# Construcción y evaluación de un índice cantonal de vulnerabilidad socioambiental en la Amazonía ecuatoriana mediante ciencia de datos y aprendizaje automático no supervisado

**Autora: María Susana Espinosa Soto**

**Tutor: Edison Vicente Mora Londoño**

**Pontificia Universidad Católica del Ecuador**

**Maestría en Sistemas de Información con mención en Ciencia de Datos**

**2026**

# Agradecimientos

Mi gratitud a la Amazonía por la contundencia de la vida y de las causas que merecen ser vividas. A ella y a las personas que me han forjado como ser humano, gracias infinitas por el apoyo y cariño incondicionales.

# Índice General

**Agradecimientos** ..... 2
**Índice General** ..... 3
**Índice de Figuras** ..... 4
**Índice de Tablas** ..... 5
**Índice de Anexos** ..... 5
**1. Capítulo 1. Introducción** ..... 6
1.1. Introducción ..... 6
1.2. Objetivos ..... 9
1.2.1. Objetivo general ..... 10
1.2.2. Objetivos específicos ..... 10
1.3. Justificación ..... 10
1.4. Preguntas de investigación ..... 15
**2. Capítulo 2. Desarrollo** ..... 16
2.1. Marco teórico y conceptual ..... 16
2.2. Metodología ..... 34
2.2.1. Diseño, alcance y unidad de análisis ..... 35
2.2.2. Fuentes de datos y adquisición ..... 36
2.2.3. Extracción, transformación, carga y organización del data mart ..... 38
2.2.4. Uso de inteligencia artificial generativa en el procesamiento ..... 41
2.2.5. Homologación territorial y tratamiento de Morona y Sevilla Don Bosco ..... 42
2.2.6. Construcción y selección de indicadores ..... 42
2.2.7. Preprocesamiento, normalización y análisis de componentes principales ..... 44
2.2.8. Clustering y caracterización de perfiles territoriales ..... 45
2.2.9. Construcción del índice y clasificación relativa ..... 46
2.2.10. Validación, interpretabilidad y análisis espacial ..... 47
2.2.11. Herramientas técnicas y reproducibilidad ..... 48
2.3. Resultados y análisis ..... 49
2.3.1. Construcción del índice y distribución territorial ..... 50
2.3.2. Estructura dimensional de la vulnerabilidad ..... 56
2.3.3. Composición e interpretabilidad del índice ..... 61
2.3.4. Perfiles territoriales identificados mediante clustering ..... 64
2.3.5. Comparación de ponderaciones y sensibilidad metodológica ..... 68

2.3.6. Patrón espacial del índice ..... 72
2.3.7. Lectura socioterritorial e implicaciones para la priorización ..... 74
2.3.8. Discusión de resultados en relación con los antecedentes y las decisiones metodológicas ..... 76
2.3.9. Alcances, desafíos y oportunidades de fortalecimiento ..... 78
**3. Capítulo 3. Conclusiones y Recomendaciones** ..... 81
3.1. Conclusiones ..... 81
3.2. Limitaciones ..... 84
3.3. Recomendaciones ..... 88
Referencias ..... 91
Anexos ..... 97
Anexo A. Inventario y arquitectura del data mart territorial ..... 97
Anexo B. Matriz de fuentes y archivos definitivos ..... 98
Anexo C. Diccionario maestro de variables ..... 98
Anexo D. Catálogo y evaluación de indicadores ..... 99
Anexo E. Resultados detallados del análisis de componentes principales ..... 99
Anexo F. Resultados detallados del clustering ..... 100
Anexo G. Interpretabilidad y sensibilidad por cantón ..... 100
Anexo H. Productos cartográficos y visualizaciones complementarias ..... 101
Anexo I. Controles de calidad y auditoría geoespacial ..... 101
Anexo J. Código y reproducibilidad ..... 102
Anexo K. Declaración y bitácora de uso de inteligencia artificial generativa ..... 103

## Índice de Figuras

Figura 1. Ranking completo de vulnerabilidad socioambiental ..... p. 54
Figura 2. Distribución territorial del índice de vulnerabilidad socioambiental ..... p. 55
Figura 3. Perfil dimensional de los cantones ordenados por el índice global ..... p. 60
Figura 4. Distribución territorial de los seis subíndices de vulnerabilidad ..... p. 61
Figura 5. Dimensión con mayor puntuación en cada cantón ..... p. 62
Figura 6. Composición dimensional de los diez cantones con mayor índice ..... p. 64
Figura 7. Configuración dimensional media de los perfiles territoriales ..... p. 66
Figura 8. Distribución territorial de los perfiles identificados mediante clustering ..... p. 67

Figura 9. Cantones con mayores cambios entre métodos de ponderación .....p. 69
Figura 10. Distribución territorial de la sensibilidad metodológica .....p. 71
Figura 11. Asociación espacial local del índice de vulnerabilidad socioambiental .....p. 73

## Índice de Tablas

Tabla 1. Matriz simplificada de fuentes y uso final .....p. 36
Tabla 2. Estructura del data mart territorial .....p. 39
Tabla 3. Indicadores del modelo principal .....p. 43
Tabla 4. Ranking completo del índice de vulnerabilidad socioambiental .....p. 51
Tabla 5. Estadísticos descriptivos de los seis subíndices .....p. 56
Tabla 6. Puntuaciones más altas por dimensión .....p. 58
Tabla 7. Promedios dimensionales de los perfiles territoriales .....p. 65
Tabla 8. Pesos dimensionales comparados .....p. 68
Tabla 9. Cantones con sensibilidad metodológica alta .....p. 70
Tabla 10. Asociaciones espaciales locales significativas .....p. 72
Tabla 11. Correspondencia entre objetivos específicos, procedimiento, evidencia y conclusión.....p. 80

## Índice de Anexos

Anexo A. Inventario y arquitectura del data mart territorial.....p. 78
Anexo B. Matriz de fuentes y archivos definitivos.....p. 79
Anexo C. Diccionario maestro de variables.....p. 79
Anexo D. Catálogo y evaluación de indicadores.....p. 80
Anexo E. Resultados detallados del PCA.....p. 80
Anexo F. Resultados detallados del clustering.....p. 81
Anexo G. Interpretabilidad y sensibilidad por cantón.....p. 82
Anexo H. Productos cartográficos y visualizaciones complementarias.....p. 82
Anexo I. Controles de calidad y auditoría geoespacial.....p. 83
Anexo J. Código y reproducibilidad.....p. 84
Anexo K. Declaración y bitácora de uso de inteligencia artificial generativa.....p. 85

# 1. Capítulo 1. Introducción

## 1.1. Introducción

La Amazonía ecuatoriana constituye un territorio de alta relevancia ecológica, cultural, social y política. En sus seis provincias convergen extensas áreas de bosque tropical, zonas de conservación, pueblos y nacionalidades indígenas, asentamientos rurales y urbanos, actividades extractivas, transformaciones de la cobertura del suelo y desigualdades persistentes en el acceso a servicios.

Estas condiciones no se distribuyen homogéneamente y tampoco pueden ser comprendidas mediante una única variable. La vulnerabilidad socioambiental se configura por la interacción entre exposición a presiones ambientales, condiciones sociodemográficas, disponibilidad de servicios, accesibilidad territorial y capacidades institucionales diferenciadas para responder a esas presiones (Adger, 2006; Cutter et al., 2003).

Esta complejidad plantea un desafío para las instituciones públicas, académicas, sociales, educativas y eclesiales que intervienen en la Amazonía. Aunque existe una cantidad creciente de datos censales, administrativos y geoespaciales, la información suele encontrarse dispersa entre diversas fuentes, períodos, formatos y escalas territoriales. La existencia de datos abiertos no garantiza por sí misma su transformación en conocimiento útil: es necesario evaluar su calidad, homologar sus unidades territoriales, reconstruir sus metadatos, documentar las transformaciones y desarrollar métodos que permitan analizar conjuntamente dimensiones que tradicionalmente se presentan de forma separada (Banco Mundial, 2021; Kitchin, 2014).

Frente a este problema, el presente Proyecto de Titulación tuvo como objetivo construir un índice cantonal de vulnerabilidad socioambiental para la Amazonía ecuatoriana mediante la integración de datos territoriales y la aplicación de técnicas de ciencia de datos, análisis geoespacial y aprendizaje automático no supervisado.

El índice fue concebido como una herramienta de apoyo para la priorización de intervenciones educativas, pastorales y de gestión del conocimiento. Su finalidad no es automatizar decisiones ni sustituir el conocimiento territorial de los actores locales, sino organizar información heterogénea, identificar patrones comparables y aportar evidencia para procesos deliberativos de planificación y asignación de recursos.

La investigación adoptó un enfoque cuantitativo, aplicado, exploratorio y no experimental. La unidad definitiva de análisis fue el cantón, debido a que esta escala ofreció la mayor compatibilidad entre las fuentes prioritarias. Se construyó un data mart territorial que preservó los datos originales, los productos procesados, la base integrada, los indicadores, los resultados de modelado, las visualizaciones y los cuadernos de código. Esta arquitectura permitió mantener la trazabilidad desde cada fuente hasta los productos analíticos finales y documentar decisiones relacionadas con valores faltantes, homologación territorial, selección de variables y agregación geoespacial.

El marco territorial incluyó los 42 cantones correspondientes a la división político-administrativa vigente. Sin embargo, el índice pudo calcularse para 40. Morona y Sevilla Don Bosco permanecieron visibles en la base y en los productos cartográficos, pero no recibieron puntuación debido a la incompatibilidad estructural entre el Censo de Población y Vivienda 2022 y la configuración territorial posterior. En lugar de redistribuir o imputar información no observada, se preservaron ambos casos como faltantes estructurales, priorizando la transparencia sobre la aparente completitud del modelo.

A partir de 34 indicadores candidatos se seleccionaron 13 para el modelo principal. Estos se organizaron en seis dimensiones: sociodemográfica, educativa, sanitaria, accesibilidad territorial, ambiental y capacidad institucional. Las variables fueron orientadas para que los valores más altos representaran mayor vulnerabilidad, transformadas cuando presentaron asimetrías pronunciadas y normalizadas en una escala común.

El índice principal utilizó ponderación equilibrada entre dimensiones, mientras que el análisis de componentes principales se empleó como procedimiento exploratorio y como alternativa de sensibilidad, no como sustituto del marco conceptual. Esta decisión responde a la necesidad de distinguir entre el peso estadístico derivado de la varianza y la relevancia sustantiva de las dimensiones en un índice compuesto (Organisation for Economic Co-operation and Development & European Commission, Joint Research Centre, 2008; Saltelli, 2007).

El aprendizaje automático no supervisado se utilizó de manera complementaria al índice. Mientras el índice produjo un ordenamiento relativo de los cantones, el clustering permitió identificar configuraciones multidimensionales semejantes. Se compararon representaciones basadas en los indicadores estandarizados y en componentes principales, así como diferentes algoritmos y números de agrupamientos. La solución seleccionada distinguió tres perfiles territoriales de tamaños equilibrados, evaluados mediante métricas internas, estabilidad por submuestreo e interpretabilidad sustantiva. Los perfiles no fueron convertidos en categorías ordinales de mayor o menor vulnerabilidad, porque describen combinaciones diferentes de condiciones territoriales.

Los resultados mostraron una diferenciación importante entre los cantones. Logroño, Taisha y Aguarico ocuparon las primeras posiciones del índice, aunque sus puntuaciones respondieron a configuraciones distintas. La dimensión sanitaria presentó el promedio más alto y un efecto de techo que produjo numerosos empates; la accesibilidad territorial fue especialmente relevante en territorios extensos y con menor articulación vial; y la dimensión ambiental mostró un patrón parcialmente independiente de las brechas sociales y de servicios.

El análisis espacial registró una autocorrelación positiva moderada y permitió identificar una concentración Alto-Alto en el centro-sur de Morona Santiago. Al mismo tiempo, las pruebas de sensibilidad evidenciaron que la posición de algunos cantones varía

cuando se modifican los pesos o se excluyen dimensiones, por lo que el ranking debe comunicarse junto con su nivel de robustez.

El estudio incorporó también consideraciones éticas e interculturales. La presencia de población indígena, pueblos y nacionalidades o territorios indígenas no fue tratada como una carencia ni como un indicador negativo. Estas variables se conservaron para contextualizar los resultados y orientar su interpretación, en concordancia con los principios CARE para la gobernanza responsable de datos indígenas (Carroll et al., 2020). Esta decisión reconoce que la vulnerabilidad no deriva de la identidad cultural, sino de relaciones históricas, desigualdades territoriales, presiones externas y capacidades diferenciadas de respuesta.

El aporte del proyecto es simultáneamente metodológico, tecnológico y aplicado. En el plano metodológico, integra fuentes heterogéneas dentro de una estructura documentada y reproducible; en el plano analítico, combina un índice compuesto, aprendizaje no supervisado, sensibilidad y análisis espacial; en el plano tecnológico, genera bases, cuadernos, mapas y un visualizador territorial; y en el plano aplicado, propone una herramienta para apoyar decisiones institucionales sin reducir la complejidad amazónica a una puntuación única. La prioridad de intervención debe considerar conjuntamente el índice global, los subíndices, el perfil territorial, la sensibilidad metodológica y el conocimiento situado.

El documento se organiza en tres capítulos. El primero presenta los objetivos, la justificación, las preguntas de investigación y los fundamentos del estudio. El segundo desarrolla el marco teórico y conceptual, la metodología, la arquitectura del data mart, la construcción de indicadores, las técnicas de modelado y el análisis de resultados. El tercer capítulo sintetiza las conclusiones, expone las limitaciones técnicas y metodológicas y formula recomendaciones para el uso, actualización y fortalecimiento futuro del índice.

### 1.2. Objetivos

### 1.2.1. Objetivo general

Construir un índice cantonal de vulnerabilidad socioambiental para la Amazonía ecuatoriana, mediante la integración de datos territoriales y técnicas de ciencia de datos y aprendizaje automático no supervisado, como herramienta de priorización para intervenciones educativas, pastorales y sociales.

### 1.2.2. Objetivos específicos

a. Construir una base territorial cantonal consolidada de la Amazonía ecuatoriana mediante la limpieza, homologación e integración de fuentes abiertas y registros administrativos.

b. Construir un índice cantonal de vulnerabilidad socioambiental a partir de indicadores normalizados y subíndices dimensionales, e identificar perfiles territoriales mediante reducción de dimensionalidad y aprendizaje automático no supervisado.

c. Evaluar la consistencia, estabilidad e interpretabilidad del índice y de los perfiles territoriales mediante métricas de validación, análisis de sensibilidad, descomposición de contribuciones y visualización geoespacial.

## 1.3. Justificación

La Amazonía ecuatoriana constituye un territorio estratégico por su diversidad biológica, cultural, lingüística y socioambiental. En esta región convergen pueblos y nacionalidades indígenas, áreas de alta riqueza ecológica, zonas de conservación, presiones extractivas, desigualdades históricas y brechas persistentes en el acceso a educación, salud, conectividad e infraestructura. Esta complejidad exige herramientas analíticas que permitan comprender el territorio de manera integrada y apoyar procesos de toma de decisiones basados en evidencia (Gudynas, 2011; Red Amazónica de Información Socioambiental Georreferenciada [RAISG], 2020; United Nations Office for Disaster Risk Reduction, 2022).

La vulnerabilidad socioambiental en territorios amazónicos no puede explicarse únicamente por la presencia de amenazas ambientales ni por indicadores sociales aislados. Se configura a partir de la interacción entre exposición a presiones ambientales, sensibilidad social, desigualdades territoriales y capacidades diferenciadas de respuesta institucional o comunitaria. Por ello, analizar la vulnerabilidad requiere integrar dimensiones sociodemográficas, educativas, sanitarias, de accesibilidad territorial, capacidad institucional y transformación ambiental en una lectura territorial articulada (Cutter et al., 2003; Comisión Económica para América Latina y el Caribe [CEPAL], 2022).

En el caso de la Amazonía ecuatoriana, esta necesidad es especialmente relevante porque las instituciones que trabajan en educación, acción pastoral, investigación aplicada, intervención social, gestión del conocimiento e incidencia sociopolítica requieren criterios claros para priorizar territorios. La toma de decisiones institucionales suele depender de información fragmentada, experiencia acumulada por los equipos o demandas emergentes desde los territorios. Aunque estos insumos son valiosos, pueden fortalecerse mediante herramientas de análisis que permitan comparar unidades territoriales, identificar patrones de vulnerabilidad y orientar de manera más sistemática la asignación de recursos y acciones (Banco Mundial, 2021; Provost & Fawcett, 2013).

El proyecto se justifica, en primer lugar, por su pertinencia territorial. La Amazonía ecuatoriana presenta diferencias entre provincias y cantones en términos de acceso a servicios, capacidad institucional, accesibilidad territorial y transformación ambiental. El estudio concentró su análisis en la escala cantonal, debido a que esta fue la unidad con mayor compatibilidad entre las fuentes integradas. La deforestación, la pérdida de cobertura boscosa y el cambio de uso del suelo son procesos que modifican las condiciones ecológicas y sociales de los territorios, por lo que constituyen variables relevantes para comprender

dinámicas de vulnerabilidad socioambiental (Food and Agriculture Organization of the United Nations [FAO], 2020; MapBiomas Ecuador, 2024).

En segundo lugar, el estudio se justifica por su relevancia institucional. Las organizaciones educativas, pastorales, eclesiales, sociales y académicas que trabajan en la Amazonía necesitan instrumentos que les permitan focalizar programas, diseñar intervenciones contextualizadas, fortalecer procesos de acompañamiento territorial y sustentar decisiones ante actores internos y externos. Un índice de vulnerabilidad socioambiental puede aportar una base analítica para priorizar acciones de formación, investigación aplicada, salud territorial, incidencia y gestión del conocimiento.

En tercer lugar, el proyecto tiene una justificación metodológica. La información sobre la Amazonía se encuentra dispersa en diferentes fuentes, como censos nacionales, registros administrativos, plataformas geoespaciales, datos satelitales y bases sectoriales. La integración de estas fuentes en una matriz territorial consolidada permitió superar una lectura fragmentada de los indicadores y avanzar hacia un modelo reproducible de análisis territorial.

Para ello, los sistemas de información geográfica ofrecen procedimientos que permiten relacionar atributos estadísticos con unidades espaciales, armonizar distintas representaciones del territorio y ejecutar operaciones de superposición, unión, medición y agregación espacial. Estas capacidades resultan fundamentales para convertir fuentes heterogéneas en información territorial comparable y analíticamente consistente (Kitchin, 2014; Longley et al., 2015).

El estudio integró información censal, registros administrativos sectoriales y datos geoespaciales procedentes de instituciones públicas y plataformas especializadas. Las fuentes aportaron información sobre población, educación, salud, infraestructura vial, capacidad institucional, cobertura y uso del suelo, actividades extractivas, territorios indígenas y otros atributos contextuales de los cantones amazónicos. Estas fuentes permitieron construir

variables sociodemográficas, educativas, sanitarias, de accesibilidad territorial, capacidad institucional y transformación ambiental para caracterizar los territorios amazónicos del Ecuador. Las variables interculturales se utilizaron para contextualizar e interpretar los resultados, sin incorporarlas como indicadores directos de vulnerabilidad.

La integración de estos datos responde a la necesidad de transformar información disponible en conocimiento accionable para la toma de decisiones (Instituto Nacional de Estadística y Censos, 2024; Ministerio de Educación, 2025; Ministerio de Salud Pública, 2026; Agencia de Aseguramiento de la Calidad de los Servicios de Salud y Medicina Prepagada, 2026; MapBiomas Ecuador, 2024; Red Amazónica de Información Socioambiental Georreferenciada, s.f.).

En cuarto lugar, el proyecto se justifica por su aporte tecnológico y analítico. La aplicación de técnicas de ciencia de datos, análisis geoespacial y aprendizaje automático permite construir indicadores, normalizar variables, reducir dimensionalidad, identificar patrones territoriales y clasificar unidades de análisis según niveles relativos de vulnerabilidad. Técnicas como el análisis de componentes principales y el clustering aportan herramientas para explorar la estructura interna de los datos y reducir la arbitrariedad en la construcción del índice (Jolliffe & Cadima, 2016; Pedregosa et al., 2011).

La contribución metodológica del estudio consistió en integrar fuentes heterogéneas, construir un índice compuesto por dimensiones y complementar su evaluación con técnicas estadísticas, aprendizaje automático no supervisado y análisis espacial. Esta aproximación permite ir más allá de una descripción aislada de indicadores y avanzar hacia una lectura integrada de perfiles territoriales. El índice no solo permite ordenar territorios según niveles de vulnerabilidad, sino también identificar qué variables tienen mayor incidencia en la configuración de dichos niveles.

Desde el punto de vista académico, el proyecto articula componentes centrales de la ciencia de datos aplicada: integración de bases de datos abiertas, procesamiento de datos estructurados y geoespaciales, construcción de indicadores, modelado mediante aprendizaje automático, evaluación de resultados, interpretabilidad y visualización. Esta articulación permite conectar el campo de los sistemas de información y la ciencia de datos con problemas concretos de gestión territorial, sostenibilidad y toma de decisiones institucionales.

