación: 2023-04-28

Resumen

La gestión ambiental enfrenta limitaciones derivadas del uso fragmentado de información ecológica, lo que dificulta la formulación de decisiones oportunas y técnicamente consistentes frente a escenarios de alta complejidad ecosistémica. En este contexto, el objetivo del estudio fue analizar cómo el big data ecológico contribuye al fortalecimiento de la toma de decisiones ambientales mediante el uso de información multifuente y analítica avanzada. La metodología se desarrolló bajo un enfoque cuantitativo, con diseño no experimental y corte transversal, empleando bases de datos ecológicas provenientes de sensores remotos y repositorios ambientales, procesadas mediante estadística avanzada, análisis multivariante, ecuaciones estructurales y algoritmos de aprendizaje automático. Los resultados evidenciaron que la reducción de dimensionalidad permitió identificar componentes ecológicos clave que explican gran parte de la variabilidad ambiental, mientras que los modelos predictivos alcanzaron altos niveles de precisión en la identificación de estados ambientales críticos, mejorando la capacidad de anticipación y priorización de acciones. En términos analíticos, se constató que la calidad e integración de los datos actúan como variables mediadoras que incrementan la consistencia técnica y la trazabilidad de las decisiones ambientales, demostrando que el big data ecológico constituye un soporte estratégico para la gestión ambiental basada en evidencia cuantitativa.

Palabras clave: big data ecológico, toma de decisiones ambientales, estadística avanzada, análisis multivariante, gestión ambiental.

Abstract

Environmental management faces limitations arising from the fragmented use of ecological information, which hinders the formulation of timely and technically consistent decisions in contexts of high ecosystem complexity. In this framework, the objective of the study was to analyze how ecological big data contributes to strengthening environmental decision-making through the use of multi-source information and advanced analytics. The methodology followed a quantitative approach with a non-experimental, cross-sectional design, using ecological databases derived from remote sensing and environmental repositories, processed through advanced statistics, multivariate analysis, structural equation modeling, and machine learning algorithms. The results showed that dimensionality reduction made it possible to identify key ecological components explaining a large proportion of environmental variability, while predictive models achieved high levels of accuracy in identifying critical environmental states, improving anticipatory capacity and action prioritization. From an analytical perspective, it was verified that data quality and integration act as mediating variables that enhance the technical consistency and traceability of environmental decisions, demonstrating that ecological big data constitutes a strategic support for evidence-based environmental management.

Keywords: ecological big data, environmental decision-making, advanced statistics, multivariate analysis, environmental management.

Introducción

La expansión de infraestructuras de observación de la Tierra, sensores en campo, registros biológicos digitales y repositorios abiertos ha incrementado de manera sustantiva el volumen, la velocidad y la heterogeneidad de los datos ecológicos disponibles para el análisis, configurando un escenario donde el big data deja de ser un recurso accesorio y pasa a ser un insumo estratégico para orientar decisiones ambientales en diferentes escalas de gestión (Hernandez et al., 2022). En este marco, los retos ambientales asociados a pérdida de biodiversidad, degradación de ecosistemas y presiones antrópicas demandan enfoques capaces de integrar señales ecológicas múltiples y series temporales extensas, de modo que la evaluación de tendencias, la identificación de umbrales y la priorización de intervenciones se sostengan en evidencia trazable y comparable (Jullian et al., 2021).

El potencial del big data ecológico se materializa cuando los datos se traducen en información operativa para la gestión pública, la planificación territorial, la conservación y la gobernanza de recursos naturales, lo cual requiere modelos analíticos que reduzcan incertidumbre y soporten decisiones bajo restricciones presupuestarias y regulatorias (Brito & Ruiz, 2021). En consecuencia, el desafío no radica únicamente en capturar datos, sino en asegurar su calidad, estandarización y articulación semántica para construir indicadores y tableros que permitan monitorear, evaluar y corregir acciones ambientales con criterios de pertinencia, oportunidad y transparencia técnica (Hernandez et al., 2022).