El proyecto también se justifica por su potencial de visualización y comunicación. La construcción de mapas temáticos, gráficos y un prototipo de visualizador territorial permitió presentar los resultados de manera comprensible para usuarios institucionales. Esta dimensión es importante porque la utilidad de un modelo de ciencia de datos no depende únicamente de su consistencia técnica, sino también de su capacidad para comunicar hallazgos y apoyar procesos deliberativos.

Asimismo, el estudio incorpora una justificación ética e intercultural. La presencia de pueblos y nacionalidades indígenas no fue tratada como sinónimo de vulnerabilidad ni como indicador negativo. Fue incorporada como una variable de pertinencia territorial e intercultural, considerando principios de gobernanza responsable de datos indígenas y evitando usos estigmatizantes o extractivos de la información (Carroll et al., 2020; Walsh, 2009).

Esta precaución es fundamental porque los datos territoriales pueden producir efectos no deseados si son interpretados fuera de contexto. Aunque el índice busca apoyar decisiones institucionales, no pretende sustituir el conocimiento comunitario, la deliberación territorial ni las lecturas cualitativas de los actores locales. Su finalidad es ofrecer una herramienta complementaria que permita identificar patrones, contrastar información y fortalecer procesos de priorización basados en evidencia.

En consecuencia, este Proyecto de Titulación es pertinente porque responde a una necesidad real de integración y análisis de información territorial amazónica. Su aporte consiste en transformar datos abiertos y registros dispersos en un modelo analítico reproducible, capaz de generar un índice de vulnerabilidad socioambiental y una clasificación territorial útil para la toma de decisiones. De esta manera, el proyecto contribuye tanto al campo académico de la ciencia de datos aplicada como a los procesos institucionales orientados a la educación, la acción pastoral, la gestión del conocimiento y la acción socioambiental en la Amazonía ecuatoriana.

### 1.4. Preguntas de investigación

1. ¿Qué indicadores sociodemográficos, educativos, sanitarios, de accesibilidad territorial, capacidad institucional y transformación ambiental pueden integrarse para caracterizar la vulnerabilidad socioambiental de los cantones de la Amazonía ecuatoriana?
2. ¿Cómo puede construirse un índice cantonal de vulnerabilidad socioambiental a partir de la integración, transformación, normalización y agregación de fuentes abiertas heterogéneas?
3. ¿Qué perfiles de cantones amazónicos pueden identificarse mediante técnicas de reducción de dimensionalidad y aprendizaje automático no supervisado, y qué nivel de consistencia presentan estos agrupamientos?
4. ¿Qué dimensiones e indicadores contribuyen en mayor medida a la puntuación del índice y cómo se distribuyen territorialmente los diferentes niveles de vulnerabilidad socioambiental?

# 2. Capítulo 2. Desarrollo

## 2.1. Marco teórico y conceptual

### Antecedentes y marco referencial

El análisis de problemas territoriales complejos ha adquirido una importancia creciente en el campo de la ciencia de datos, especialmente cuando se trata de estudiar regiones donde convergen desigualdades sociales, presiones ambientales, brechas de acceso a servicios y procesos históricos de exclusión. En este contexto, los datos abiertos, los sistemas de información geográfica y las técnicas de aprendizaje automático ofrecen herramientas metodológicas para integrar información dispersa, identificar patrones y generar insumos útiles para la toma de decisiones (Kitchin, 2014; Provost & Fawcett, 2013).

En territorios amazónicos, esta aproximación resulta particularmente relevante. La Amazonía no puede ser comprendida únicamente como un espacio natural o como una unidad administrativa. Se trata de un territorio socioambiental en el que interactúan biodiversidad, pueblos indígenas, actividades extractivas, disputas por el uso del suelo, desigualdades históricas, limitaciones de conectividad, brechas educativas y dificultades de acceso a servicios básicos (Gudynas, 2011; RAISG, 2020).

El uso de información territorial integrada permite pasar de diagnósticos fragmentados a modelos analíticos más consistentes. La información proveniente de censos, registros administrativos, plataformas geoespaciales, imágenes satelitales y sistemas de monitoreo ambiental puede ser procesada para caracterizar condiciones de vulnerabilidad diferenciadas entre unidades territoriales. Esta integración es especialmente útil cuando las instituciones requieren priorizar intervenciones, asignar recursos, diseñar programas o monitorear transformaciones territoriales (Banco Mundial, 2021).

A nivel internacional, los índices compuestos han sido utilizados para sintetizar fenómenos multidimensionales como pobreza, desarrollo humano, vulnerabilidad, riesgo,

desigualdad territorial o exposición ambiental. Su principal utilidad consiste en transformar múltiples variables en una medida comparable que permita ordenar unidades de análisis, identificar brechas y apoyar procesos de priorización (Programa de las Naciones Unidas para el Desarrollo, 2020; United Nations Office for Disaster Risk Reduction, 2022).

Sin embargo, los índices compuestos también presentan desafíos metodológicos. La selección de variables, su normalización, la asignación de pesos y la definición de umbrales pueden depender de decisiones expertas que introducen subjetividad. Por ello, la incorporación de técnicas estadísticas y de aprendizaje automático puede contribuir a reducir la arbitrariedad, explorar patrones internos en los datos y evaluar la consistencia de las clasificaciones territoriales (Jolliffe & Cadima, 2016; Pedregosa et al., 2011).

En América Latina existen diversas iniciativas de producción y visualización de datos socioambientales que resultan relevantes para estudios territoriales. MapBiomas Ecuador ha desarrollado series históricas de cobertura y uso del suelo que permiten analizar procesos de deforestación, transformación de coberturas y cambios ambientales en la región amazónica (MapBiomas Ecuador, 2024). Por su parte, RAISG ha producido información geoespacial sobre territorios indígenas, áreas protegidas, infraestructura y presiones socioambientales en la Amazonía (RAISG, 2020).

En el caso ecuatoriano, el Censo de Población y Vivienda 2022 del Instituto Nacional de Estadística y Censos (INEC) constituye una fuente central para caracterizar población, vivienda, ruralidad, conectividad, educación y autoidentificación étnica (Instituto Nacional de Estadística y Censos, 2024). A ello se suman registros administrativos del Ministerio de Educación, información de establecimientos de salud del Ministerio de Salud Pública y ACESS, y fuentes ambientales como MapBiomas Ecuador y RAISG. La combinación de estas fuentes permite construir una mirada multidimensional de los territorios amazónicos del Ecuador.

## Antecedentes empíricos de la construcción de índices territoriales de vulnerabilidad

Los estudios empíricos sobre vulnerabilidad territorial han empleado índices compuestos para integrar variables sociales, económicas, ambientales e institucionales en medidas comparables. Aunque difieren en sus escalas, dimensiones y métodos de ponderación, estos trabajos muestran la utilidad de combinar información estadística y geográfica para identificar desigualdades territoriales, establecer prioridades de intervención y examinar los factores asociados con la vulnerabilidad.

Uno de los principales antecedentes es el Índice de Vulnerabilidad Social —Social Vulnerability Index, SoVI— desarrollado por Cutter et al. (2003) para los condados de Estados Unidos. El estudio partió de 42 variables socioeconómicas, demográficas y relacionadas con el entorno construido, que fueron reducidas mediante análisis factorial a 11 componentes independientes. Estos componentes explicaron aproximadamente el 76 % de la varianza y fueron agregados en una puntuación sintética de vulnerabilidad social. El trabajo estableció una referencia metodológica para el uso de técnicas multivariadas en la construcción de índices territoriales, aunque su aplicación directa a la Amazonía ecuatoriana estaría limitada por las diferencias sociales, institucionales y ambientales entre ambos contextos.

En Ecuador, Fernandez et al. (2015) construyeron un indicador compuesto de vulnerabilidad frente al cambio climático a escala cantonal. El modelo integró variables climáticas y socioeconómicas relacionadas con exposición, sensibilidad y capacidad adaptativa. Los autores utilizaron el comportamiento estadístico de los indicadores para determinar su ponderación y asignaron pesos inversamente proporcionales a su varianza entre cantones. El estudio también señaló que la localización, armonización y disponibilidad de datos de calidad constituyen dificultades centrales para las evaluaciones locales. Sus

resultados identificaron diferencias importantes entre cantones y mostraron que la vulnerabilidad no dependía únicamente del tamaño poblacional, sino de la combinación de condiciones sociales, climáticas e institucionales (Fernandez et al., 2015).

Este antecedente resulta especialmente relevante para el presente proyecto porque utilizó el cantón como unidad de análisis y construyó una medición multidimensional para el territorio ecuatoriano. Sin embargo, se concentró en la vulnerabilidad frente al cambio climático y no integró de manera específica variables sobre oferta educativa, disponibilidad de servicios sanitarios, accesibilidad vial, capacidad operativa municipal, actividades extractivas y transformación de la cobertura del suelo. Tampoco tuvo como propósito organizar las fuentes mediante un repositorio analítico que preservara la trazabilidad completa de los procesos de limpieza, homologación e integración.

En el contexto amazónico, Mansur et al. (2016) evaluaron la vulnerabilidad urbana en 41 ciudades del delta y estuario del Amazonas, utilizando información correspondiente a 2.938 sectores censales. El índice integró tres dimensiones: exposición a inundaciones, sensibilidad socioeconómica y condiciones de infraestructura. Entre los indicadores se incluyeron servicios de saneamiento, características de la vivienda, ingreso de los hogares, población, riesgo de inundación y asentamientos no planificados. Los resultados mostraron que entre el 60 % y el 90 % de la población urbana analizada vivía en condiciones de vulnerabilidad moderada o alta, con diferencias tanto entre ciudades como dentro de ellas (Mansur et al., 2016).

El trabajo de Mansur et al. (2016) evidenció la importancia de integrar exposición ambiental, condiciones sociales e infraestructura en el análisis de territorios amazónicos. También mostró que una puntuación agregada puede ocultar desigualdades internas cuando se utiliza una unidad territorial extensa. No obstante, su objeto de estudio estuvo delimitado a espacios urbanos y al riesgo de inundación, por lo que no abordó integralmente otras

transformaciones socioambientales, como la pérdida acumulada de bosque, los cambios recientes de cobertura, la minería o la actividad petrolera.

La aplicación del aprendizaje automático al estudio de la vulnerabilidad fue abordada por Abarca-Alvarez et al. (2019) en la región española de Andalucía. Los autores utilizaron datos censales y demográficos para identificar perfiles territoriales mediante mapas autoorganizados —Self-Organizing Maps—, una técnica de aprendizaje no supervisado. Posteriormente, construyeron un modelo basado en árboles de inferencia condicional para reconocer el perfil asociado con mayores condiciones de vulnerabilidad social. La combinación de aprendizaje automático y sistemas de información geográfica permitió localizar áreas con características semejantes y relacionar los perfiles demográficos con las condiciones de las viviendas (Abarca-Alvarez et al., 2019).

Este estudio mostró que el aprendizaje no supervisado puede complementar los índices agregados mediante la identificación de perfiles que no necesariamente siguen un orden lineal de menor a mayor vulnerabilidad. Los agrupamientos permiten reconocer configuraciones diferenciadas de los indicadores, mientras que el índice facilita la comparación ordinal entre unidades territoriales. Esta distinción es relevante porque un clúster representa similitud multivariada entre territorios, pero no constituye automáticamente una categoría de vulnerabilidad baja, media, alta o muy alta.

En conjunto, los antecedentes examinados muestran tres aproximaciones complementarias: la reducción de variables para construir índices sintéticos, la integración de dimensiones sociales y ambientales en unidades territoriales, y la aplicación de aprendizaje no supervisado para identificar perfiles. También evidencian que los resultados dependen de la escala seleccionada, la calidad de las fuentes, la definición de las dimensiones, el método de normalización y las reglas de ponderación.

El presente Proyecto de Titulación articuló estos aportes en un modelo analítico aplicado a los cantones de la Amazonía ecuatoriana. El estudio integró indicadores sociodemográficos, educativos, sanitarios, de accesibilidad territorial, capacidad institucional y transformación ambiental. Además, utilizó variables interculturales para contextualizar los resultados, sin convertir la identidad o la territorialidad indígena en indicadores negativos de vulnerabilidad. La construcción del índice se complementó con análisis de componentes principales, aprendizaje automático no supervisado, evaluación de estabilidad, análisis de sensibilidad y representación geoespacial.

El aporte específico consistió en combinar la construcción de un índice compuesto con la identificación de perfiles cantonales, manteniendo una separación metodológica entre ambos productos. La información procedente de fuentes estadísticas, administrativas y geoespaciales fue organizada en un data mart territorial documentado, lo que permitió conservar la procedencia de las variables y reconstruir las transformaciones realizadas. De esta manera, el estudio atendió tanto la dimensión analítica del problema como los requisitos de calidad, trazabilidad y reproducibilidad propios de un proyecto de ciencia de datos.

**Vulnerabilidad socioambiental**

El concepto de vulnerabilidad socioambiental permite analizar la relación entre condiciones sociales y presiones ambientales. Desde esta perspectiva, un territorio no es vulnerable únicamente porque esté expuesto a amenazas ambientales, sino porque esa exposición interactúa con desigualdades sociales, capacidades institucionales limitadas, brechas de acceso a servicios y condiciones históricas de exclusión (Cutter et al., 2003).

Cutter et al. (2003) plantean que la vulnerabilidad social frente a amenazas ambientales depende de factores como condiciones demográficas, acceso a recursos, infraestructura, capacidades de respuesta y distribución desigual de riesgos. Esta aproximación resulta pertinente para territorios amazónicos, donde la vulnerabilidad no se

expresa de manera aislada, sino mediante la combinación de múltiples factores sociales, ambientales e institucionales.

Desde una perspectiva socioambiental, la vulnerabilidad puede entenderse como una condición relacional. No se reduce a la presencia de pobreza, deforestación o aislamiento geográfico, sino a la interacción entre esas dimensiones. Así, un cantón amazónico puede presentar alta vulnerabilidad si combina baja conectividad, limitada disponibilidad de servicios de salud, rezagos educativos, alta ruralidad, transformación acelerada de cobertura boscosa y débil capacidad institucional de respuesta.

La CEPAL (2022) ha señalado que la desigualdad territorial constituye una dimensión clave de las brechas estructurales en América Latina. Esta desigualdad se expresa en diferencias de acceso a servicios, infraestructura, oportunidades educativas, conectividad y capacidades institucionales. En territorios amazónicos, estas brechas suelen intensificarse por la dispersión poblacional, las dificultades de conectividad física y digital, y la presencia de dinámicas extractivas que transforman el territorio.

La vulnerabilidad socioambiental también debe ser entendida en relación con la exposición a procesos de transformación ambiental. En este estudio, la deforestación, la pérdida de cobertura boscosa y el cambio de uso del suelo se consideran variables centrales porque expresan modificaciones estructurales en el territorio. Estas variables permiten observar no solo la pérdida de bosque, sino también la transición hacia otros usos, como áreas agropecuarias, infraestructura o zonas intervenidas (FAO, 2020; MapBiomas Ecuador, 2024).

Esta aproximación permite superar una lectura exclusivamente ambiental o exclusivamente social de la vulnerabilidad. La vulnerabilidad socioambiental se produce cuando las presiones sobre el territorio se combinan con condiciones sociales que limitan la capacidad de respuesta, adaptación o protección de las poblaciones.

**Análisis territorial y Amazonía ecuatoriana**

El análisis territorial parte de la premisa de que los fenómenos sociales, ambientales y económicos no se distribuyen de forma homogénea en el espacio. Las desigualdades tienen expresión territorial, por lo que deben analizarse considerando escalas, localización, relaciones espaciales, concentración de servicios, accesibilidad y patrones de transformación del entorno (Harvey, 2006).

En este proyecto, el territorio se entiende como una construcción socioespacial. No se trata solamente de una superficie delimitada administrativamente, sino de un espacio donde interactúan población, instituciones, ecosistemas, infraestructura, relaciones económicas, dinámicas culturales y procesos históricos. Esta comprensión es especialmente importante para el caso amazónico, donde los límites administrativos no siempre capturan la complejidad ecológica, cultural y política del territorio.

La Amazonía ecuatoriana comprende las provincias de Sucumbíos, Orellana, Napo, Pastaza, Morona Santiago y Zamora Chinchipe. Esta región se caracteriza por una alta diversidad ecológica, presencia de pueblos y nacionalidades indígenas, amplias áreas de bosque tropical, zonas de conservación, dinámicas extractivas, desigualdades de acceso a servicios y dificultades de conectividad física y digital (RAISG, 2020; Instituto Nacional de Estadística y Censos, 2024).

Desde una mirada socioambiental, la Amazonía ecuatoriana debe ser comprendida como un territorio de interdependencias. Las condiciones ecológicas están vinculadas con formas de vida, dinámicas comunitarias, economías locales, derechos colectivos, relaciones interculturales y disputas por el uso del suelo. Por ello, el análisis de vulnerabilidad no puede limitarse a medir carencias sociales ni únicamente a registrar pérdidas ambientales.

Gudynas (2011) plantea que las discusiones sobre desarrollo, ambiente y buen vivir en América Latina obligan a cuestionar enfoques que separan artificialmente sociedad y

naturaleza. En territorios amazónicos, esta separación resulta especialmente problemática, porque las condiciones de vida de las comunidades, la conservación de los ecosistemas y las presiones extractivas forman parte de una misma trama territorial.

El enfoque territorial permite justificar la elección de unidades de análisis como cantones o parroquias. Estas unidades permiten integrar información estadística, administrativa y geoespacial de manera comparable. No obstante, también implican límites metodológicos, porque las dinámicas socioambientales no siempre se ajustan a divisiones político-administrativas.

En este Proyecto de Titulación, el análisis se realizó inicialmente a nivel cantonal. La posibilidad de trabajar a escala parroquial fue examinada durante la preparación de los datos, pero se descartó debido a la falta de compatibilidad, cobertura y denominadores consistentes en ese nivel.

**Datos abiertos, sistemas de información geográfica y ciencia de datos aplicada**

Los datos abiertos constituyen una base fundamental para el desarrollo de modelos analíticos aplicados a la toma de decisiones públicas, sociales e institucionales. El Banco Mundial (2021) señala que los datos pueden contribuir al desarrollo cuando son accesibles, reutilizables, confiables y transformados en información útil para orientar decisiones. Sin embargo, la disponibilidad de datos no garantiza automáticamente su aprovechamiento.

Kitchin (2014) advierte que los datos no son neutrales ni autosuficientes. Su uso requiere procesos de limpieza, contextualización, interpretación y validación. En estudios territoriales, esta advertencia es clave, porque las bases abiertas suelen provenir de instituciones distintas, con criterios de producción, escalas, formatos y años de actualización diferentes.

La ciencia de datos aplicada ofrece un conjunto de procedimientos para transformar información dispersa en conocimiento accionable. Provost y Fawcett (2013) sostienen que la

ciencia de datos no se limita a usar algoritmos, sino que implica formular problemas, seleccionar variables, preparar datos, construir modelos, evaluar resultados y comunicar hallazgos de manera útil para la toma de decisiones.

En este estudio, la ciencia de datos se utilizó para integrar fuentes heterogéneas, construir indicadores, normalizar variables, modelar patrones territoriales y generar visualizaciones. El proceso no se reduce a una operación técnica, sino que implica decisiones metodológicas sobre qué variables incluir, cómo tratarlas, cómo ponderarlas y cómo interpretar sus resultados.

Los sistemas de información geográfica aportaron una dimensión central al proceso de integración territorial. Un SIG permite vincular datos alfanuméricos con localizaciones y geometrías, representar entidades espaciales y analizar sus relaciones de posición, proximidad, superposición y pertenencia territorial. En este estudio, estas capacidades permitieron relacionar los límites cantonales con información sobre cobertura y uso del suelo, territorios indígenas, áreas protegidas, infraestructura vial, actividades extractivas, establecimientos de servicios y otras presiones socioambientales (Longley et al., 2015).

La integración de datos estadísticos, administrativos y geoespaciales permitió construir una matriz territorial consolidada a escala cantonal. Su elaboración requirió procesos de extracción, transformación y carga, homologación de códigos territoriales, control de sistemas de referencia, uniones espaciales, recortes, intersecciones, cálculo de superficies y agregación de resultados por cantón. Esta matriz constituyó la base para la construcción de indicadores, el cálculo del índice de vulnerabilidad socioambiental, la identificación de perfiles territoriales y la elaboración de productos cartográficos.

Las fuentes principales fueron el Censo de Población y Vivienda 2022 del INEC, registros del Ministerio de Educación, información del Ministerio de Salud Pública y ACESS, datos de MapBiomas Ecuador y capas geoespaciales de RAISG. Estas fuentes permitieron

construir variables sociodemográficas, educativas, sanitarias, de accesibilidad territorial, de capacidad institucional y ambientales, además de variables digitales e interculturales utilizadas de forma complementaria o contextual (Agencia de Aseguramiento de la Calidad de los Servicios de Salud y Medicina Prepagada, 2026; Instituto Nacional de Estadística y Censos, 2024; MapBiomas Ecuador, 2024; Ministerio de Educación, 2025; Ministerio de Salud Pública, 2026; Red Amazónica de Información Socioambiental Georreferenciada, s.f.).

**Índices compuestos y construcción de indicadores territoriales**

Los índices compuestos permiten sintetizar múltiples variables en una medida única o en un conjunto reducido de dimensiones interpretables. Su utilidad radica en que facilitan la comparación entre unidades territoriales, la identificación de brechas y la priorización de intervenciones. Por esta razón, han sido utilizados en mediciones de desarrollo humano, vulnerabilidad, riesgo, pobreza multidimensional y desigualdad territorial (Programa de las Naciones Unidas para el Desarrollo, 2020; United Nations Office for Disaster Risk Reduction, 2022).

La construcción de un índice implica varias decisiones metodológicas. En primer lugar, se deben seleccionar las dimensiones teóricas que explican el fenómeno. En este estudio, la vulnerabilidad socioambiental se organiza en seis dimensiones: sociodemográfica, educativa, sanitaria, accesibilidad territorial, ambiental y capacidad institucional. Cada dimensión captura un componente específico del fenómeno analizado. El índice se expresa en una escala de 0 a 100, que permite comparar territorios según niveles relativos de vulnerabilidad

En segundo lugar, se deben definir variables observables para cada dimensión. Por ejemplo, la dimensión sociodemográfica puede incluir población total, densidad, ruralidad, estructura etaria y condiciones de vivienda. La dimensión educativa puede considerar matrícula, instituciones educativas o relación estudiantes/docente. La dimensión ambiental

puede incorporar pérdida de cobertura boscosa, cambio de uso del suelo y transformación territorial.

En tercer lugar, se requiere normalizar las variables. Este paso es necesario porque los indicadores pueden encontrarse en escalas distintas. Algunas variables pueden expresarse como porcentajes, otras como conteos absolutos, tasas por población o áreas geográficas. Técnicas como Min-Max Scaling o estandarización Z-score permiten transformar las variables para que sean comparables.

Una de las decisiones más sensibles en la construcción de índices es la ponderación. Asignar pesos iguales a todas las variables puede ser simple, pero no necesariamente refleja la estructura estadística de los datos. Por ello, el análisis de componentes principales puede utilizarse como técnica para reducir dimensionalidad, identificar variables correlacionadas y explorar combinaciones que expliquen mayor proporción de varianza (Jolliffe & Cadima, 2016).

En este trabajo, el índice de vulnerabilidad socioambiental se expresa en una escala de 0 a 100. Esta escala permite comparar territorios y clasificarlos según niveles relativos de vulnerabilidad. No obstante, el índice no debe interpretarse como una medición absoluta ni como una representación total de la realidad territorial, sino como una herramienta analítica de priorización.

**Aprendizaje automático aplicado al análisis socioambiental**

El aprendizaje automático comprende un conjunto de técnicas que permiten identificar patrones, clasificar observaciones o realizar predicciones a partir de datos. En el campo del análisis territorial, estas técnicas pueden utilizarse para agrupar unidades con características similares, detectar estructuras internas en los datos y evaluar la importancia relativa de las variables (Pedregosa et al., 2011).

En este Proyecto de Titulación se priorizó el aprendizaje no supervisado, debido a que no se contó inicialmente con una variable objetivo etiquetada que indique de antemano qué territorios son de vulnerabilidad baja, media, alta o crítica. La clasificación se construyó a partir del índice y de los patrones identificados en los datos, no a partir de una etiqueta externa preexistente.

El clustering es una técnica de aprendizaje no supervisado que permite agrupar observaciones según su similitud interna. En este caso, las observaciones fueron cantones y la similitud se calculó a partir de los indicadores territoriales normalizados. El objetivo fue identificar grupos de territorios con perfiles socioambientales semejantes (MacQueen, 1967; Pedregosa et al., 2011).

K-means es uno de los algoritmos de clustering más utilizados. Su lógica consiste en dividir las observaciones en un número definido de grupos, minimizando la distancia interna entre los elementos de cada grupo. En este estudio, K-means se evaluó como una de las alternativas para identificar conjuntos de cantones con perfiles semejantes de vulnerabilidad, siempre que la estructura de los datos permita definir un número adecuado de clusters.

El clustering jerárquico complementó este análisis porque permite observar relaciones de proximidad entre territorios sin exigir inicialmente una cantidad fija de grupos. Este método puede resultar útil para explorar la estructura de similitudes y diferencias entre cantones. Además, permite generar dendrogramas que facilitan la interpretación de agrupamientos.

DBSCAN se utilizó como diagnóstico exploratorio para examinar posibles agrupamientos por densidad y observaciones atípicas. Su desempeño se interpretó con cautela debido al reducido número de unidades territoriales. La calidad de los agrupamientos puede evaluarse mediante métricas como silhouette score y Davies-Bouldin index. Estas métricas permiten estimar qué tan cohesionados y separados están los grupos. Su uso es importante

porque evita que la clasificación territorial dependa únicamente de una lectura visual o de una decisión arbitraria.

Esta distinción es metodológicamente importante. El estudio no parte de una clasificación previa de vulnerabilidad, por lo que el núcleo del análisis corresponde al aprendizaje no supervisado.

**Interpretabilidad del índice y de los perfiles territoriales para la toma de decisiones**

La interpretabilidad es un componente clave en proyectos de ciencia de datos aplicados a decisiones institucionales. No basta con producir un índice o clasificar territorios; también es necesario explicar por qué determinados territorios aparecen como más vulnerables y qué variables inciden en esa clasificación. Sin esta explicación, el modelo puede convertirse en una "caja negra" difícil de justificar ante actores institucionales o territoriales.

En este trabajo, la interpretabilidad cumplió una función metodológica, comunicativa y ética. Desde el punto de vista metodológico, permitió comprobar la consistencia del índice, reconstruir la puntuación de cada cantón e identificar las dimensiones e indicadores que explican su posición. En el caso del clustering, permitió caracterizar los perfiles mediante sus configuraciones dimensionales, sin convertirlos automáticamente en categorías de vulnerabilidad.

Desde el punto de vista comunicativo, facilitó la traducción de resultados técnicos en información comprensible para usuarios institucionales. Desde el punto de vista ético, contribuyó a prevenir interpretaciones simplificadoras o estigmatizantes de los territorios y de las poblaciones que los habitan.

La utilidad institucional del índice dependerá de su capacidad para traducir datos complejos en información comprensible y accionable. Para organizaciones educativas,

pastorales, sociales o académicas, una herramienta de priorización territorial puede apoyar decisiones sobre focalización de programas, diseño de intervenciones, investigación aplicada, acompañamiento comunitario e incidencia socioambiental.

Sin embargo, la utilidad del índice no debe confundirse con una capacidad de decisión automática. El modelo debe ser entendido como una herramienta de apoyo, no como sustituto del juicio institucional, el diálogo territorial o el conocimiento situado. Su valor radica en ordenar información, hacer visibles patrones y aportar evidencia para procesos deliberativos.

**Consideraciones éticas e interculturales sobre datos territoriales e indígenas**

El uso de datos sobre territorios amazónicos e indígenas exige una reflexión ética específica. La información territorial puede contribuir a la toma de decisiones, pero también puede producir riesgos si se interpreta de manera descontextualizada, extractiva o estigmatizante. Por ello, el análisis de datos debe incorporar criterios de responsabilidad, pertinencia intercultural y cuidado en la comunicación de resultados.

Los principios CARE para la gobernanza de datos indígenas plantean que los datos vinculados con pueblos indígenas deben considerar beneficios colectivos, autoridad para controlar, responsabilidad y ética (Carroll et al., 2020). Estos principios son relevantes para este estudio porque recuerdan que la información sobre pueblos y nacionalidades indígenas no debe ser tratada únicamente como una variable técnica.

En este trabajo, la presencia de población indígena o territorios indígenas no fue interpretada como indicador negativo ni como sinónimo de vulnerabilidad. Se la incorporó como una variable de pertinencia intercultural y territorial. Esta decisión busca evitar que el índice reproduzca lecturas deficitarias sobre los pueblos indígenas o que asocie automáticamente identidad indígena con carencia.

La interculturalidad implica reconocer la coexistencia de diversas culturas, saberes y formas de relación con el territorio. Walsh (2009) entiende la interculturalidad no solo como

convivencia entre culturas, sino como un horizonte crítico frente a relaciones históricas de desigualdad y colonialidad. Desde esta perspectiva, el análisis territorial debe evitar reducir los territorios indígenas a objetos de medición.

El índice de vulnerabilidad socioambiental debe ser interpretado como una herramienta complementaria. No reemplaza la consulta, el diálogo comunitario, la deliberación territorial ni los saberes locales. Su uso debe estar orientado a fortalecer procesos institucionales responsables, no a imponer decisiones externas sobre los territorios.

Además, el trabajo debe evitar la divulgación de información sensible que pueda afectar negativamente a comunidades específicas. Aunque se usen datos abiertos, la integración de múltiples fuentes puede generar nuevas formas de exposición territorial. Por ello, la presentación de resultados debe cuidar el nivel de agregación, el lenguaje utilizado y las implicaciones de la clasificación.

**Marco conceptual operativo**

Para efectos de este Proyecto de Titulación, se emplean los siguientes conceptos operativos:

**Vulnerabilidad socioambiental.** Se entiende como la condición territorial derivada de la interacción entre brechas sociales, limitaciones de acceso a servicios, desigualdades educativas, de accesibilidad, capacidad institucional, sanitaria, deforestación o transformación de cobertura del suelo, presiones ambientales y capacidades diferenciadas de respuesta institucional o comunitaria (CEPAL, 2022; Cutter et al., 2003).

**Territorio amazónico.** Se refiere a las unidades geográficas ubicadas en las provincias amazónicas del Ecuador: Sucumbíos, Orellana, Napo, Pastaza, Morona Santiago y Zamora Chinchipe. Estas unidades se caracterizan por su diversidad ecológica, cultural y social, así como por la presencia de pueblos y nacionalidades indígenas y dinámicas

socioambientales específicas (RAISG, 2020; Red Amazónica de Información Socioambiental Georreferenciada, s.f.).

**Deforestación.** Es la pérdida o disminución de cobertura boscosa en un territorio determinado, identificada mediante datos satelitales o fuentes geoespaciales. En este estudio es considerada una variable ambiental central para observar procesos de transformación territorial (FAO, 2020).

**Cambio de uso del suelo.** Se refiere a la transformación de una cobertura territorial hacia otra, por ejemplo, de bosque a áreas agropecuarias, infraestructura, zonas intervenidas u otras categorías. Esta variable permite observar dinámicas ambientales más amplias que la pérdida de bosque en sentido estricto (MapBiomas Ecuador, 2024).

**Índice de vulnerabilidad socioambiental.** Es una medida sintética que integra variables normalizadas de distintas dimensiones para comparar territorios según niveles relativos de vulnerabilidad. En este trabajo, el índice se construyó en una escala de 0 a 100 y se utilizó como herramienta de priorización territorial de la intervención social, educativa y pastoral (United Nations Office for Disaster Risk Reduction, 2022).

**Priorización territorial.** Es el proceso mediante el cual se identifican territorios que requieren mayor atención institucional, considerando criterios sociodemográficos, educativos, sanitarios, de accesibilidad territorial, ambientales y de capacidad institucional, complementados por información intercultural de carácter contextual. Su finalidad es orientar la asignación de recursos, el diseño de intervenciones y la toma de decisiones basada en evidencia (Banco Mundial, 2021).

**Datos abiertos.** Son datos publicados por instituciones públicas, privadas o sociales que pueden ser utilizados, reutilizados y redistribuidos bajo determinadas condiciones de acceso, atribución o licencia. En este estudio, los datos abiertos fueron la base para construir una matriz territorial integrada (Banco Mundial, 2021; Kitchin, 2014).

**Ciencia de datos aplicada.** Se entiende como el conjunto de procedimientos orientados a transformar datos en conocimiento útil para la toma de decisiones. Incluye la formulación del problema, integración de fuentes, limpieza, transformación, análisis, modelado, evaluación e interpretación de resultados (Provost & Fawcett, 2013).

**Aprendizaje automático.** Es un conjunto de técnicas computacionales que permiten identificar patrones, clasificar observaciones o realizar predicciones a partir de datos. En este trabajo se utilizó principalmente aprendizaje no supervisado para identificar perfiles territoriales de vulnerabilidad (Pedregosa et al., 2011).

**Clustering.** Es una técnica de aprendizaje no supervisado que agrupa observaciones con características similares. En este estudio, el clustering permitió identificar grupos de cantones amazónicos con perfiles semejantes de vulnerabilidad socioambiental (MacQueen, 1967).

**Análisis de componentes principales.** Es una técnica estadística de reducción de dimensionalidad que transforma variables correlacionadas en componentes no correlacionados. Su uso permite explorar la estructura de los datos, reducir redundancias y apoyar la construcción del índice (Jolliffe & Cadima, 2016).

**Interpretabilidad del índice y de los perfiles territoriales.** Para efectos del presente estudio, se entiende como la capacidad de reconstruir y explicar de manera transparente cómo los indicadores, sus transformaciones, orientaciones y ponderaciones producen las puntuaciones del índice, así como la forma en que las semejanzas multivariadas sustentan la asignación de los cantones a perfiles territoriales. La interpretabilidad del índice se obtuvo mediante su descomposición aditiva, identificando para cada cantón la dimensión dominante y los indicadores con mayor contribución en puntos. La interpretación de los perfiles se realizó a partir de sus promedios dimensionales y de la comparación de sus configuraciones, sin equipararlos con niveles ordinales de vulnerabilidad.

**Interculturalidad.** Es un enfoque que reconoce la coexistencia de diversas culturas, saberes y formas de relación con el territorio. En este proyecto orienta el tratamiento ético de variables vinculadas con pueblos y nacionalidades indígenas, evitando interpretaciones estigmatizantes o reduccionistas (Walsh, 2009).

**Síntesis del enfoque teórico-conceptual del estudio**

El enfoque teórico de este Proyecto de Titulación articula vulnerabilidad socioambiental, análisis territorial, ciencia de datos aplicada, aprendizaje automático e interculturalidad. Esta articulación permite estudiar la Amazonía ecuatoriana como un territorio complejo, donde las condiciones sociales y ambientales se expresan de manera diferenciada en el espacio.

La vulnerabilidad socioambiental aporta el marco para comprender la interacción entre brechas sociales, limitaciones institucionales y presiones ambientales. El análisis territorial permite ubicar esas condiciones en unidades espaciales comparables. La ciencia de datos y el aprendizaje automático ofrecen herramientas para integrar fuentes heterogéneas, identificar patrones y construir una clasificación territorial reproducible.

El índice resultante no pretende capturar toda la complejidad amazónica. Su finalidad es generar una herramienta analítica que permita apoyar procesos de priorización educativa, pastoral y social. Por ello, los resultados deberán interpretarse con cautela, considerando tanto la evidencia cuantitativa como el contexto territorial, ético e intercultural en el que se producen.

## 2.2. Metodología

El estudio se desarrolló con un enfoque cuantitativo, aplicado y exploratorio. Integró datos abiertos, registros administrativos y capas geoespaciales para construir un índice cantonal de vulnerabilidad socioambiental en la Amazonía ecuatoriana y caracterizar perfiles territoriales mediante aprendizaje automático no supervisado.

El proceso se organizó en ocho etapas: adquisición y evaluación de fuentes; extracción, transformación y carga de datos; homologación territorial; construcción de la base maestra cantonal; selección y transformación de indicadores; análisis de componentes principales; clustering y construcción del índice; validación, interpretabilidad y visualización. Esta secuencia respondió a los tres objetivos específicos del estudio.

### 2.2.1. Diseño, alcance y unidad de análisis

La investigación utilizó un diseño no experimental y transversal, basado en fuentes secundarias correspondientes a distintos períodos de referencia. Su alcance fue exploratorio porque no existía una clasificación oficial de vulnerabilidad socioambiental para los cantones amazónicos que pudiera utilizarse como variable objetivo.

La unidad territorial definitiva fue el cantón. El marco de referencia incluyó los 42 cantones de Sucumbíos, Orellana, Napo, Pastaza, Morona Santiago y Zamora Chinchipe, conforme a la división político-administrativa vigente utilizada en la base territorial.

Durante la preparación de los datos se examinó la posibilidad de construir el índice a escala parroquial. Esta alternativa se descartó porque varias fuentes prioritarias solo estaban disponibles a nivel cantonal y otras contenían variables cantonales repetidas en archivos parroquiales. Utilizar esos valores como mediciones parroquiales habría producido una falsa desagregación estadística.

Tampoco se trasladaron valores parroquiales al nivel cantonal ni se distribuyeron valores cantonales entre parroquias mediante proporciones de población, superficie u otros supuestos. En consecuencia, el índice se presentó exclusivamente a escala cantonal. La construcción de una versión parroquial quedó condicionada a la disponibilidad de nuevas fuentes con cobertura, llaves territoriales y denominadores compatibles en ese nivel.

### 2.2.2. Fuentes de datos y adquisición

Los datos se obtuvieron de repositorios institucionales, registros administrativos y plataformas geoespaciales. No se aplicaron encuestas, entrevistas ni otros instrumentos de levantamiento primario, porque el estudio se concentró en integrar y analizar información pública existente.

Cada fuente se evaluó según su institución de procedencia, período, formato, nivel territorial, llave de integración, cobertura y restricciones de uso. La matriz detallada de archivos definitivos registró 35 productos procesados y se presenta en el Anexo B. La Tabla 1 presenta una síntesis de los bloques utilizados y de su función final.

**Tabla 1**

*Matriz simplificada de fuentes y uso final*

| **Bloque de información** | **Institución o fuente** | **Período** | **Nivel original** | **Uso final en el estudio** |
|---|---|---|---|---|
| Base territorial | INEC / DPA vigente | 2024 | Cantonal y parroquial | Esqueleto de integración; códigos, nombres, geometrías y superficies. |
| Población, vivienda y conectividad | INEC, Censo de Población y Vivienda 2022 | 2022 | Cantonal y parroquial | Indicadores sociodemográficos; conectividad censal como información complementaria; control de Morona y Sevilla Don Bosco. |
| Educación escolar | Ministerio de Educación | Serie histórica y corte 2025–2026 | Cantonal; parroquial en algunos productos | Indicadores educativos del modelo principal; se utilizaron únicamente los |

| | | | | |
|---|---|---|---|---|
| | | | | archivos finales priorizados. |
| Salud | Ministerio de Salud Pública y ACESS | Cortes disponibles hasta 2026 | Puntos georreferenciados y cantonal | Indicadores de disponibilidad sanitaria por población y superficie; tasas condicionadas por denominadores censales compatibles. |
| Cobertura y uso del suelo | MapBiomas Ecuador | 1985–2024 | Geoespacial | Pérdida de bosque y transición reciente de coberturas naturales. |
| Accesibilidad, conservación y presiones | RAISG y capas geoespaciales oficiales procesadas | Periodos variables | Geoespacial | Red vial, territorios indígenas, áreas protegidas, minería, actividad petrolera, minería ilegal, hidroelectricidad y unidades hidrográficas; uso principal, contextual o de sensibilidad según el indicador. |
| Capacidad institucional | Consejo Nacional de Competencias | 2024 | Cantonal | Índice de Capacidad Operativa municipal incorporado al modelo principal con orientación inversa. |

| | | | | |
|---|---|---|---|---|
| Educación superior | Fuente administrativa procesada | 2025 | Cantonal | Información complementaria; las tasas heredaron las restricciones de los denominadores censales. |
| Bienestar y protección social | Registros administrativos de programas sociales y Registro Social | 2025–2026 / IRS25 | Cantonal y parroquial | Banco de variables y análisis de sensibilidad; no ingresó al índice principal por cobertura y ambigüedad interpretativa. |
| Primera infancia | Secretaría Técnica Ecuador Crece sin Desnutrición Infantil | Segundo trimestre de 2025 | Cantonal | Análisis descriptivo y de contraste; no se utilizó como medida de prevalencia ni en el índice principal. |
| Acceso a internet de telecomunicaciones | ARCOTEL | Corte revisado | No territorializable | Fuente excluida por ausencia de una llave territorial verificable. |

*Nota*. La matriz completa de archivos definitivos incluyó el nombre canónico de cada producto, la ruta, la llave de integración, la cobertura, los controles de calidad, el destino y la decisión de uso. Dicha matriz detallada se incorporó como anexo.

### 2.2.3. Extracción, transformación, carga y organización del data mart

La preparación de los datos siguió un proceso ETL —extracción, transformación y carga—. Los archivos originales se conservaron sin sobreescritura; las transformaciones se ejecutaron en módulos temáticos; y los productos aprobados se integraron en una base maestra mediante llaves territoriales controladas.

El data mart territorial se organizó en capas diferenciadas de documentación, datos crudos, datos procesados, base integrada, indicadores, modelado, visualizaciones y código. Esta separación preservó la procedencia de cada archivo y permitió reconstruir el recorrido desde la fuente original hasta el resultado analítico (Kimball & Caserta, 2004; Kimball & Ross, 2013).