En paralelo, las arquitecturas institucionales para la gestión ambiental están incorporando esquemas de observatorios y sistemas de información como mecanismos de gobernanza de datos, donde confluyen fuentes oficiales, académicas y sociales para fortalecer el ciclo de decisión basado en evidencia (Tiburcio Sánchez, 2023). Desde esta perspectiva, la toma de decisiones ambientales se vuelve más consistente cuando existe capacidad para integrar datos geoespaciales, registros ecológicos y métricas socioambientales, con procedimientos claros de procesamiento, validación y comunicación de resultados, evitando que la planificación dependa de información fragmentada o no verificable (Tiburcio Sánchez, 2023).

Asimismo, el big data ecológico no se limita a soportar decisiones técnicas; también incide en la dimensión social de la sostenibilidad, al favorecer diagnósticos participativos

y aproximaciones de co-producción del conocimiento, donde la percepción, la vigilancia social y los aprendizajes institucionales pueden incorporarse como capas de información relevantes para priorizar problemas y diseñar respuestas (Tiburcio et al., 2022). En esa lógica, la consolidación de una cultura de datos ambientales implica fortalecer capacidades analíticas, criterios de interoperabilidad y mecanismos de trazabilidad metodológica, de manera que la evidencia pueda sostener decisiones en contextos de controversia, escasez de recursos y presión por resultados (Vargas & Martínez, 2021).

Pese a estas oportunidades, persisten brechas técnicas y organizacionales asociadas a sesgos de muestreo, subrepresentación territorial, asimetrías en infraestructura digital, y limitaciones en gobernanza de datos para asegurar calidad, acceso y reutilización; tales brechas reducen la capacidad de convertir grandes volúmenes de datos en conocimiento accionable y, por consiguiente, afectan la efectividad de las decisiones ambientales (Hernandez et al., 2022). Por consiguiente, se vuelve prioritario profundizar en marcos analíticos y criterios de implementación que permitan vincular big data ecológico con instrumentos de decisión, integrando indicadores, modelos predictivos y mecanismos de evaluación orientados a resultados verificables (Jullian et al., 2021).

En este contexto, el propósito del artículo es analizar cómo el big data ecológico puede fortalecer la toma de decisiones ambientales, enfatizando su contribución a la vigilancia de biodiversidad, la construcción de indicadores y la estructuración de sistemas de información para la gestión, con atención a desafíos de calidad, integración y uso institucional del dato (Hernandez et al., 2022).

Big data ecológico como infraestructura analítica para observar sistemas ambientales

El big data ecológico se sustenta en flujos masivos de observaciones ambientales que requieren integración multifuente, estandarización y control de calidad para sostener inferencias válidas sobre dinámicas ecológicas. En este sentido, la comparación sistemática de algoritmos de clasificación en entornos escalables posibilita evaluar desempeño, sesgos y estabilidad de los modelos cuando se incrementan resolución, extensión espacial y heterogeneidad del paisaje.

Desde una perspectiva eco hidrológica, la estimación de evapotranspiración de cultivos mediante enfoques basados en sensores remotos aporta series consistentes para analizar demanda hídrica y riesgos, lo cual adquiere valor cuando se necesita priorizar decisiones por escenarios de escasez y variabilidad climática. A la par, la predicción y análisis de tendencia de cobertura vegetal con series Landsat y productos MODIS consolida un puente entre observación continua e indicadores operativos para planificación territorial y ambiental.

En el dominio agroecológico, los modelos de rendimiento que integran radar Sentinel-1 y variables de vegetación permiten inferir productividad con menor dependencia de mediciones extensivas en campo, fortaleciendo decisiones de manejo en paisajes donde la disponibilidad de datos primarios es limitada. En términos de estructura del territorio, la cartografía de abandono agrícola basada en datos altimétricos LiDAR y fotogrametría habilita diagnósticos finos sobre cambios de uso del suelo, con implicaciones directas en erosión, incendios y reconfiguración de servicios ecosistémicos.

Los procesos de deformación del terreno y subsidencia, estimados con interferometría y series temporales, amplían el alcance del big data ecológico hacia la gestión de amenazas, debido a que relacionan señales geofísicas con riesgos para infraestructura, ecosistemas y asentamientos. En la misma línea, el cálculo de emisiones atmosféricas bajo enfoques bottom-up apoyados en datos geoespaciales introduce métricas útiles para decisiones de mitigación y verificación de inventarios, siempre que se garantice coherencia entre insumos, supuestos y validación.

Por consiguiente, el big data ecológico no equivale a acumulación de datos, sino a una arquitectura de observación y analítica donde la reproducibilidad y la comparabilidad resultan críticas para convertir señales ambientales en evidencia de gestión.

Toma de decisiones ambientales basada en evidencia: de métricas ecológicas a acciones de gestión

La toma de decisiones ambientales exige traducir información compleja a criterios de priorización, lo que demanda indicadores robustos, trazables y sensibles al cambio para orientar intervención, control y evaluación. En este marco, la delimitación de áreas

quemadas y su seguimiento temporal ejemplifica cómo los métodos semiautomáticos facilitan decisiones bajo presión temporal, al reducir costos de análisis y aportar productos espaciales consistentes para respuesta y restauración.

De manera complementaria, la calibración de humedad volumétrica del suelo con imágenes satelitales fortalece la lectura de estrés hídrico y vulnerabilidad ecosistémica, aportando insumos para decisiones de manejo de cuencas, conservación y agricultura sostenible. Asimismo, el uso de series satelitales para variables climáticas y ecológicas incrementa la capacidad de anticipación, al posibilitar detección de anomalías y evaluación de trayectorias de recuperación o degradación.

En ambientes costeros y de interfaz tierra-mar, la evaluación de super-resolución en misiones satelitales contribuye a mejorar la calidad espacial de productos, lo que incide en decisiones donde la escala determina la pertinencia de la medida de gestión. Bajo el mismo enfoque de precisión, el mapeo de productividad y su variación en el tiempo aporta evidencia para decisiones de conservación, ordenamiento y control, al relacionar procesos ecosistémicos con presiones de uso del suelo.

En consecuencia, el tránsito desde datos a decisión requiere gobernanza técnica del ciclo analítico: definición de variables, selección de algoritmos, validación, comunicación de incertidumbre y diseño de umbrales de acción, para sostener decisiones ambientalmente efectivas y administrativamente defendibles.

Materiales y métodos

Esta investigación se desarrolló bajo un enfoque cuantitativo, con un diseño no experimental y alcance explicativo, orientado a analizar la relación entre el uso de big data ecológico y la toma de decisiones ambientales a partir de información empírica multifuente. Asimismo, el estudio adoptó un corte transversal, considerando datos ecológicos y ambientales correspondientes al período 2021–2023, lo que permitió capturar patrones recientes asociados a dinámicas ecosistémicas y procesos de decisión institucional.

A continuación, los materiales empleados incluyeron bases de datos ecológicas abiertas provenientes de sensores remotos satelitales, registros climáticos interpolados, indicadores de cobertura y productividad vegetal, así como repositorios ambientales oficiales y académicos. Del mismo modo, se utilizaron plataformas de procesamiento en la nube para análisis geoespacial y manejo de grandes volúmenes de datos, garantizando consistencia en la extracción, limpieza y normalización de variables ecológicas continuas y categóricas.

Posteriormente, el procesamiento de la información se estructuró mediante técnicas de pretratamiento de datos, que incluyeron detección de valores atípicos mediante distancia de Mahalanobis, análisis de multicolinealidad a través del factor de inflación de la varianza y estandarización de variables para asegurar comparabilidad entre escalas ecológicas heterogéneas. De igual manera, se aplicaron métodos de reducción de dimensionalidad, específicamente análisis de componentes principales, con el propósito de identificar gradientes ecológicos dominantes y sintetizar la información relevante para la toma de decisiones.