**Tabla 2**

*Estructura del Data Mart Territorial*

| **Carpeta** | **Contenido principal** | **Función** |
|---|---|---|
| 00_documentacion | bitacora_procesamiento_general_v1; diccionario_maestro_variables_v1; matriz_archivos_definitivos_v1; README_data_mart 00_uso_ia_generativa | Documentación transversal sobre procedencia, variables, decisiones, actividades y uso del repositorio. |
| 01_datos_crudos | 19 módulos temáticos: base territorial, territorios indígenas, SNAP, vías, zonas mineras, petróleo, minería ilegal, hidroeléctricas, unidades hidrográficas, MapBiomas, Censo 2022, educación, salud, bienestar social, educación superior, acceso a internet, ICO municipal, alertas de menores de 24 meses y Registro Social de hogares. | Conservación de las fuentes originales sin sobrescritura ni transformaciones analíticas. |
| 02_datos_procesados | 19 módulos equivalentes a los datos crudos y 99_paquetes_respaldo. | Almacenamiento de datos depurados, homologados, agregados territorialmente, diagnósticos, controles y respaldos comprimidos. |
| 03_base_integrada | 00_documentacion_base_maestra; 01_base_maestra_cantonal; 02_complementarios_base_maestra; 99_paquetes_respaldo. | Consolidación de la base canónica cantonal, su documentación específica, productos complementarios y respaldos. |
| 04_indicadores | 01_catalogo_candidatos; 02_evaluacion_preseleccion; | Construcción, evaluación y selección de indicadores, además de |

| | | |
|---|---|---|
| | 03_matriz_modelo_principal;<br>04_indicadores_sensibilidad. | matrices para el modelo principal y los análisis de sensibilidad. |
| 05_modelado | 01_preprocesamiento_pca; 02_clustering;<br>03_indice_subindices;<br>04_validacion_interpretabilidad. | Resultados del preprocesamiento, PCA, agrupamiento, construcción del índice, validación, interpretabilidad y análisis territorial. |
| 06_visualizaciones | 01_preprocesamiento_pca; 02_clustering;<br>03_indice_subindices;<br>04_validacion_interpretabilidad. | Gráficos y mapas finales organizados según la etapa analítica que los produjo. |
| 07_codigo | 00_reproducibilidad;<br>01_etl; 02_indicadores;<br>03_preprocesamiento_pca; 04_clustering;<br>05_indice_subindices;<br>06_validacion_interpretabilidad. | Cuadernos de Google Colab, incluidas las versiones ejecutadas, ordenados según la secuencia de reproducción. |

El inventario completo del repositorio y la descripción de su arquitectura se encuentran disponibles en el Anexo A.

Los códigos oficiales de provincia y cantón funcionaron como llaves principales. Los nombres territoriales se mantuvieron como atributos descriptivos y controles adicionales. Las uniones tabulares utilizaron archivos CSV de indicadores, mientras que las capas GPKG y otros productos geoespaciales se conservaron como evidencia de los recortes, intersecciones y agregaciones.

La integración distinguió los valores iguales a cero de los datos faltantes. El cero se conservó únicamente cuando la cobertura de la fuente permitió confirmar la ausencia del fenómeno. Los registros sin cobertura suficiente, con incompatibilidad territorial o sin llave verificable se mantuvieron como faltantes o fueron excluidos de la base analítica.

La trazabilidad se documentó mediante la matriz de fuentes, la matriz de archivos definitivos, el diccionario maestro retrospectivo, controles de calidad, manifiestos de ejecución y cuadernos de procesamiento. El diccionario consolidó 924 variables canónicas y distinguió su nivel territorial, fuente, definición, unidad, rol, restricciones y confianza documental. El diccionario maestro completo puede consultarse en el Anexo C.

La calidad se examinó en relación con el uso previsto y consideró completitud, consistencia, exactitud y trazabilidad, de acuerdo con el modelo de calidad de datos ISO/IEC 25012 (International Organization for Standardization, 2008). Los principios FAIR orientaron la incorporación de metadatos, procedencia, identificadores y reglas de transformación (Wilkinson et al., 2016). Los reportes de control de calidad, manifiestos de ejecución y auditoría geoespacial se incluyen en el Anexo I.

### 2.2.4. Uso de inteligencia artificial generativa en el procesamiento

Durante el desarrollo se utilizó ChatGPT como herramienta de apoyo para estructurar y depurar scripts de Python, documentar cuadernos de procesamiento y preparar tablas derivadas de los resultados, explorar fuentes bibliográficas y estándares para el adecuado manejo de los datos dentro del estudio (OpenAI, 2026). Su participación se limitó a asistencia técnica y editorial de un proceso supervisado.

La autora definió las decisiones metodológicas, seleccionó las fuentes, estableció los criterios de inclusión y exclusión, ejecutó el código, verificó los archivos de salida y realizó la interpretación sustantiva. Los cuadernos ejecutados, los controles y productos exportados constituyeron la evidencia reproducible del procedimiento. La declaración ampliada, la bitácora de uso y una muestra representativa de las interacciones empleadas durante el proceso se presentan en el Anexo K.

### 2.2.5. Homologación territorial y tratamiento de Morona y Sevilla Don Bosco

La integración utilizó la división político-administrativa vigente como marco territorial común. Esta decisión permitió representar la configuración actual de los 42 cantones amazónicos, pero generó una incompatibilidad con fuentes anteriores a la creación del cantón Sevilla Don Bosco.

El Censo de Población y Vivienda 2022 y otras fuentes previas registraron el territorio de Sevilla Don Bosco dentro del cantón Morona. No se dispuso de una correspondencia oficial suficientemente detallada para redistribuir los valores censales entre ambos cantones con exactitud.

Por esta razón, los valores de Morona no se dividieron ni se imputaron a Sevilla Don Bosco. Los ceros generados por la inexistencia del nuevo código cantonal en las fuentes antiguas se trataron como faltantes estructurales y no como ausencia real del fenómeno.

La incompatibilidad también se trasladó a los indicadores que utilizaron población censal como denominador. Las tasas de servicios de salud, educación superior u otras variables quedaron excluidas cuando el numerador y el denominador no correspondieron a la misma configuración territorial.

La base maestra conservó los 42 cantones y banderas de control asociadas. La matriz principal de modelado incluyó 40 cantones con información completa; Morona y Sevilla Don Bosco permanecieron sin puntuación en el índice y se mantuvieron en los productos descriptivos y geográficos para preservar la transparencia territorial.

### 2.2.6. Construcción y selección de indicadores

A partir de la base maestra se construyeron 34 indicadores candidatos expresados como porcentajes, tasas, razones y densidades. Su evaluación combinó criterios conceptuales, documentales y estadísticos, conforme a las recomendaciones para índices compuestos

(Organisation for Economic Co-operation and Development & European Commission, Joint Research Centre, 2008).

Los criterios conceptuales examinaron la relación con la vulnerabilidad socioambiental y la interpretación de cada indicador. Los criterios documentales consideraron procedencia, período, nivel original, fórmula, unidad y restricciones. Los criterios estadísticos incluyeron cobertura, faltantes, variabilidad, concentración de ceros, valores atípicos y correlaciones.

La evaluación preseleccionó 13 indicadores para el modelo principal, reservó 14 para la sensibilidad, mantuvo cinco como variables contextuales o de estratificación, dejó uno pendiente de definición y excluyó uno por variabilidad insuficiente. Los indicadores principales se distribuyeron en seis dimensiones. El catálogo de los 34 indicadores candidatos, sus criterios de evaluación y la decisión adoptada para cada uno se presentan en el Anexo D.

**Tabla 3**

*Indicadores del Modelo Principal*

| **Dimensión** | **Indicadores incorporados** | **Tratamiento de orientación** |
|---|---|---|
| Sociodemográfica | Porcentaje de población rural; índice aproximado de dependencia demográfica. | Directa: valores altos representaron mayor vulnerabilidad. |
| Educativa | Estudiantes por docente; instituciones educativas por 1.000 km²; instituciones educativas por 10.000 habitantes. | La relación estudiantes/docente fue directa; los indicadores de oferta se invirtieron. |
| Sanitaria | Establecimientos de salud por 10.000 habitantes; establecimientos por 1.000 km². | Inversa: menor disponibilidad se transformó en mayor vulnerabilidad. |
| Accesibilidad territorial | Densidad de carreteras. | Inversa: menor densidad se transformó en mayor vulnerabilidad. |

| | | |
|---|---|---|
| Ambiental | Pérdida de bosque 1985–2024; transición reciente de coberturas naturales 2020–2024; superficie asociada con zonas mineras; superficie vinculada con actividad petrolera. | Directa: valores altos representaron mayor presión o transformación. |
| Capacidad institucional | Índice de Capacidad Operativa municipal 2024. | Inversa: menor capacidad operativa se transformó en mayor vulnerabilidad. |

La orientación se armonizó antes de la normalización, de modo que los valores altos representaran consistentemente mayor vulnerabilidad. Los valores originales se conservaron y las transformaciones se almacenaron en columnas separadas para mantener la trazabilidad.

Las variables interculturales —población indígena, pueblos y nacionalidades, superposición con territorios indígenas y oferta de educación intercultural bilingüe— se utilizaron como elementos contextuales. No recibieron una orientación positiva o negativa ni se incorporaron como carencias, en concordancia con los principios CARE para la gobernanza de datos indígenas (Carroll et al., 2020).

La dimensión digital no formó parte del índice principal. La fuente de ARCOTEL no contenía una llave territorial verificable y no se asignó artificialmente a los cantones. La conectividad censal se conservó como información complementaria, sin incorporarla a los 13 indicadores modelados.

Los registros de bienestar, transferencias y Registro Social se reservaron para análisis de sensibilidad. Su interpretación resultó ambivalente porque una mayor cobertura podía reflejar mayor necesidad social como mayor capacidad de respuesta institucional.

### 2.2.7. Preprocesamiento, normalización y análisis de componentes principales

La matriz analítica principal incluyó 40 cantones y 13 indicadores completos. Antes del modelado se revisaron distribuciones, asimetrías, ceros, valores atípicos y correlaciones.

Los valores atípicos no se eliminaron automáticamente, porque podían representar condiciones territoriales reales.

Los indicadores con asimetría absoluta igual o superior a 1 se transformaron mediante Yeo-Johnson. El umbral de asimetría se utilizó como criterio operativo para identificar distribuciones claramente asimétricas y decidir de manera reproducible qué variables requerían transformación; no se interpretó como una prueba estadística de normalidad. Después se construyeron dos matrices: una estandarizada con puntuaciones Z y otra normalizada entre 0 y 1 mediante Min-Max (Yeo & Johnson, 2000).

La retención de componentes consideró el análisis paralelo, el criterio de Kaiser, la varianza acumulada, el gráfico de sedimentación y la interpretabilidad de las cargas. Las cargas, comunalidades o coordenadas territoriales se examinaron antes y después de una rotación Varimax (Horn, 1965; Kaiser, 1958; Jolliffe & Cadima, 2016).

Se comparó un PCA con los 13 indicadores y una variante con 11 indicadores que excluyó temporalmente el Índice de Capacidad Operativa y la relación estudiantes por docente. Esta comparación permitió evaluar la influencia de variables con menor adecuación individual sin eliminarlas del análisis sustantivo. El PCA se utilizó como herramienta exploratoria de reducción y diagnóstico. No sustituyó el marco conceptual ni determinó por sí solo la composición del índice. Los diagnósticos, cargas, comunalidades, varianza explicada y demás resultados detallados del análisis de componentes principales se encuentran en el Anexo E.

### 2.2.8. Clustering y caracterización de perfiles territoriales

El aprendizaje no supervisado se aplicó para identificar perfiles de cantones con configuraciones semejantes. Se compararon tres representaciones de los datos: los 13 indicadores estandarizados, tres componentes del PCA con 13 indicadores y tres componentes del PCA reducido con 11 indicadores.

Se evaluaron K-means y clustering jerárquico aglomerativo con enlace Ward para soluciones entre dos y seis grupos. DBSCAN se utilizó como diagnóstico exploratorio de agrupamientos por densidad y de posibles observaciones atípicas.

La comparación consideró silhouette score, índice Davies-Bouldin, criterio Calinski-Harabasz, tamaño y equilibrio de los grupos. La estabilidad se examinó mediante 50 submuestras aleatorias del 80% de los cantones y el Índice de Rand ajustado (Rousseeuw, 1987; Davies & Bouldin, 1979; Caliński & Harabasz, 1974; Hubert & Arabie, 1985).

No se aceptaron como solución principal configuraciones con grupos menores de cuatro cantones. El mínimo de cuatro observaciones por grupo se definió como una restricción pragmática previa a la selección, orientada a evitar perfiles residuales de difícil interpretación dada la reducida muestra de 40 cantones; no constituye un umbral estadístico universal. La selección combinó calidad interna, estabilidad, equilibrio e interpretabilidad, y favoreció la alternativa más parsimoniosa cuando varias soluciones presentaron un desempeño comparable. En consecuencia, ninguna métrica se utilizó como criterio único de selección: silhouette, Davies-Bouldin y Calinski-Harabasz se interpretaron comparativamente entre las soluciones evaluadas, mientras que la estabilidad por submuestreo, el tamaño de los grupos y su interpretabilidad actuaron como criterios complementarios.

Los grupos se interpretaron como perfiles territoriales multidimensionales. No se equipararon automáticamente con niveles de vulnerabilidad, porque el clustering describió configuraciones internas y el índice produjo un ordenamiento relativo. La comparación de algoritmos, métricas de validación, pruebas de estabilidad y asignación cantonal definitiva se documenta en el Anexo F.

### 2.2.9. Construcción del índice y clasificación relativa

La versión principal del índice se construyó mediante ponderación equilibrada por dimensiones. Cada una de las seis dimensiones recibió el mismo peso y, dentro de cada

dimensión, los indicadores recibieron pesos iguales. Los subíndices se calcularon con valores Min-Max y el resultado global se reescaló entre 0 y 100.

La ponderación equilibrada evitó que las dimensiones con mayor número de indicadores dominaran el resultado. Como alternativa estadística se construyó un índice basado en las cargas al cuadrado de tres componentes principales, ponderadas por su varianza explicada. Dicha ponderación se adoptó como escenario principal por su transparencia conceptual y porque evita que el número de indicadores o la varianza estadística determine automáticamente el peso de una dimensión. No se asumió que esta estructura constituyera una ponderación óptima; por esta razón se contrastó mediante alternativas basadas en PCA y otras pruebas de sensibilidad.

También se calcularon dos variantes de sensibilidad: el promedio simple de los 13 indicadores y un índice basado en el PCA reducido de 11 indicadores. La comparación entre métodos permitió evaluar la dependencia del ranking respecto de la estrategia de ponderación.

Los 40 cantones modelados se clasificaron por cuartiles en niveles relativos de vulnerabilidad baja, media, alta y muy alta. Estas categorías describieron la posición dentro de la muestra y no representaron umbrales absolutos ni diagnósticos equivalentes para otros territorios.

### 2.2.10. Validación, interpretabilidad y análisis espacial

La validación verificó la integridad aritmética y territorial de los resultados. Se comprobó que la suma de las contribuciones de indicadores y dimensiones reprodujera exactamente la puntuación global y que la capa final conservara los 42 cantones, con puntuaciones para los 40 incluidos en el modelo.

La interpretabilidad se obtuvo mediante la descomposición directa del índice. Para cada cantón se identificaron la dimensión dominante y los indicadores con mayor

contribución en puntos. Esta estrategia resultó coherente con la estructura aditiva del índice y no requirió un modelo supervisado adicional.

La sensibilidad se examinó mediante la comparación de ponderaciones, la exclusión sucesiva de dimensiones y perturbaciones simuladas de los pesos. Para cada cantón se estimaron cambios de ranking, amplitud de posiciones y probabilidad de permanencia en cada categoría. La sensibilidad metodológica se interpretó separadamente de la vulnerabilidad. Los resultados completos de interpretabilidad y sensibilidad para cada cantón se presentan en el Anexo G.

La dependencia espacial se evaluó mediante el índice global de Moran y los indicadores locales de asociación espacial. Las pruebas de permutación permitieron contrastar si los patrones observados diferían de una distribución espacial aleatoria (Anselin, 1995).

Los resultados se representaron mediante mapas del índice, niveles relativos, subíndices, perfiles de clustering y sensibilidad metodológica. También se generó un mapa interactivo y se exportaron capas enriquecidas en formatos GeoJSON y GPKG. Las versiones complementarias de los productos cartográficos, gráficos y el mapa interactivo se encuentran disponibles en el Anexo H.

### 2.2.11. Herramientas técnicas y reproducibilidad

El procesamiento se ejecutó en Python mediante cuadernos Jupyter y Google Colab. Se utilizaron Pandas y NumPy para manipulación de datos; SciPy y Scikit-learn para transformaciones, PCA, clustering y métricas; Matplotlib para gráficos; GeoPandas para datos espaciales; y Folium para visualización interactiva. Los cuadernos ejecutados y los archivos necesarios para reproducir el procesamiento y modelado se encuentran disponibles en el repositorio indicado en el Anexo J.

Cada etapa exportó matrices, controles gráficos, manifiestos y paquetes versionados. Los cuadernos conservaron el código y las salidas de ejecución, lo que permitió revisar el procedimiento y reproducir sus resultados (Peng, 2011; Rule et al., 2019).

El índice se interpretó como una herramienta de apoyo para la priorización institucional. No sustituyó el conocimiento situado, la deliberación territorial ni las decisiones de los actores que trabajan en la Amazonía. Se suma como una herramienta complementaria para la toma de decisiones de procesos de intervención social.

### 2.3. Resultados y análisis

El análisis integra tres productos complementarios: el índice global de vulnerabilidad socioambiental, los seis subíndices que explican su composición y los perfiles territoriales obtenidos mediante clustering. El índice ordena a los cantones según su vulnerabilidad relativa; los subíndices permiten identificar las dimensiones que sostienen cada puntuación; y los perfiles describen configuraciones multivariadas que no deben interpretarse como niveles ordinales. La lectura conjunta de estas salidas evita reducir la complejidad territorial a una sola posición de ranking y permite vincular los resultados cuantitativos con los procesos sociales, institucionales y ambientales diferenciados.

La vulnerabilidad se entiende aquí como una condición relacional y multidimensional. No resulta únicamente de la exposición a presiones ambientales ni de una suma de carencias individuales, sino de la interacción entre desigualdades sociodemográficas, acceso a servicios, capacidades institucionales, conectividad territorial y transformaciones ambientales. Esta perspectiva coincide con los enfoques que conciben la vulnerabilidad como producto de condiciones sociales e institucionales que configuran capacidades desiguales para anticipar, afrontar y transformar situaciones adversas (Adger, 2006; Cutter et al., 2003).

### 2.3.1. Construcción del índice y distribución territorial

El índice se calculó para 40 de los 42 cantones que conforman actualmente la Amazonía ecuatoriana. Morona y Sevilla Don Bosco permanecieron en la base territorial, pero no recibieron puntuación debido a la incompatibilidad estructural entre las fuentes censales de 2022 y la división político-administrativa vigente. No se dividieron, heredaron ni imputaron valores entre ambos territorios, pues cualquiera de esas alternativas habría introducido información no observada y afectado artificialmente la normalización, ponderación y el agrupamiento.

La versión principal utiliza una ponderación equilibrada: cada una de las seis dimensiones aporta 16,67% al índice global y, dentro de cada dimensión, los indicadores reciben pesos iguales. Las puntuaciones se expresan en una escala de 0 a 100, donde un valor mayor representa una posición de vulnerabilidad comparativamente más alta dentro del universo analizado. La media fue 54,33; el mínimo, 34,37; y el máximo, 76,91. La amplitud de 42,54 puntos confirma que existe diferenciación sustantiva entre los cantones, aunque las distancias entre posiciones consecutivas no son uniformes.

Las categorías Baja, Media, Alta y Muy alta se construyeron mediante cuartiles, por lo que cada grupo contiene diez cantones. Este equilibrio numérico es una propiedad del método de clasificación y no una evidencia de que la vulnerabilidad se distribuya en cuatro bloques naturales. En consecuencia, el ranking y la puntuación continua deben conservarse como referencias principales, mientras la categoría cumple una función comunicativa y comparativa.