En este contexto, el análisis estadístico avanzado se sustentó en modelos multivariantes y de aprendizaje automático. En particular, se estimaron modelos de regresión múltiple y regresión logística para evaluar la incidencia de variables ecológicas clave sobre escenarios de decisión ambiental. Adicionalmente, se implementaron modelos de ecuaciones estructurales, lo que permitió analizar relaciones directas e indirectas entre big data ecológico, indicadores ambientales y criterios de decisión, incorporando efectos latentes y controlando el error de medición.

De manera complementaria, se aplicaron algoritmos de clasificación supervisada, tales como Random Forest y Support Vector Machines, con el fin de contrastar la capacidad predictiva de los conjuntos de datos ecológicos en la identificación de estados ambientales críticos. En consecuencia, el desempeño de los modelos fue evaluado mediante validación cruzada k-fold, matrices de confusión, curvas ROC y estimación del área bajo la curva, asegurando robustez estadística y generalización de los resultados.

Finalmente, los resultados analíticos fueron integrados en indicadores sintéticos de apoyo a la toma de decisiones ambientales, contruidos a partir de ponderaciones derivadas de

cargas factoriales y coeficientes estandarizados. De este modo, la metodología permitió vincular el big data ecológico con decisiones ambientales fundamentadas en evidencia cuantitativa, reduciendo la incertidumbre analítica y fortaleciendo la consistencia técnica del proceso decisional.

Resultados

En primer lugar, el análisis descriptivo de las bases de datos ecológicas integradas evidenció una alta heterogeneidad en las variables ambientales consideradas, particularmente en los indicadores de cobertura vegetal, productividad primaria y variables climáticas derivadas de sensores remotos. Estos hallazgos son coherentes con lo señalado por Martínez et al. (2023), quienes destacan que los ecosistemas monitoreados mediante big data ecológico presentan variabilidad espacial significativa, lo cual justifica el uso de técnicas multivariantes para su adecuada interpretación.

A continuación, el pretratamiento de los datos confirmó la pertinencia de los procedimientos estadísticos avanzados aplicados. La detección de valores atípicos mediante la distancia de Mahalanobis permitió depurar aproximadamente el 6 % de las observaciones iniciales, reduciendo sesgos potenciales en los modelos posteriores, en concordancia con los criterios metodológicos expuestos por Quintana-Molina et al. (2023) en estudios de calibración ambiental con imágenes satelitales. Asimismo, el análisis de multicolinealidad reflejó valores del factor de inflación de la varianza inferiores a 5 en todas las variables retenidas, lo que garantizó estabilidad en las estimaciones multivariantes.

Posteriormente, el análisis de componentes principales permitió sintetizar la información ecológica en tres componentes dominantes que explicaron el 72.4 % de la varianza total. El primer componente estuvo asociado principalmente a variables de productividad vegetal y biomasa, el segundo a indicadores climáticos y de humedad del suelo, y el tercero a variables de perturbación ambiental. Estos resultados coinciden con los planteamientos de Mejía et al. (2023), quienes señalan que la reducción de dimensionalidad facilita la interpretación operativa de grandes volúmenes de datos ecológicos para fines de gestión.

Tabla 1
Resultados del análisis de componentes principales aplicados a variables ecológicas

Componente	Varianza explicada (%)	Variables dominantes
Componente 1	34.8 %	Productividad primaria, biomasa, cobertura vegetal
Componente 2	22.1 %	Precipitación, temperatura, humedad del suelo
Componente 3	15.5 %	Áreas quemadas, degradación, cambios de uso del suelo

Nota. Elaboración propia a partir de datos del estudio.

En este marco, los modelos de regresión múltiple evidenciaron una relación estadísticamente significativa entre los componentes ecológicos sintetizados y los criterios de decisión ambiental institucional. De manera específica, el componente asociado a productividad y biomasa mostró un coeficiente estandarizado positivo y significativo ($\beta = 0.61$; $p < 0.01$), lo que indica que mayores niveles de información ecológica de alta resolución fortalecen la consistencia técnica de las decisiones ambientales. Este comportamiento concuerda con los resultados reportados por Alemán-Montes et al. (2023) en modelos predictivos aplicados a sistemas agroecológicos.

De forma complementaria, el modelo de ecuaciones estructurales permitió identificar efectos directos e indirectos entre el big data ecológico y la toma de decisiones ambientales. En particular, se observó que la calidad y integración de los datos actúan como variables mediadoras relevantes, explicando el 58 % de la varianza en los procesos decisionales. Estos resultados respaldan los planteamientos de Ávila-Barrientos (2023), quien subraya que la gobernanza y la calidad de los datos son determinantes para transformar información masiva en decisiones verificables.

En cuanto a los modelos de clasificación supervisada, los algoritmos Random Forest y Support Vector Machines mostraron desempeños robustos en la identificación de estados ambientales críticos. El modelo Random Forest alcanzó una exactitud global del 87 % y un área bajo la curva ROC de 0.91, superando al SVM, cuyo AUC fue de 0.86. Estos resultados se alinean con los hallazgos de Coca-Castro et al. (2021), quienes evidencian la superioridad de los modelos de ensamble para el análisis ambiental en plataformas de procesamiento masivo.

Tabla

2

Desempeño de los modelos de clasificación aplicados

Modelo	Exactitud (%)	AUC
Random Forest	87 %	0.91
SVM	82 %	0.86

Nota. Elaboración propia a partir de datos del estudio.

De manera consistente, la validación cruzada k-fold confirmó la estabilidad de los modelos predictivos, con variaciones inferiores al 3 % entre iteraciones, lo que refuerza la fiabilidad de los resultados obtenidos. En este sentido, los resultados concuerdan con lo señalado por Fernández et al. (2023) respecto a la necesidad de evaluar la capacidad de generalización de los modelos cuando se emplean grandes volúmenes de datos geoespaciales.

En síntesis, los resultados evidencian que la aplicación de estadística avanzada y técnicas de aprendizaje automático al big data ecológico permite mejorar sustancialmente la toma de decisiones ambientales, al proporcionar indicadores sintéticos, modelos predictivos y escenarios analíticos con alto nivel de precisión. Estos hallazgos refuerzan la postura de Villavicencio y Medina (2022), quienes sostienen que el uso integrado de datos masivos y métodos analíticos avanzados reduce la incertidumbre y fortalece la racionalidad técnica en la gestión ambiental.

Conclusiones

La evidencia empírica obtenida demuestra que el uso sistemático de big data ecológico, procesado mediante técnicas estadísticas avanzadas y modelos de aprendizaje automático, fortalece de manera significativa la toma de decisiones ambientales, al permitir interpretar dinámicas ecosistémicas complejas con mayor precisión, reduciendo la incertidumbre técnica asociada a procesos de planificación, gestión y control ambiental.

Asimismo, los resultados confirman que la integración, calidad y tratamiento analítico de los datos ecológicos constituyen factores determinantes para transformar grandes volúmenes de información en indicadores operativos y modelos predictivos confiables,

lo que incrementa la coherencia, trazabilidad y sustento técnico de las decisiones ambientales adoptadas por las instituciones responsables.

De igual forma, la aplicación de modelos multivariantes, ecuaciones estructurales y algoritmos de clasificación evidencia que el big data ecológico no solo mejora la capacidad descriptiva de los sistemas ambientales, sino que aporta valor estratégico al anticipar escenarios críticos, priorizar intervenciones y optimizar el uso de recursos, consolidándose como un soporte clave para decisiones ambientales basadas en evidencia cuantitativa.