**Tabla 4**

*Ranking completo del Índice de Vulnerabilidad Socioambiental*

| Pos. | Cantón | Provincia | Índice | Nivel | Perfil | Dimensión dominante |
|---|---|---|---|---|---|---|
| 1 | Logroño | Morona Santiago | 76,91 | Muy alta | Perfil 3 | Sanitaria |
| 2 | Taisha | Morona Santiago | 75,32 | Muy alta | Perfil 3 | Accesibilidad territorial |
| 3 | Aguarico | Orellana | 71,09 | Muy alta | Perfil 2 | Sanitaria |
| 4 | Huamboya | Morona Santiago | 67,94 | Muy alta | Perfil 3 | Sanitaria |
| 5 | Palora | Morona Santiago | 67,72 | Muy alta | Perfil 2 | Capacidad institucional |
| 6 | Tiwintza | Morona Santiago | 64,33 | Muy alta | Perfil 3 | Sanitaria |
| 7 | Pablo Sexto | Morona Santiago | 63,60 | Muy alta | Perfil 2 | Sanitaria |
| 8 | Pastaza | Pastaza | 61,71 | Muy alta | Perfil 2 | Accesibilidad territorial |
| 9 | Arajuno | Pastaza | 61,47 | Muy alta | Perfil 2 | Sanitaria |
| 10 | Carlos Julio Arosemena Tola | Napo | 61,44 | Muy alta | Perfil 3 | Sanitaria |
| 11 | Palanda | Zamora Chinchipe | 61,37 | Alta | Perfil 3 | Sanitaria |
| 12 | Putumayo | Sucumbíos | 60,74 | Alta | Perfil 2 | Capacidad institucional |
| 13 | Limón Indanza | Morona Santiago | 60,07 | Alta | Perfil 3 | Sanitaria |
| 14 | Sucúa | Morona Santiago | 58,60 | Alta | Perfil 1 | Educativa |
| 15 | Santiago | Morona Santiago | 57,67 | Alta | Perfil 3 | Sanitaria |
| 16 | Loreto | Orellana | 57,31 | Alta | Perfil 3 | Sanitaria |
| 17 | Yacuambi | Zamora Chinchipe | 56,96 | Alta | Perfil 3 | Accesibilidad territorial |
| 18 | Mera | Pastaza | 56,29 | Alta | Perfil 1 | Educativa |
| 19 | Archidona | Napo | 56,27 | Alta | Perfil 2 | Sanitaria |
| 20 | Sucumbíos | Sucumbíos | 56,05 | Alta | Perfil 2 | Sanitaria |
| 21 | San Juan Bosco | Morona Santiago | 55,24 | Media | Perfil 3 | Sanitaria |
| 22 | Gualaquiza | Morona Santiago | 55,12 | Media | Perfil 3 | Sanitaria |
| 23 | Chinchipe | Zamora Chinchipe | 54,96 | Media | Perfil 3 | Sanitaria |
| 24 | Cascales | Sucumbíos | 53,16 | Media | Perfil 1 | Educativa |
| 25 | Cuyabeno | Sucumbíos | 52,19 | Media | Perfil 2 | Accesibilidad territorial |

| | | | | | | |
|---|---|---|---|---|---|---|
| 26 | Nangaritza | Zamora Chinchipe | 51,54 | Media | Perfil 2 | Accesibilidad territorial |
| 27 | Gonzalo Pizarro | Sucumbíos | 51,49 | Media | Perfil 2 | Educativa |
| 28 | Tena | Napo | 51,21 | Media | Perfil 1 | Accesibilidad territorial |
| 29 | Francisco de Orellana | Orellana | 46,19 | Media | Perfil 1 | Educativa |
| 30 | Zamora | Zamora Chinchipe | 44,60 | Media | Perfil 1 | Educativa |
| 31 | El Pangui | Zamora Chinchipe | 42,85 | Baja | Perfil 1 | Educativa |
| 32 | Centinela del Cóndor | Zamora Chinchipe | 42,54 | Baja | Perfil 1 | Ambiental |
| 33 | Santa Clara | Pastaza | 42,40 | Baja | Perfil 1 | Ambiental |
| 34 | El Chaco | Napo | 42,32 | Baja | Perfil 2 | Accesibilidad territorial |
| 35 | Shushufindi | Sucumbíos | 42,04 | Baja | Perfil 1 | Educativa |
| 36 | Paquisha | Zamora Chinchipe | 41,89 | Baja | Perfil 1 | Capacidad institucional |
| 37 | Yantzaza | Zamora Chinchipe | 41,73 | Baja | Perfil 1 | Educativa |
| 38 | Lago Agrio | Sucumbíos | 37,34 | Baja | Perfil 1 | Ambiental |
| 39 | La Joya de los Sachas | Orellana | 37,24 | Baja | Perfil 1 | Ambiental |
| 40 | Quijos | Napo | 34,37 | Baja | Perfil 2 | Educativa |
| N/A | Morona | Morona Santiago | — | — | — | — |
| N/A | Sevilla Don Bosco | Morona Santiago | — | — | — | — |

*Nota.* Morona y Sevilla Don Bosco se mantienen visibles sin puntuación por incompatibilidad territorial estructural.

Logroño ocupó la primera posición con 76,91 puntos, seguido por Taisha con 75,32 y Aguarico con 71,09. Huamboya, Palora, Tiwintza y Pablo Sexto completaron las siete primeras posiciones. La presencia de seis cantones de Morona Santiago entre los diez primeros muestra una concentración territorial relevante, pero no autoriza a interpretar la provincia como una unidad homogénea ni a atribuir causalidad a la pertenencia provincial. El resumen provincial es descriptivo: Morona Santiago alcanzó una media de 63,87 puntos y

nueve de sus once cantones modelados se ubicaron en los niveles alto o muy alto. Este patrón debe examinarse mediante las dimensiones y las relaciones espaciales específicas.

En el extremo inferior se ubicaron Quijos (34,37), La Joya de los Sachas (37,24), Lago Agrio (37,34), Yantzaza (41,73) y Paquisha (41,89). Estas posiciones no equivalen a ausencia de vulnerabilidad. Indican solamente que, frente al conjunto amazónico modelado, esos cantones presentaron una combinación relativa menos elevada de los 13 indicadores utilizados. Algunos de ellos registraron, no obstante, puntuaciones altas en dimensiones particulares; por ejemplo, La Joya de los Sachas alcanzó 68,42 en el subíndice ambiental y Yantzaza 69,95 en el educativo. Esta divergencia confirma que la posición global no sustituye el análisis de la composición interna.

**Figura 1**

*Ranking completo de vulnerabilidad socioambiental*

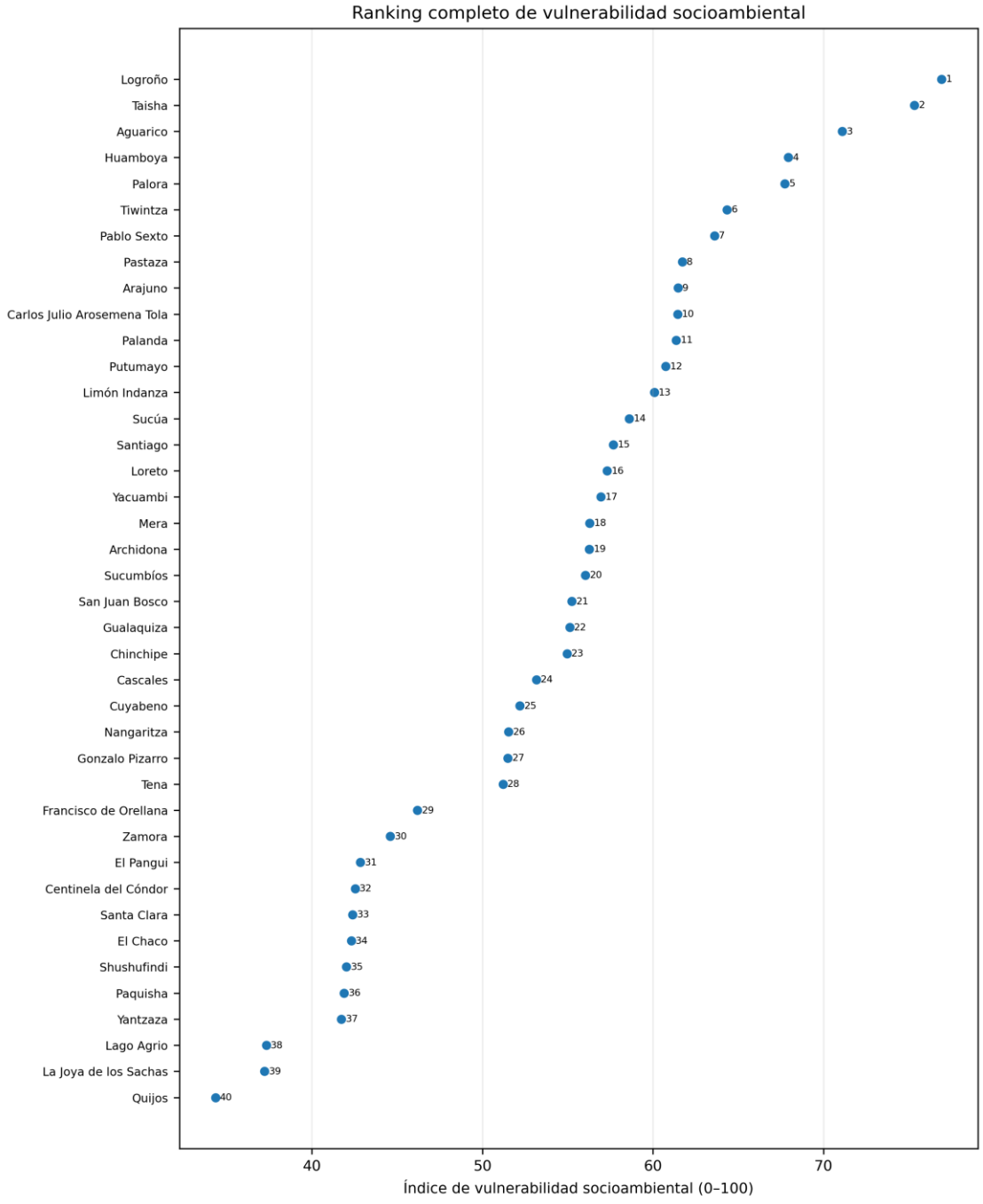
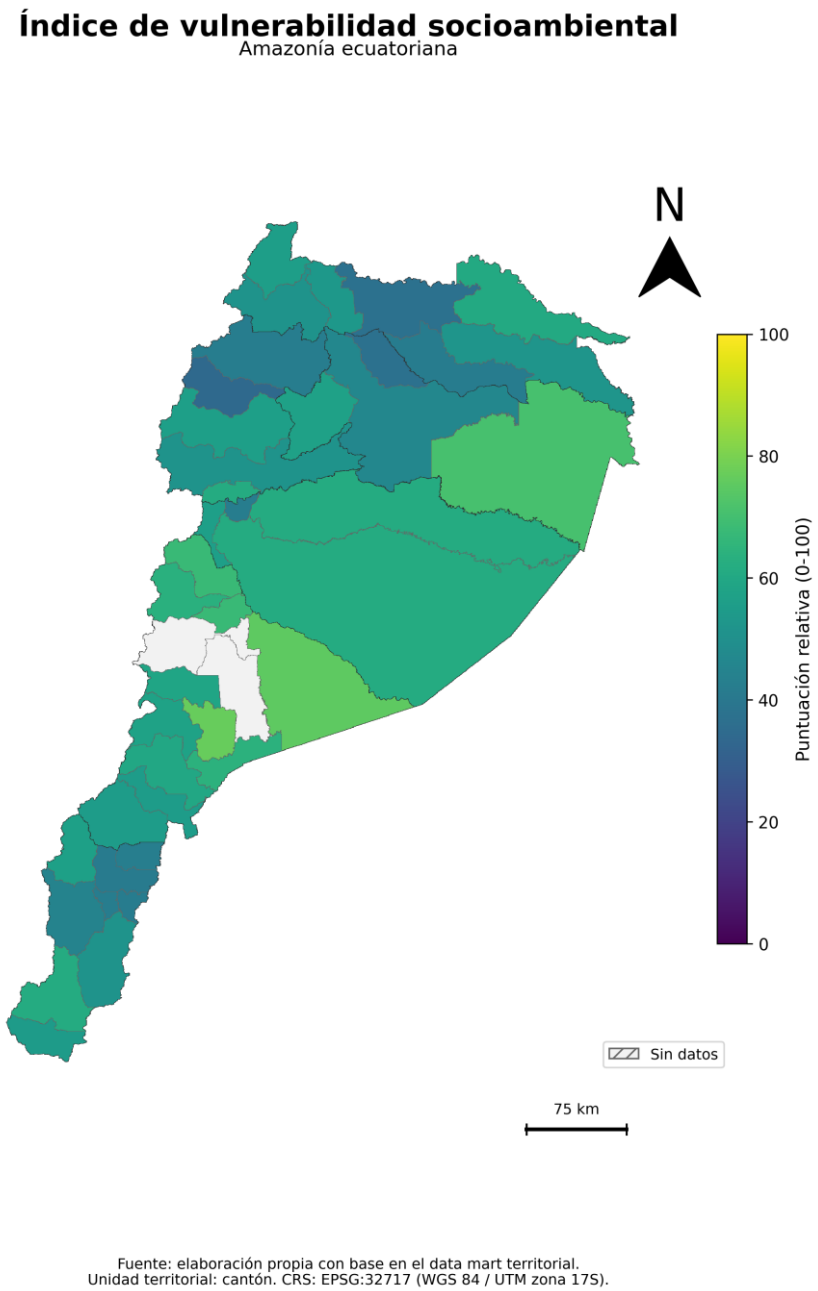


**Figura 2**

*Distribución territorial del índice de vulnerabilidad socioambiental*




*Nota*. Unidad territorial: cantón; CRS EPSG:32717 (WGS 84 / UTM zona 17S).

La auditoría cartográfica confirmó la validez de las 42 geometrías y la conservación de 1.632.802 vértices, sin aplicar simplificación. Se registró únicamente un micro-solapamiento de 342,40 m² entre Taisha y Tiwintza —equivalente a 0,000342 km²—,

insuficiente para alterar la representación o los resultados y conservado en el reporte de control sin modificar la geometría canónica.

### 2.3.2. Estructura dimensional de la vulnerabilidad

La descomposición dimensional revela que los cantones no alcanzan puntuaciones similares por las mismas razones. La media más elevada correspondió al subíndice sanitario (67,98), seguido por el educativo (56,90) y el de accesibilidad territorial (55,71). Las dimensiones sociodemográfica, capacidad institucional y ambiental registraron medias de 49,60; 48,89 y 46,90, respectivamente. Estas medias no deben interpretarse como una comparación absoluta entre sectores, porque cada subíndice fue normalizado dentro del conjunto de cantones y depende de la distribución de sus indicadores.

**Tabla 5**

*Estadísticos descriptivos de los seis subíndices*

| Dimensión | Media | Mediana | D.E. | Mín. | Máx. | ρ con índice | Máximo relativo |
|---|---|---|---|---|---|---|---|
| Sociodemográfica | 49,60 | 46,10 | 20,60 | 11,44 | 98,74 | 0,66 | Taisha |
| Educativa | 56,90 | 60,59 | 13,52 | 23,10 | 79,47 | -0,02 | Palora |
| Sanitaria | 67,98 | 67,85 | 27,64 | 7,28 | 100,00 | 0,79 | Limón Indanza |
| Accesibilidad territorial | 55,71 | 56,83 | 27,01 | 0,00 | 100,00 | 0,56 | Pablo Sexto |
| Ambiental | 46,90 | 49,55 | 17,37 | 11,22 | 75,69 | -0,11 | Tiwintza |
| Capacidad institucional | 48,89 | 50,55 | 20,45 | 0,00 | 100,00 | 0,55 | Palora |

El subíndice sanitario presentó un rasgo que debe hacerse explícito: doce cantones alcanzaron la puntuación máxima de 100. Esta acumulación de empates sugiere un efecto de techo derivado de la distribución y normalización de los indicadores sanitarios. Por ello, un ranking estricto dentro de esta dimensión sería artificial para los casos empatados. La interpretación adecuada consiste en reconocer un grupo amplio con la mayor vulnerabilidad relativa observada, no en imponer diferencias inexistentes entre sus miembros.

La dimensión sociodemográfica mostró una concentración destacada en Morona Santiago. Taisha (98,74), Tiwintza (88,02) y Huamboya (87,05) ocuparon las tres primeras posiciones, seguidos por Aguarico (82,43) y Logroño (77,25). Este patrón no debe asociarse mecánicamente con la composición étnica de los territorios. Las variables interculturales no fueron utilizadas para puntuar vulnerabilidad; la interpretación se refiere a las condiciones sociodemográficas incluidas en la matriz analítica y no a la identidad indígena como atributo deficitario.

En educación, las puntuaciones más altas correspondieron a Palora (79,47), Mera (78,53), Aguarico (74,14), Francisco de Orellana (73,99) y Pastaza (72,16). La accesibilidad territorial, por su parte, concentró sus mayores valores en Pablo Sexto (100), Arajuno (99,67), Taisha (99,13), Aguarico (96,66) y Pastaza (90,77). En estos casos, la accesibilidad no debe reducirse a distancia geométrica. Expresa desigualdades en la articulación con redes viales, centros de servicios y circuitos institucionales; sus efectos dependen además de la dispersión poblacional, la extensión territorial y las formas de movilidad fluvial o terrestre.

La dimensión ambiental presentó sus mayores valores en Tiwintza (75,69), Huamboya (75,58), Centinela del Cóndor (69,40), Logroño (68,76) y la Joya de los Sachas (68,42). Su distribución muestra que la presión ambiental no sigue exactamente el mismo patrón que las brechas de servicios. Aguarico, por ejemplo, se ubicó tercero en el índice global y mostró valores muy altos en salud, accesibilidad y condiciones sociodemográficas, pero una puntuación ambiental comparativamente baja de 13,15. Esta diferencia evita confundir vulnerabilidad socioambiental con presión ambiental exclusivamente.

En capacidad institucional, Palora alcanzó 100 puntos, seguido por Putumayo (85,96), Sucumbíos (83,03), Palanda (77,74) y Taisha (68,27). Estos resultados no constituyen una evaluación moral de la gestión municipal. La capacidad operativa observada está condicionada por recursos fiscales, escala territorial, densidad poblacional, trayectoria

administrativa y disponibilidad de información. El subíndice identifica una condición relativa que requiere contextualización institucional.

**Tabla 6**

*Puntuaciones más altas por dimensión*

| Dimensión | Pos. | Cantón | Provincia | Puntuación |
|---|---|---|---|---|
| Accesibilidad territorial | 1 | Pablo Sexto | Morona Santiago | 100,00 |
| Accesibilidad territorial | 2 | Arajuno | Pastaza | 99,67 |
| Accesibilidad territorial | 3 | Taisha | Morona Santiago | 99,13 |
| Accesibilidad territorial | 4 | Aguarico | Orellana | 96,66 |
| Accesibilidad territorial | 5 | Pastaza | Pastaza | 90,77 |
| Ambiental | 1 | Tiwintza | Morona Santiago | 75,69 |
| Ambiental | 2 | Huamboya | Morona Santiago | 75,58 |
| Ambiental | 3 | Centinela del Cóndor | Zamora Chinchipe | 69,40 |
| Ambiental | 4 | Logroño | Morona Santiago | 68,76 |
| Ambiental | 5 | La Joya de los Sachas | Orellana | 68,42 |
| Capacidad institucional | 1 | Palora | Morona Santiago | 100,00 |
| Capacidad institucional | 2 | Putumayo | Sucumbíos | 85,96 |
| Capacidad institucional | 3 | Sucumbíos | Sucumbíos | 83,03 |
| Capacidad institucional | 4 | Palanda | Zamora Chinchipe | 77,74 |
| Capacidad institucional | 5 | Taisha | Morona Santiago | 68,27 |
| Educativa | 1 | Palora | Morona Santiago | 79,47 |
| Educativa | 2 | Mera | Pastaza | 78,53 |
| Educativa | 3 | Aguarico | Orellana | 74,14 |
| Educativa | 4 | Francisco de Orellana | Orellana | 73,99 |
| Educativa | 5 | Pastaza | Pastaza | 72,16 |

| | | | | |
|---|---|---|---|---|
| Sanitaria | 1 | Aguarico | Orellana | 100,00 |
| Sanitaria | 1 | Arajuno | Pastaza | 100,00 |
| Sanitaria | 1 | Carlos Julio Arosemena Tola | Napo | 100,00 |
| Sanitaria | 1 | Chinchipe | Zamora Chinchipe | 100,00 |
| Sanitaria | 1 | Huamboya | Morona Santiago | 100,00 |
| Sanitaria | 1 | Limón Indanza | Morona Santiago | 100,00 |
| Sanitaria | 1 | Logroño | Morona Santiago | 100,00 |
| Sanitaria | 1 | Pablo Sexto | Morona Santiago | 100,00 |
| Sanitaria | 1 | San Juan Bosco | Morona Santiago | 100,00 |
| Sanitaria | 1 | Santiago | Morona Santiago | 100,00 |
| Sanitaria | 1 | Sucumbíos | Sucumbíos | 100,00 |
| Sanitaria | 1 | Tiwintza | Morona Santiago | 100,00 |
| Sociodemográfica | 1 | Taisha | Morona Santiago | 98,74 |
| Sociodemográfica | 2 | Tiwintza | Morona Santiago | 88,02 |
| Sociodemográfica | 3 | Huamboya | Morona Santiago | 87,05 |
| Sociodemográfica | 4 | Aguarico | Orellana | 82,43 |
| Sociodemográfica | 5 | Logroño | Morona Santiago | 77,25 |

*Nota*. En la dimensión sanitaria existen empates adicionales en 100 puntos; la posición se representa únicamente como referencia de ordenación.

**Figura 3**

*Perfil dimensional de los cantones ordenados por el índice global*

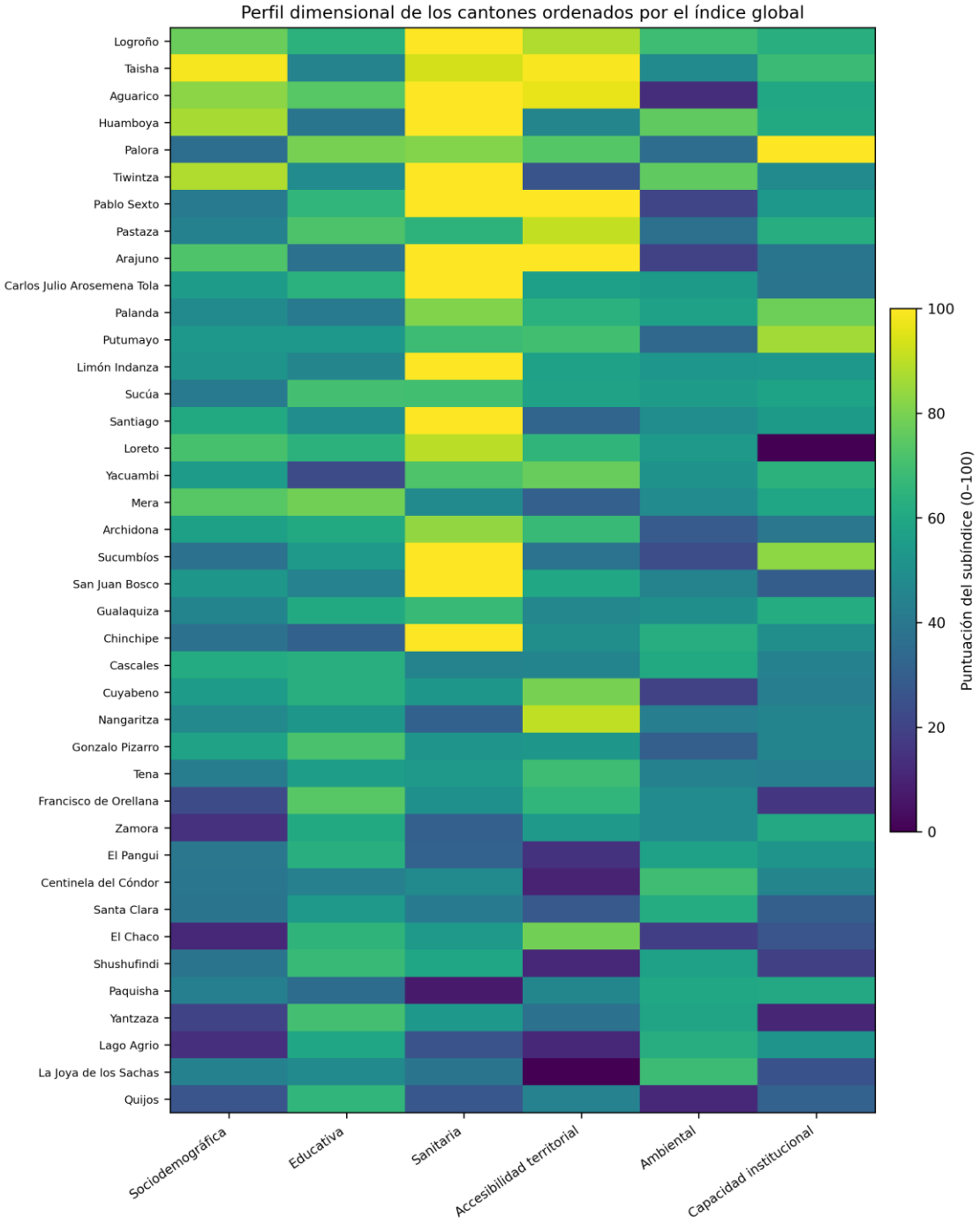
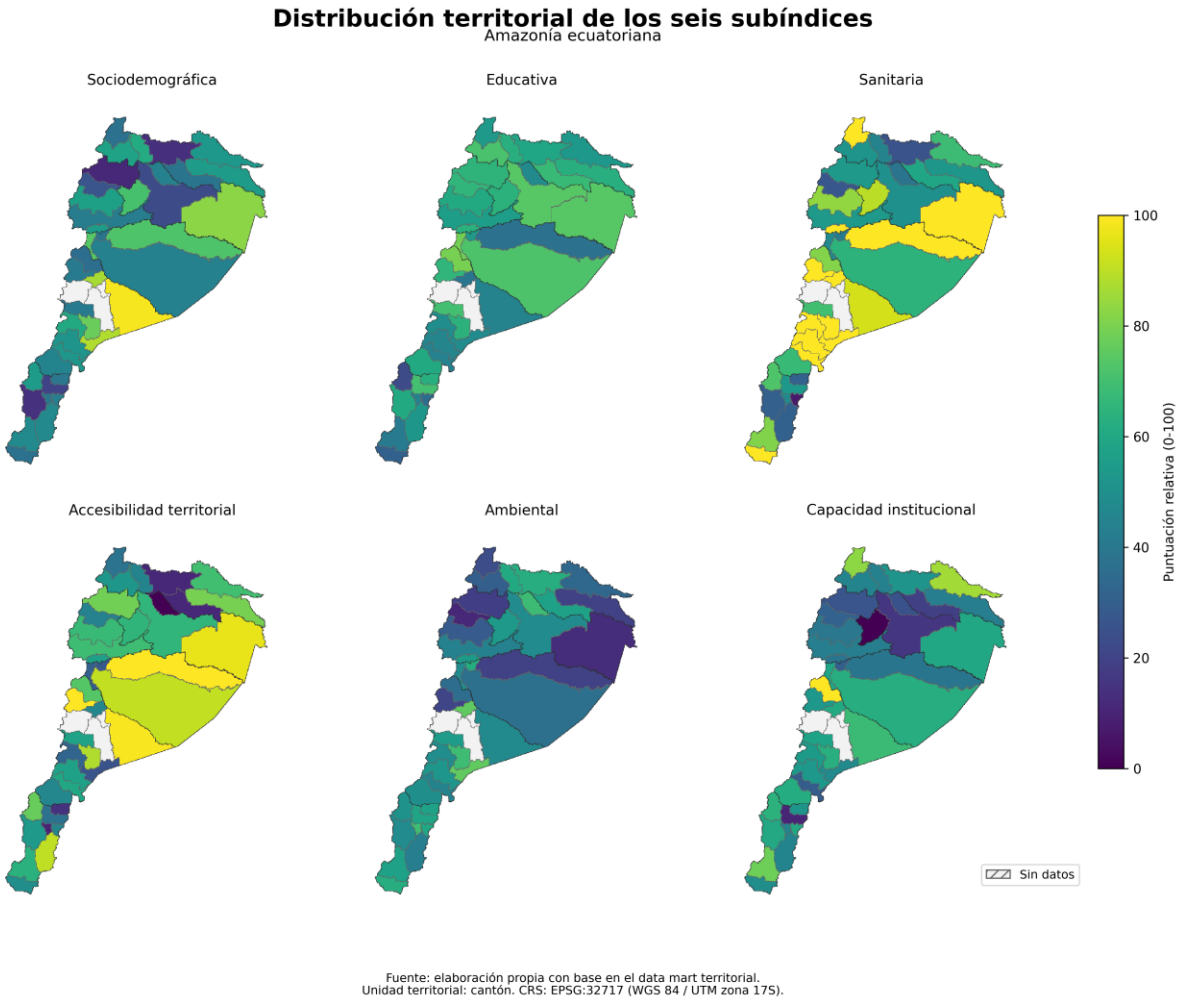


**Figura 4**

*Distribución territorial de los seis subíndices de vulnerabilidad*




*Nota.* Unidad territorial: cantón; CRS EPSG:32717 (WGS 84 / UTM zona 17S). La escala cromática es común de 0 a 100.

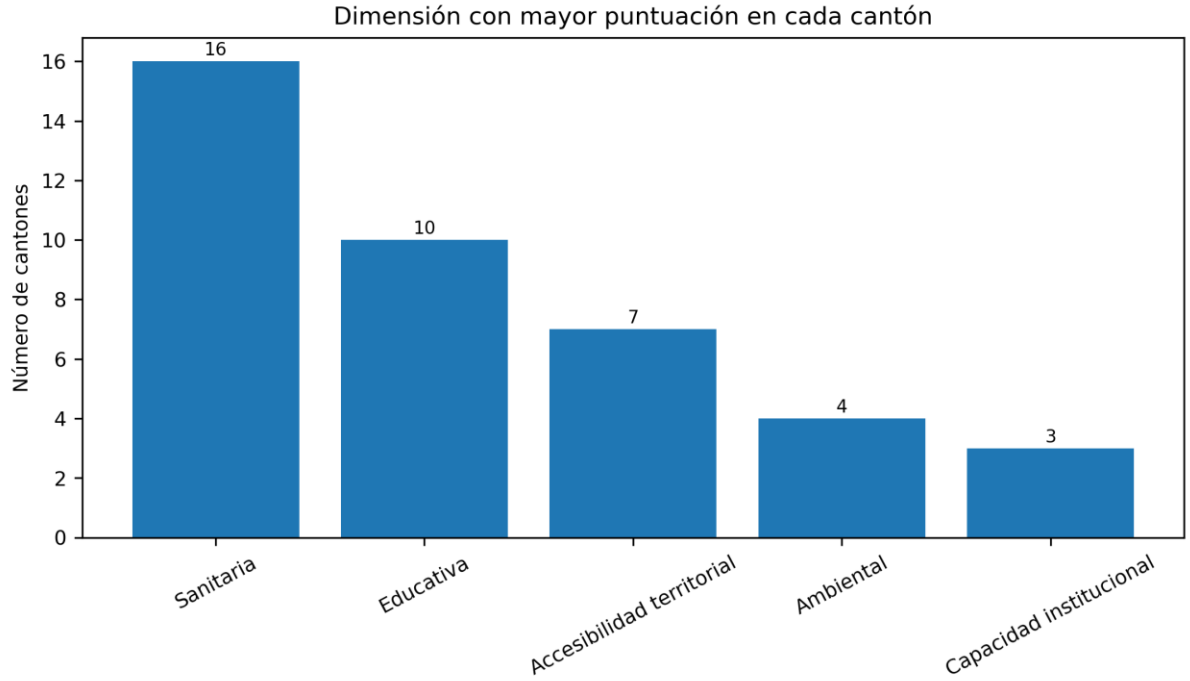
### 2.3.3. Composición e interpretabilidad del índice

La ponderación equilibrada permite descomponer cada puntuación global de manera exacta: la contribución de una dimensión equivale a su subíndice dividido para seis. La suma de las seis contribuciones reproduce el índice global, lo que garantiza consistencia aritmética y facilita una interpretación transparente. De este modo, la clasificación no funciona como una salida de caja negra, porque es posible identificar qué dimensiones sostienen la posición de cada cantón.

La dimensión de mayor puntuación fue la sanitaria en 16 cantones, la educativa en 10, la accesibilidad territorial en 7, la ambiental en 4 y la capacidad institucional en 3. La dimensión sociodemográfica no fue dominante en ningún cantón, aunque alcanzó valores muy elevados en varios de los primeros puestos. Esto no significa que carezca de importancia; en territorios como Taisha, Tiwintza y Huamboya, su aporte es alto, pero es superado por salud o accesibilidad. La frecuencia de dominancia debe leerse junto con los valores absolutos de cada subíndice.

**Figura 5**

*Dimensión con mayor puntuación en cada cantón*



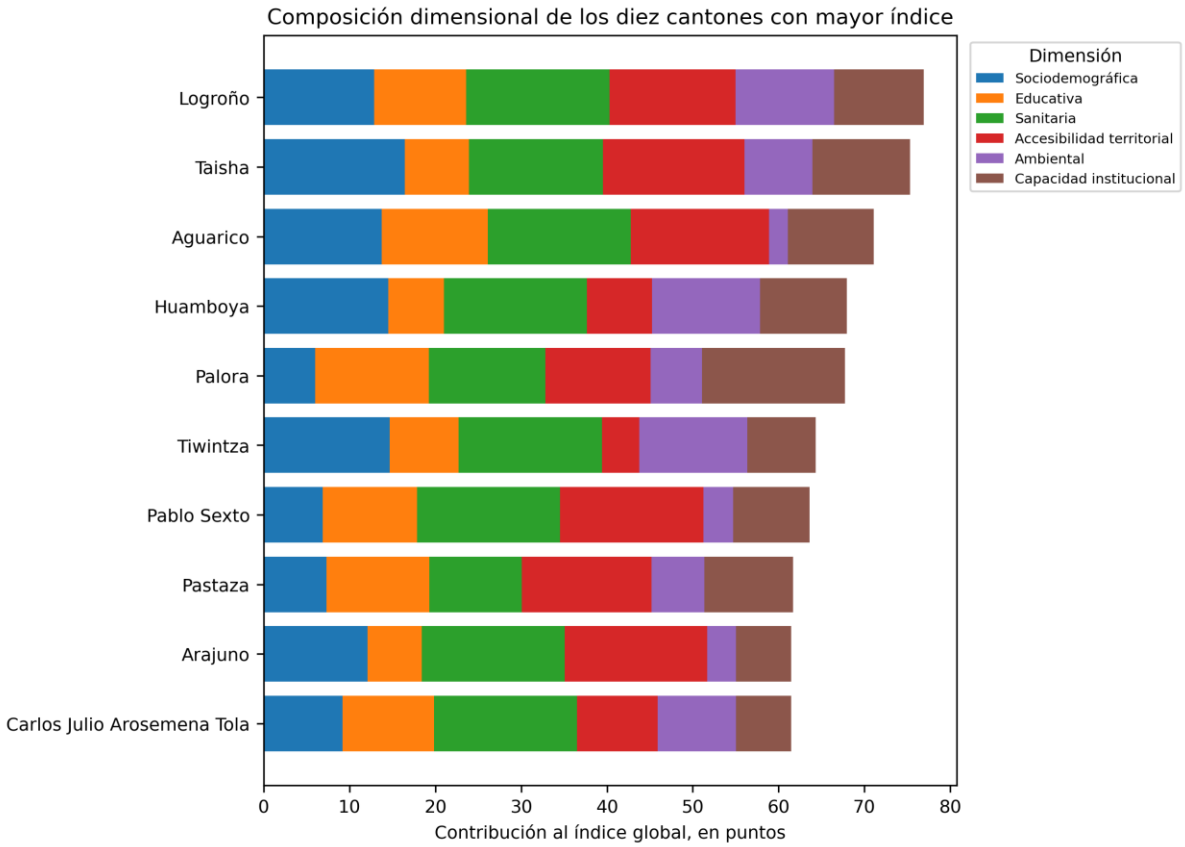

Entre los cantones de mayor índice se observan configuraciones claramente distintas. Logroño combina valores altos en las seis dimensiones y alcanza 100 en salud, 88,23 en accesibilidad, 77,25 en condiciones sociodemográficas y 68,76 en ambiente. Taisha presenta una combinación marcada por accesibilidad (99,13), condiciones sociodemográficas (98,74) y salud (93,63). Aguarico alcanza 100 en salud y valores superiores a 74 en educación y 96 en accesibilidad, pero 13,15 en ambiente. Palora en cambio, se ubica quinta debido principalmente a capacidad institucional (100), salud (81,46), educación (79,47) y

accesibilidad (73,65), pese a tener puntuaciones menores en las dimensiones sociodemográfica y ambiental.

Estas diferencias tienen implicaciones para la priorización. Dos cantones con puntuaciones cercanas no necesariamente requieren las mismas respuestas. Una intervención centrada en acceso a servicios podría ser pertinente para un territorio del Perfil 2, mientras otro cantón con presión ambiental e institucional requeriría una combinación diferente de acompañamiento, incidencia y fortalecimiento de capacidades. Por ello, la puntuación global debe utilizarse como puerta de entrada al diagnóstico y no como sustituto de la lectura territorial.

**Figura 6**

*Composición dimensional de los diez cantones con mayor índice*




### 2.3.4. Perfiles territoriales identificados mediante clustering

El clustering produjo tres perfiles de tamaños equilibrados: 14 cantones en el Perfil 1 y 13 en cada uno de los Perfiles 2 y 3. Esta distribución reduce el riesgo de que la solución esté determinada por un grupo residual de pocos casos y permite caracterizar configuraciones territoriales recurrentes. No obstante, los perfiles son categorías descriptivas construidas a partir de similitudes multivariadas; no constituyen etapas de desarrollo ni niveles ordenados de vulnerabilidad.

La lectura conjunta de las métricas impide interpretar la solución como una segmentación inequívoca del territorio. El valor de silhouette muestra que existe solapamiento entre los perfiles, mientras que la estabilidad obtenida por submuestreo indica que la asignación presenta una consistencia comparativamente alta frente a las perturbaciones

evaluadas. En consecuencia, los perfiles se utilizaron como tipología exploratoria y descriptiva y no como clasificación definitiva de los cantones. Los resultados comparativos completos de las configuraciones evaluadas se encuentran disponibles en el Anexo F.

**Tabla 7**

*Promedios dimensionales de los perfiles territoriales*

| Perfil | Sociodemográfica | Educativa | Sanitaria | Accesibilidad | Ambiental | Capacidad institucional | Índice medio |
|---|---|---|---|---|---|---|---|
| Perfil 1 | 38,10 | 60,09 | 42,92 | 34,28 | 57,08 | 40,99 | 45,58 |
| Perfil 2 | 47,87 | 62,77 | 70,34 | 75,55 | 25,76 | 54,90 | 56,20 |
| Perfil 3 | 63,70 | 47,61 | 92,61 | 58,97 | 57,08 | 51,40 | 61,90 |

El Perfil 1 presenta un índice medio de 45,58 y se distingue por una presión ambiental relativa elevada (57,08) junto con menores puntuaciones en accesibilidad (34,28), salud (42,92) y condiciones sociodemográficas (38,10). En términos comparativos, reúne territorios con mejor acceso relativo, pero con señales ambientales o extractivas que no deben quedar ocultas por su puntuación global más baja. Esta configuración muestra que la menor vulnerabilidad agregada puede coexistir con presiones sectoriales importantes.

El Perfil 2 registra un índice medio de 56,20. Sus rasgos principales son las brechas de accesibilidad (75,55), salud (70,34) y educación (62,77), combinadas con la menor puntuación ambiental media de los tres perfiles (25,76). Se trata de una configuración en la que la distancia a servicios y articulación territorial tienen mayor peso que la presión ambiental medida. Aguarico, Arajuno, Pastaza y Pablo Sexto ilustran este patrón, aunque cada uno presenta una combinación particular.

El Perfil 3 alcanza el mayor índice medio, 61,90, y combina vulnerabilidad sanitaria muy alta (92,61), condiciones sociodemográficas elevadas (63,70) y presión ambiental de 57,08. Es el perfil más cercano a una acumulación multidimensional de desventajas. Sin

embargo, incluso dentro de este grupo existen contrastes: Loreto tiene capacidad institucional normalizada en cero, mientras Palanda registra 77,74; Tiwintza presenta 75,69 en ambiente y 26,05 en accesibilidad. La pertenencia al perfil no elimina la necesidad de revisar el detalle cantonal.

La solución seleccionada correspondió a K-means aplicado sobre los tres componentes del PCA reducido de 11 indicadores, con tres grupos. Esta configuración registró un *silhouette score* de 0,303, un índice Davies–Bouldin de 1,117, un criterio Calinski–Harabasz de 19,284 y una estabilidad media medida mediante el índice de Rand ajustado de 0,847 en las pruebas de submuestreo. En conjunto, estos resultados respaldaron su selección como solución exploratoria para la caracterización territorial, considerando simultáneamente calidad interna, estabilidad, equilibrio e interpretabilidad (véase Anexo F).