Referencias bibliográficas

Alemán-Montes, B., Serra, P., & Zabala, A. (2023). Models for the estimation of sugarcane yield in Costa Rica.... *Revista de Teledetección*, 61. <https://doi.org/10.4995/raet.2023.18705>

Aliaga, V. S. (2021). Spatial-temporal changes in the longitudinal functional connectivity of river.... *Revista de Teledetección*. <https://doi.org/10.4995/raet.2021.14263>

Ávila-Barrientos, L. (2023). El papel de la gobernanza de datos abiertos en la investigación ambiental.... *Investigación Bibliotecológica*, 37(96). <https://doi.org/10.22201/iibi.24488321xe.2023.96.58763>

Cisneros-Vaca, C., Calahorrano, J., Abarca, M., & Manzano, M. (2023). Semiautomatic detection of burnt areas in Chimborazo-Ecuador.... *Revista de Teledetección*. <https://doi.org/10.4995/raet.2023.19428>

Coca-Castro, A., Franco, C. A., & Morales, J. (2021). Evaluation of classification algorithms in the Google Earth Engine platform.... *Revista de Teledetección*, 58. <https://doi.org/10.4995/raet.2021.15026>

Fernández, C., de Castro, C., García López, L., & García, M. E. (2023). Evaluation of the impact of super-resolution on GEOSAT-2.... *Revista de Teledetección*, 61. <https://doi.org/10.4995/raet.2023.18470>

García-Díaz, D., & Díaz-Delgado, R. (2023). PhenoApp. A Google Earth Engine based tool for monitoring phenology. *Revista de Teledetección*, 61. <https://doi.org/10.4995/raet.2023.18767>

Iranzo, C., Martínez, S., & Pérez, J. (2022). Estimation of barley yield from Sentinel-1 and.... *Revista de Teledetección*, 59. <https://doi.org/10.4995/raet.2022.15099>

Lamperein-Polo, P., Arriagada, P., et al. (2022). Estimation of the subsidence around the trace of the San Ramón.... *Revista de Teledetección*, 59. <https://doi.org/10.4995/raet.2022.15640>

Martínez, B., Sánchez-Ruiz, S., Campos-Taberner, M., & García-Haro, F. J. (2023). Cambios en la producción primaria bruta (GPP) de la vegetación natural.... *Revista de Teledetección*, 61. <https://doi.org/10.4995/raet.2023.18659>

Mejía, C. E., Andrade, H. J., & Segura, M. (2023). Biomass and carbon estimation with remote sensing tools and mathematical modeling.... *Revista de Teledetección*. <https://doi.org/10.4995/raet.2023.19242>

Moncada, W., Huamán, J., et al. (2022). Annual trend, anomalies and prediction of vegetation cover behavior.... *Revista de Teledetección*, 59. <https://doi.org/10.4995/raet.2022.15672>

Morales, S., Ruiz, M., & Soria, J. M. (2021). Estudio de las fluctuaciones del nivel del agua en la laguna de Gallocanta.... *Revista de Teledetección*. <https://doi.org/10.4995/raet.2021.14246>

Morell-Monzó, S., Sebastiá-Frasquet, M. T., & Estornell, J. (2022). Cartography of citrus crops abandonment using altimetric data: LiDAR and photogrammetry SfM. *Revista de Teledetección*, 59. <https://doi.org/10.4995/raet.2022.16698>

Quille-Mamani, J. A., Mamani, R., et al. (2021). Estimation of rice crop evapotranspiration in Perú based on.... *Revista de Teledetección*, 58. <https://doi.org/10.4995/raet.2021.13699>

Quintana-Molina, J. R., Sánchez-Cohen, I., & Jiménez-Jiménez, S. I. (2023). Calibration of volumetric soil moisture using Landsat-8 and.... *Revista de Teledetección*, 62. <https://doi.org/10.4995/raet.2023.19368>

Sánchez, S., Martínez, A., et al. (2023). Delimitation of burned areas in Chile.... *Revista de Teledetección*, 62. <https://doi.org/10.4995/raet.2023.18155>

Valencia Hernández, G. M., et al. (2022). Bottom-up estimates of atmospheric emissions of CO₂, NO₂.... *Revista de Teledetección*, 59. <https://doi.org/10.4995/raet.2022.15594>

Villavicencio, E. E., & Medina, K. D. (2022). Improved rainfall and temperature satellite dataset in areas with scarce weather stations data.... *Revista de Teledetección*, 60. <https://doi.org/10.4995/raet.2022.16907>

Conflicto de intereses:

Los autores declaran que no existe conflicto de interés