**Figura 7**

*Configuración dimensional media de los perfiles territoriales*

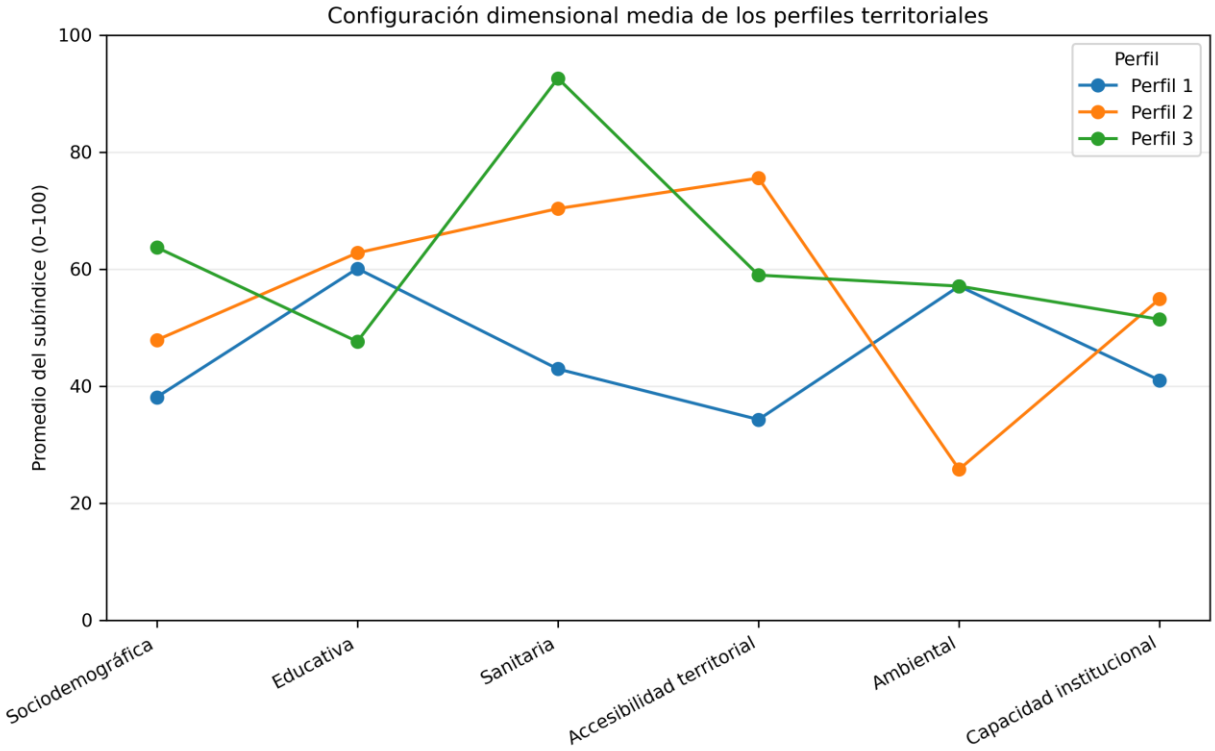
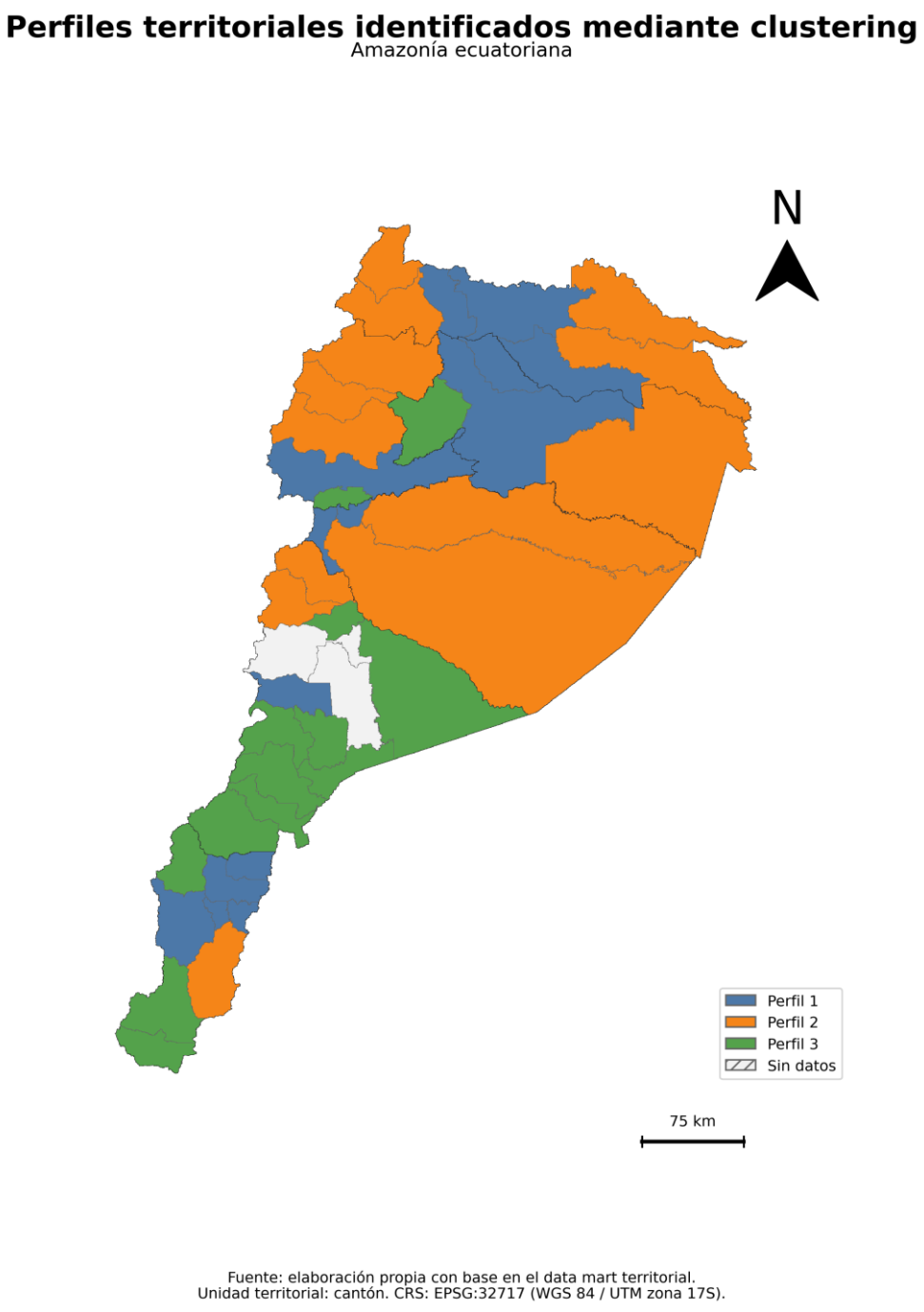


**Figura 8**

*Distribución territorial de los perfiles identificados mediante clustering.*




*Nota.* Unidad territorial: cantón; CRS EPSG:32717 (WGS 84 / UTM zona 17S). Los perfiles representan configuraciones multidimensionales y no niveles ordinales.

### 2.3.5. Comparación de ponderaciones y sensibilidad metodológica

La robustez del índice se examinó comparando la ponderación equilibrada con una alternativa derivada de las cargas de un análisis de componentes principales con 13 indicadores. La correlación de Spearman entre ambos rankings fue 0,814, lo que indica una convergencia general en el ordenamiento. Sin embargo, el cambio promedio fue de 5,65 posiciones, el máximo alcanzó 16 y el acuerdo exacto entre categorías fue de 50%. En otras palabras, la estructura general es relativamente estable, pero la decisión de ponderación modifica de forma sustantiva la posición y la categoría de varios cantones.

**Tabla 8**

*Pesos dimensionales comparados*

| Dimensión | Método equilibrado (%) | PCA-13 (%) |
|---|---|---|
| Sociodemográfica | 16,67 | 13,49 |
| Educativa | 16,67 | 23,50 |
| Sanitaria | 16,67 | 15,48 |
| Accesibilidad territorial | 16,67 | 7,85 |
| Ambiental | 16,67 | 37,07 |
| Capacidad institucional | 16,67 | 2,62 |

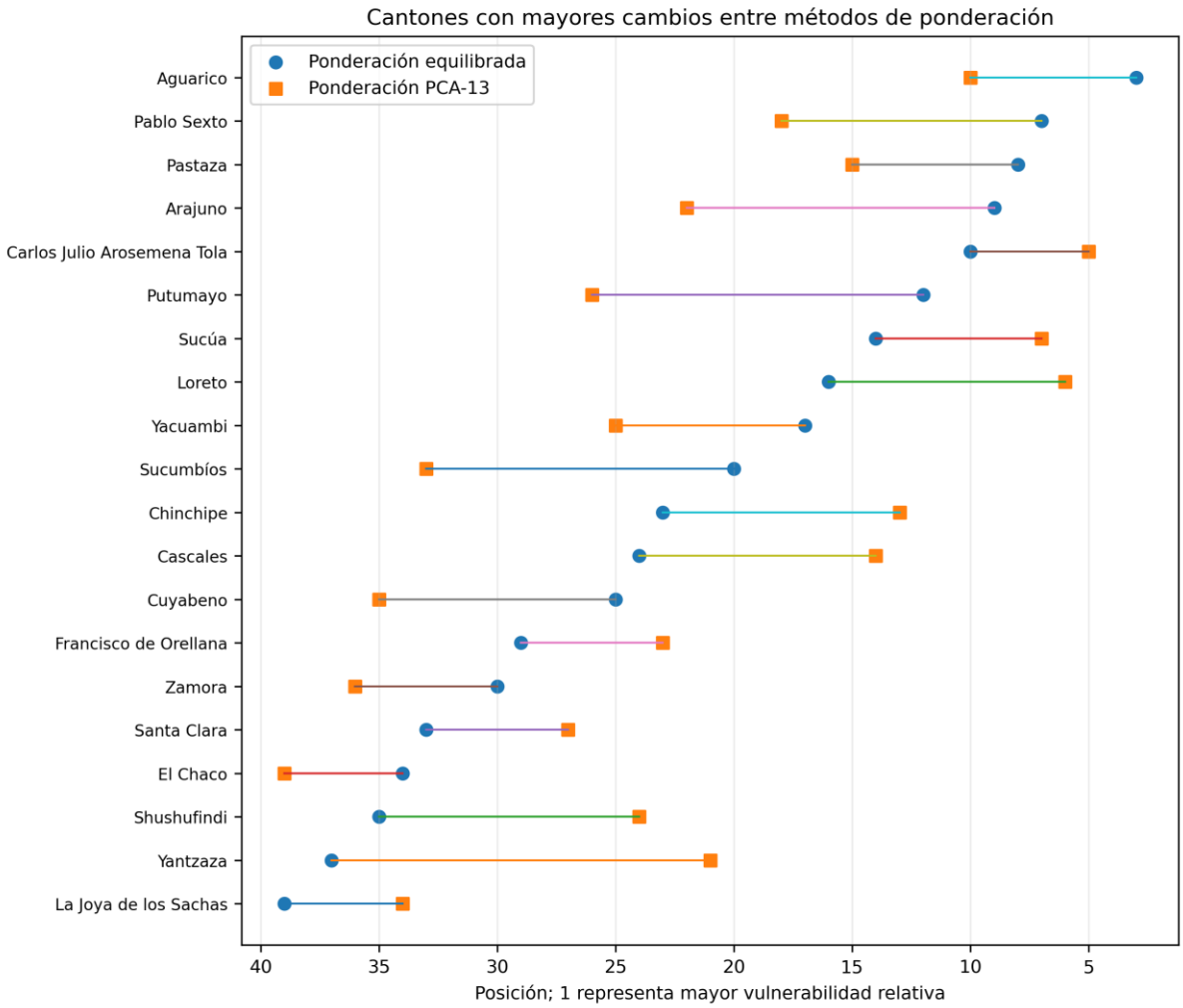
La principal diferencia reside en la concentración del método PCA: la dimensión ambiental recibe 37,07% del peso total y la educativa 23,50%, mientras la capacidad institucional desciende a 2,62% y la accesibilidad a 7,85%. Esta estructura refleja la variabilidad y correlación estadística de los datos, no necesariamente la importancia normativa de cada dimensión. Por ello se mantuvo la ponderación equilibrada como versión principal y se conservó el PCA como análisis de sensibilidad, en concordancia con las recomendaciones de transparencia y examen de incertidumbre en índices compuestos (Organisation for Economic Co-operation and Development & European Commission, Joint Research Centre, 2008; Saltelli, 2007).

Los mayores desplazamientos permiten observar el efecto práctico de la ponderación. Yantzaza pasa del puesto 37 con pesos equilibrados al 21 con PCA, porque el método

alternativo otorga mayor peso a las dimensiones ambiental y educativa, donde el cantón alcanza 58,45 y 69,95. En sentido contrario, Putumayo desciende del puesto 12 al 26, Arajuno del 9 al 22 y Sucumbíos del 20 al 33. En estos casos, los valores elevados de capacidad institucional, accesibilidad o salud pierden influencia relativa bajo la estructura de pesos del PCA.

**Figura 9**

*Cantones con mayores cambios entre métodos de ponderación*




La evaluación final de sensibilidad combinó el cambio frente al PCA, la exclusión sucesiva de dimensiones, la estabilidad del clustering, la perturbación simulada de pesos y la cercanía a los puntos de corte de categoría. Once cantones presentaron sensibilidad

metodológica alta. Los intervalos de posición derivados de las simulaciones no son intervalos de confianza muestral; representan el rango de resultados que puede producirse cuando se modifican decisiones normativas de ponderación dentro de escenarios plausibles.

**Tabla 9**

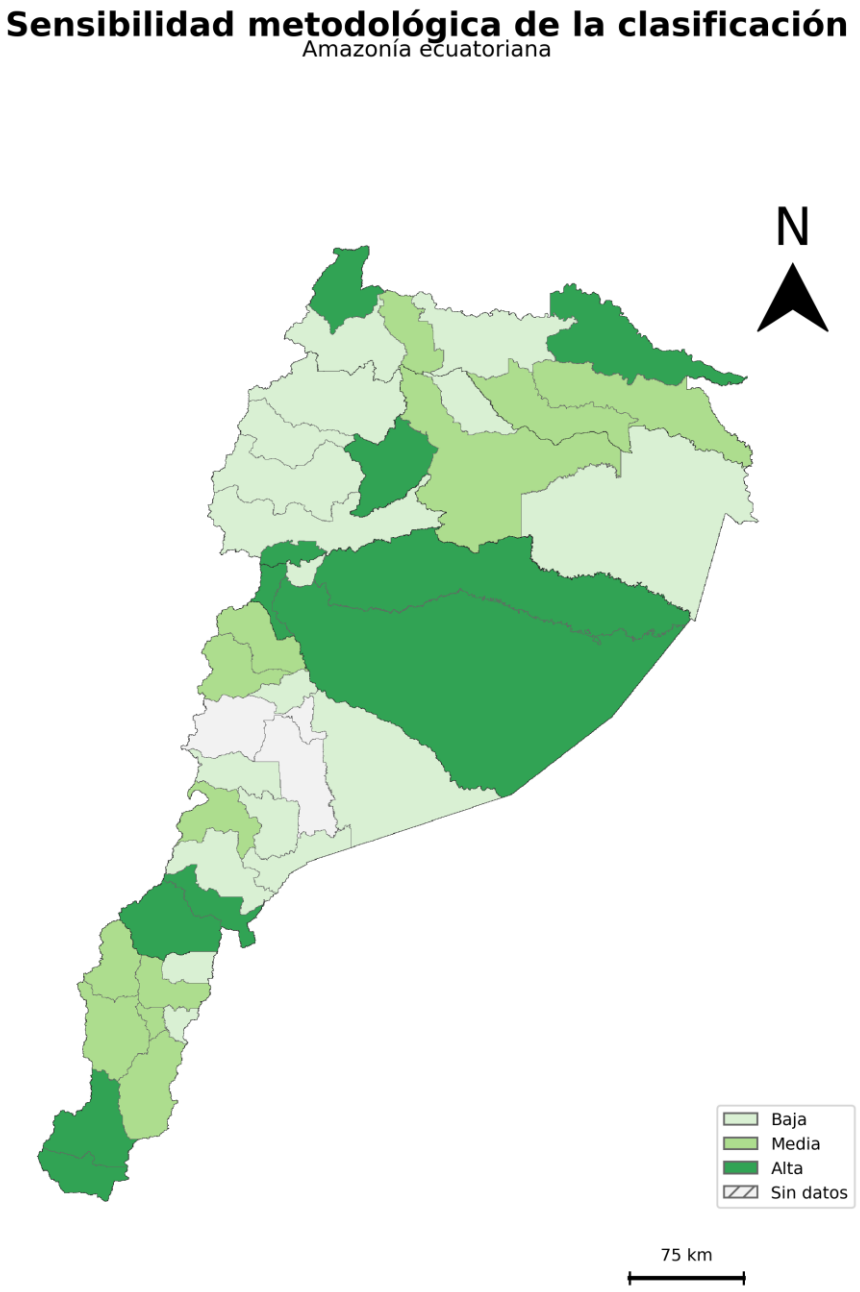
*Cantones con sensibilidad metodológica alta*

| Cantón | Alertas | Rango 5 % | Rango 95 % | Permanencia en categoría modal (%) | Categoría modal |
|---|---|---|---|---|---|
| Sucumbíos | 5 | 12 | 27 | 57,70 | Alta |
| Arajuno | 5 | 6 | 18 | 56,10 | Muy alta |
| Putumayo | 5 | 7 | 17 | 54,60 | Alta |
| Loreto | 4 | 9 | 25 | 65,50 | Alta |
| Chinchipe | 4 | 15 | 28 | 58,50 | Media |
| San Juan Bosco | 4 | 16 | 26 | 53,30 | Media |
| Mera | 3 | 10 | 25 | 57,50 | Alta |
| Pastaza | 3 | 6 | 16 | 61,70 | Muy alta |
| Gualaquiza | 3 | 17 | 24 | 58,80 | Media |
| Carlos Julio Arosemena Tola | 3 | 7 | 14 | 59,80 | Muy alta |
| Palanda | 3 | 7 | 14 | 57,70 | Muy alta |

Sucumbíos, Arajuno y Putumayo acumularon cinco alertas cada uno. La amplitud de sus rangos y probabilidades modales cercanas a 55% muestra que una comunicación responsable no debería presentar su posición o categoría como un resultado rígido. En cambio, los cantones ubicados en los extremos del ranking y con baja sensibilidad pueden considerarse más robustos para fines de priorización inicial. La sensibilidad no invalida el índice; permite distinguir qué resultados son más robustos frente a cambios metodológicos y cuáles requieren mayor cautela interpretativa.

**Figura 10**

*Distribución territorial de la sensibilidad metodológica*




*Nota.* Unidad territorial: cantón; CRS EPSG:32717 (WGS 84 / UTM zona 17S). La sensibilidad metodológica se interpreta separadamente de la vulnerabilidad.

### 2.3.6. Patrón espacial del índice

El índice presentó un I de Moran global de 0,397, con un valor *p* bilateral de permutación de 0,005. Este resultado indica autocorrelación espacial positiva: los cantones con puntuaciones semejantes tienden a localizarse próximos entre sí con mayor frecuencia de la esperada bajo una distribución espacial aleatoria. El patrón es moderado y no implica que todos los cantones formen agrupamientos significativos ni que la proximidad territorial sea una causa suficiente de vulnerabilidad.

El análisis local identificó tres asociaciones Alto-Alto en Morona Santiago —Huamboya, Santiago y Tiwintza— y una asociación Bajo-Bajo en Cascales. Los patrones Alto-Alto corresponden a cantones con valores elevados rodeados por vecinos que también registran puntuaciones altas; el patrón Bajo-Bajo señala una concentración local de valores comparativamente menores. La interpretación sigue la lógica de los indicadores locales de asociación espacial de Anselin (1995).

**Tabla 10**

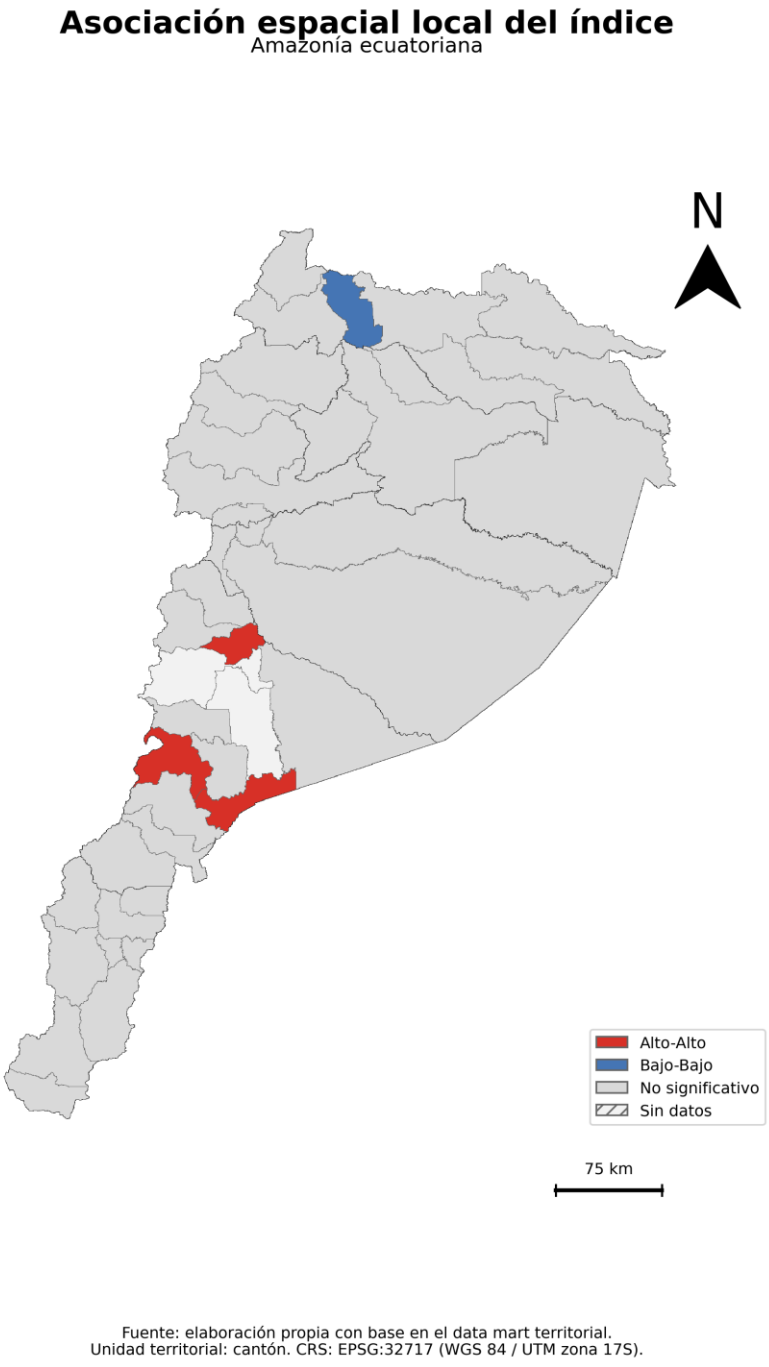
*Asociaciones espaciales locales significativas*

| Cantón | Provincia | Patrón LISA | *p* de permutación |
|---|---|---|---|
| Huamboya | Morona Santiago | Alto-Alto | 0,01 |
| Santiago | Morona Santiago | Alto-Alto | 0,04 |
| Tiwintza | Morona Santiago | Alto-Alto | 0,01 |
| Cascales | Sucumbíos | Bajo-Bajo | 0,03 |

La concentración Alto-Alto en el centro-sur de Morona Santiago refuerza la necesidad de no interpretar los cantones como unidades aisladas. La movilidad, la provisión regional de servicios, la continuidad de ecosistemas, las redes institucionales y las dinámicas extractivas pueden atravesar límites administrativos. Desde una perspectiva relacional del territorio, el espacio se configura mediante relaciones sociales y conexiones desiguales, en lugar de funcionar como un soporte neutral (Massey, 1994). Por ello, la intervención podría combinar acciones cantonales con estrategias supracantonales o provinciales.

**Figura 11**

*Asociación espacial local del Índice de Vulnerabilidad Socioambiental*




*Nota.* Unidad territorial: cantón; CRS EPSG:32717 (WGS 84 / UTM zona 17S). Se representan las asociaciones espaciales locales estadísticamente significativas y los cantones sin asociación significativa.

### 2.3.7. Lectura socioterritorial e implicaciones para la priorización

Los resultados muestran que la vulnerabilidad socioambiental amazónica no se organiza mediante una única división entre territorios favorecidos y desfavorecidos. Existen acumulaciones multidimensionales, como en Logroño, Taisha y Huamboya; configuraciones dominadas por accesibilidad y servicios, como en Arajuno, Aguarico o Pastaza; y casos con posiciones globales menores, pero presiones ambientales relevantes, como La Joya de los Sachas, Centinela del Cóndor o Lago Agrio. Esta diversidad exige respuestas diferenciadas y cuestiona el uso del ranking como mecanismo automático de asignación de recursos.

La accesibilidad territorial ocupa un lugar central porque conecta la geografía física con la organización social de los servicios. Un valor alto puede expresar tiempos de desplazamiento, dispersión de asentamientos, baja articulación vial o fluvial y dependencia de centros externos. Estas condiciones inciden en el acceso efectivo a salud, educación, administración pública y mercados, pero sus efectos no son uniformes: las comunidades pueden desarrollar redes propias, estrategias de movilidad y formas de organización que no quedan completamente captadas por los indicadores cuantitativos.

La dimensión institucional también requiere una lectura relacional. Las capacidades de los gobiernos locales no dependen exclusivamente de decisiones internas; están condicionadas por la estructura fiscal, la extensión del territorio, la densidad poblacional, las competencias asignadas y la presencia de otras entidades públicas, eclesiales y comunitarias. Una puntuación institucional alta en el índice de vulnerabilidad señala una brecha relativa de capacidad operativa, no una descalificación de los actores locales.

En el ámbito ambiental, los patrones sugieren que la presión extractiva y la transformación territorial pueden coexistir tanto con brechas sociales altas como con mejores condiciones relativas de acceso. Esto es consistente con la idea de que la vulnerabilidad surge de la interacción entre exposición, sensibilidad y capacidad de respuesta, y no de un único

factor ambiental (Adger, 2006). La dimensión ambiental debe interpretarse junto con los usos del suelo, la presencia de actividades extractivas, la cobertura vegetal y las capacidades sociales e institucionales para incidir sobre esas dinámicas.

Como hipótesis de uso institucional, los resultados permiten proponer una matriz preliminar de lectura basada en cuatro componentes: puntuación global, dimensión crítica, perfil territorial y sensibilidad metodológica. Esta propuesta no fue validada mediante una implementación institucional ni con usuarios finales, por lo que debe entenderse como un marco analítico a contrastar mediante procesos participativos antes de utilizarlo para asignar recursos o definir intervenciones.

A estos elementos debe añadirse el conocimiento situado de jurisdicciones eclesiales, organizaciones, gobiernos locales, comunidades y pueblos indígenas. Un cantón con índice alto y baja sensibilidad podría considerarse candidato para una revisión prioritaria, que debería complementarse con información cualitativa y conocimiento territorial antes de adoptar una decisión operativa; otro con índice intermedio, presión ambiental alta y relevancia territorial estratégica puede requerir una respuesta distinta; y un caso con alta sensibilidad debe someterse a verificación cualitativa antes de adoptar decisiones de asignación.

Las variables interculturales se mantuvieron como información contextual y no como indicadores directos de vulnerabilidad. Esta decisión evita convertir la identidad indígena, la presencia de pueblos y nacionalidades o la territorialidad comunitaria en atributos negativos. Su incorporación futura debe realizarse mediante procesos de interpretación participativa, respeto a la autoridad sobre los datos y búsqueda de beneficios colectivos, de acuerdo con los principios CARE de gobernanza de datos indígenas (Carroll et al., 2020).

### 2.3.8. Discusión de resultados en relación con los antecedentes y las decisiones metodológicas

A través del análisis realizado, se pudo comprobar que los resultados obtenidos coinciden con los antecedentes revisados en señalar que la vulnerabilidad territorial no puede ser representada adecuadamente mediante un único indicador sectorial. Al igual que el Social Vulnerability Index de Cutter et al. (2003) y el índice cantonal desarrollado para Ecuador por Fernandez et al. (2015), el presente estudio encontró diferencias territoriales que responden a la combinación de varias dimensiones y no exclusivamente a una característica demográfica, ambiental o institucional.

La principal diferencia metodológica reside en la estrategia de agregación: mientras esos antecedentes utilizaron procedimientos estadísticos para reducir o ponderar variables, el índice principal de este estudio adoptó una ponderación equilibrada entre dimensiones y reservó el PCA como referencia exploratoria y de sensibilidad. Esta elección no implica que los pesos iguales sean estadísticamente óptimos, sino que constituyen una regla normativa transparente que evita que la varianza o el número desigual de indicadores determine por sí mismo la importancia sustantiva de cada dimensión.

Los resultados también respaldan la precaución planteada por Mansur et al. (2016) respecto del efecto de la escala territorial. En su estudio amazónico, una medida agregada podía ocultar heterogeneidades internas aun trabajando con unidades censales considerablemente menores. En el presente proyecto, esta limitación es especialmente relevante porque el cantón fue la unidad mínima compatible entre las fuentes prioritarias. En consecuencia, las diferencias observadas entre cantones no permiten inferir homogeneidad dentro de ellos, particularmente en territorios extensos, con asentamientos dispersos, movilidad fluvial o fuertes diferencias entre cabeceras y comunidades rurales. A diferencia de Mansur et al. (2016), tampoco resulta apropiado comparar proporciones de población

clasificadas como vulnerable, porque las categorías utilizadas aquí corresponden a cuartiles relativos dentro de los 40 cantones modelados y no a umbrales absolutos de vulnerabilidad.

En cuanto al aprendizaje automático, los perfiles identificados son consistentes con la función complementaria atribuida al aprendizaje no supervisado por Abarca-Alvarez et al. (2019): reconocer configuraciones multivariadas que un índice agregado puede ocultar. Sin embargo, las métricas obtenidas requieren una interpretación matizada. El silhouette score de 0,303 indica una separación limitada o moderada entre agrupamientos, mientras que el ARI medio de 0,847 en las pruebas de submuestreo muestra una estabilidad comparativamente alta frente a las perturbaciones evaluadas. Por ello, los tres perfiles no deben presentarse como clases naturales o definitivas de la Amazonía, sino como una solución exploratoria relativamente estable dentro del conjunto de representaciones y algoritmos comparados.

El análisis de sensibilidad refuerza las advertencias formuladas por la OECD y el Joint Research Centre (2008) y por Saltelli (2007) acerca de la influencia de las decisiones de ponderación en los índices compuestos. Aunque la correlación de Spearman entre el ranking equilibrado y el derivado del PCA alcanzó 0,814, el desplazamiento medio fue de 5,65 posiciones y el acuerdo exacto entre categorías fue de 50 %. Por tanto, la convergencia general entre métodos no elimina la incertidumbre metodológica a escala cantonal. Este resultado justifica conservar la ponderación equilibrada como escenario principal por su transparencia conceptual, pero no permite considerarla superior en términos absolutos.

Finalmente, el patrón espacial aporta una dimensión adicional, pero tampoco debe sobreinterpretarse. El I de Moran de 0,397 y el valor *p* de permutación de 0,005 evidencian asociación espacial positiva en el conjunto analizado, mientras que las asociaciones locales significativas se limitaron a tres patrones Alto-Alto y uno Bajo-Bajo. Esto sugiere continuidad espacial en determinados sectores, particularmente en Morona Santiago, pero no demuestra la existencia de un mecanismo causal común ni autoriza a generalizar esa

configuración al conjunto provincial. En este sentido, el análisis espacial funciona como evidencia complementaria para formular hipótesis territoriales que requerirían contraste cualitativo o longitudinal.

### 2.3.9. Alcances, desafíos y oportunidades de fortalecimiento

El índice constituye una herramienta analítica reproducible para comparar territorios y apoyar procesos de priorización, pero no debe presentarse como una medición definitiva ni como sustituto de diagnósticos locales. Su principal fortaleza radica en integrar información heterogénea, conservar la trazabilidad de las transformaciones y hacer explícitas las decisiones de ponderación y sensibilidad. Sus limitaciones deben leerse como condiciones de uso y como agenda de fortalecimiento.

En primer lugar, las categorías son relativas a los 40 cantones modelados. Un territorio clasificado en nivel bajo puede mantener brechas relevantes, mientras un nivel muy alto significa una posición comparativamente elevada y no un umbral universal de emergencia. En segundo lugar, la escala cantonal puede ocultar desigualdades internas entre cabeceras, parroquias, comunidades ribereñas y zonas de difícil acceso. Una futura aplicación parroquial permitiría examinar esta heterogeneidad siempre que la disponibilidad y compatibilidad de datos eviten producir una falsa desagregación.

En tercer lugar, las fuentes no corresponden necesariamente al mismo período. El índice integra la mejor información compatible disponible, pero no representa una fotografía temporal perfectamente simultánea. La actualización periódica del data mart permitiría registrar cambios, evaluar tendencias y distinguir transformaciones estructurales de variaciones asociadas a la renovación de las fuentes. En cuarto lugar, algunos registros administrativos reflejan simultáneamente necesidad, cobertura institucional y capacidad de registro, por lo que deben interpretarse con cautela.

La normalización también introduce desafíos. El efecto de techo observado en salud genera numerosos empates y reduce la capacidad de discriminación interna de esa dimensión. En próximas versiones conviene revisar los indicadores sanitarios, explorar transformaciones alternativas o incorporar variables con mayor gradiente territorial. De igual manera, la diferencia entre los pesos equilibrados y PCA demuestra que ningún sistema de ponderación es neutral. La publicación conjunta de las pruebas de sensibilidad es necesaria para que los usuarios conozcan qué resultados son estables y cuáles dependen más de las decisiones metodológicas.

La validación territorial constituye la siguiente oportunidad de fortalecimiento. Talleres o entrevistas con actores locales podrían contrastar si los perfiles estadísticos coinciden con la experiencia de quienes gestionan servicios, acompañan comunidades y participan en procesos educativos, pastorales o socioambientales. Esta validación no debe limitarse a confirmar el índice, sino identificar variables ausentes, interpretaciones equivocadas y factores protectores que los datos secundarios no captan.

Finalmente, el índice puede ampliarse con información sobre tiempos reales de desplazamiento, movilidad fluvial, conectividad digital, capacidad fiscal, exposición climática, conflictos socioambientales y disponibilidad efectiva de servicios. La incorporación debe responder a criterios conceptuales, éticos y de calidad, no al simple aumento del número de variables. El objetivo de una versión futura no sería producir un índice más complejo, sino uno más pertinente, validado y útil para decisiones transparentes.

En síntesis, los resultados permiten identificar prioridades sin borrar la diversidad territorial. El ranking señala posiciones relativas; los subíndices explican su composición; los perfiles muestran configuraciones recurrentes; el análisis espacial identifica concentraciones; y la sensibilidad delimita la robustez de las clasificaciones. La combinación de estos componentes ofrece una base más sólida para el diálogo institucional y la planificación,

siempre que se mantenga abierta a la revisión, la actualización de datos y la interpretación situada.

**Tabla 11**

*Correspondencia entre objetivos específicos, procedimiento, evidencia y conclusión*

| Objetivo específico | Procedimientos principales | Evidencia de cumplimiento | Conclusión asociada |
|---|---|---|---|
| **OE1. Construir una base territorial cantonal consolidada** | Secciones 2.2.2–2.2.5: adquisición, ETL, data mart y homologación territorial | Data mart documentado; 35 productos procesados; diccionario de 924 variables canónicas; base territorial de 42 cantones; Anexos A, B, C, I y J. | El objetivo se cumplió, aunque la incompatibilidad Morona–Sevilla Don Bosco restringió a 40 cantones la matriz de modelado. |
| **OE2. Construir el índice e identificar perfiles** | Secciones 2.2.6–2.2.9 | 34 indicadores candidatos; 13 seleccionados; seis dimensiones; índice 0–100; tres perfiles territoriales; Tablas 3–7 y Anexos D–F. | El índice compuesto y los perfiles permiten dos lecturas complementarias: ordenamiento relativo y configuración multidimensional. |
| **OE3. Evaluar consistencia, estabilidad e interpretabilidad** | Sección 2.2.10 | Reconstrucción aritmética exacta; comparación de ponderaciones; sensibilidad; estabilidad del clustering; análisis espacial; Tablas 8–10 y Anexos G–I. | El índice presenta consistencia interna e interpretabilidad, pero la robustez varía por cantón y no se realizó validación externa independiente. |

# 3. Capítulo 3. Conclusiones y Recomendaciones

## 3.1. Conclusiones

El Proyecto de Titulación alcanzó su objetivo general al construir un índice cantonal de vulnerabilidad socioambiental para la Amazonía ecuatoriana y complementarlo con perfiles territoriales obtenidos mediante aprendizaje automático no supervisado. El producto resultante integra dimensiones sociales, educativas, sanitarias, territoriales, ambientales e institucionales y ofrece una base analítica para apoyar procesos de priorización. Su principal valor no radica únicamente en ordenar cantones, sino en hacer visible la composición de cada puntuación y las diferencias entre configuraciones territoriales.

El primer objetivo específico se cumplió mediante la construcción de una base territorial cantonal consolidada y de un data mart documentado. La integración requirió limpiar, homologar y agregar fuentes censales, administrativas y geoespaciales que presentaban diferentes formatos, períodos y unidades territoriales. La separación entre datos crudos, productos procesados, base integrada, indicadores, modelado, visualizaciones y código permitió conservar la procedencia de los archivos y reconstruir las transformaciones realizadas. Esta arquitectura constituye uno de los aportes más relevantes del proyecto porque convierte un conjunto disperso de fuentes en una infraestructura reproducible de información territorial.

La experiencia de integración confirmó que la calidad de un modelo de ciencia de datos depende tanto del algoritmo como del trabajo previo de gestión de datos. La incompatibilidad entre Morona y Sevilla Don Bosco demostró que las divisiones político-administrativas pueden cambiar con mayor rapidez que las fuentes estadísticas. La decisión de no dividir ni imputar valores preservó la integridad del análisis, aunque redujo la muestra modelada a 40 cantones. La ausencia de puntuación para estos dos territorios representa una limitación explícita y metodológicamente preferible a la generación de datos artificiales.

El segundo objetivo específico se alcanzó mediante la selección de 13 indicadores y su organización en seis dimensiones. La ponderación equilibrada permitió que cada dimensión tuviera la misma participación conceptual, evitando que aquellas con más indicadores o mayor varianza dominaran automáticamente el resultado. El índice principal no fue producido por un algoritmo de aprendizaje automático en sentido estricto: fue construido como un índice compuesto aditivo. El aprendizaje automático no supervisado se utilizó para explorar la estructura de los datos e identificar perfiles territoriales. Esta distinción fortalece la claridad metodológica del trabajo y evita atribuir al algoritmo decisiones que fueron conceptuales y normativas.

Los resultados evidenciaron que la vulnerabilidad socioambiental no presenta una distribución uniforme. Logroño, Taisha y Aguarico encabezaron el ranking, pero sus puntuaciones respondieron a combinaciones diferentes. La concentración de cantones de Morona Santiago en posiciones altas constituye un patrón territorial relevante, aunque no implica homogeneidad provincial ni demuestra una causa común. En el extremo inferior, los cantones conservaron vulnerabilidades específicas que no siempre se reflejan en la puntuación global. Por ello, la categoría baja debe interpretarse como una posición relativa dentro de la muestra y no como ausencia de problemas socioambientales.

La descomposición en subíndices confirmó que el valor global no es suficiente para orientar intervenciones. La dimensión sanitaria fue dominante en el mayor número de cantones, seguida por la educativa y la accesibilidad territorial. La ambiental presentó una relación más débil con el índice global, lo que muestra que las presiones ambientales no coinciden necesariamente con las brechas sociales o de servicios. Estas diferencias respaldan la necesidad de formular respuestas diferenciadas: un territorio con limitaciones sanitarias y de accesibilidad no requiere el mismo tipo de intervención que otro cuya principal presión se relaciona con transformación ambiental o capacidad institucional.

Los tres perfiles obtenidos mediante clustering aportaron una lectura complementaria. El Perfil 1 combinó mejores condiciones relativas de acceso con presiones ambientales relevantes; el Perfil 2 estuvo marcado por brechas de accesibilidad y servicios; y el Perfil 3 presentó una acumulación más intensa de vulnerabilidad sanitaria, sociodemográfica y ambiental. Estos agrupamientos fueron de tamaños equilibrados y presentaron una separación interna moderada y una estabilidad alta frente al esquema de submuestreo aplicado, por lo que se interpretaron como una tipología exploratoria y no como clases territoriales definitivas. Su utilidad consiste en reconocer configuraciones recurrentes y apoyar el diseño de intervenciones adaptadas a tipos de territorio.

El tercer objetivo específico se cumplió mediante la evaluación de consistencia, estabilidad, sensibilidad e interpretabilidad. La suma de las contribuciones reprodujo exactamente la puntuación global, lo que permitió explicar de forma directa el resultado de cada cantón. La comparación con la ponderación PCA mostró una correlación general alta, pero también desplazamientos importantes y un acuerdo de categorías limitado. Once cantones presentaron sensibilidad metodológica alta. Estos resultados demuestran que la incertidumbre no debe ocultarse: forma parte de la información necesaria para decidir con qué grado de confianza puede utilizarse cada posición. La evaluación realizada establece consistencia y robustez internas, pero no constituye una validación externa del índice, debido a la inexistencia de una clasificación independiente de vulnerabilidad contra la cual contrastarlo.

El análisis espacial mostró que las puntuaciones no se distribuyen aleatoriamente. La autocorrelación global positiva y las asociaciones Alto-Alto identificadas en Morona Santiago indican que algunas vulnerabilidades presentan continuidad entre cantones vecinos. Esto sugiere que determinadas respuestas no deberían restringirse a intervenciones cantonales

aisladas. La planificación puede requerir estrategias supracantonales relacionadas con redes de servicios, movilidad, conectividad, ecosistemas y dinámicas extractivas.

El estudio también cumplió una función ética. Las variables interculturales no fueron orientadas como carencias ni incorporadas a la puntuación. Esta decisión evitó asociar automáticamente identidad indígena y vulnerabilidad. El índice debe ser utilizado como una herramienta complementaria, sujeta a interpretación participativa, validación territorial y respeto por la autoridad de pueblos y comunidades sobre la información que los involucra.

En relación con el plan inicial, la mayor parte de los productos esperados fue alcanzada: se construyó la base territorial, el diccionario de variables, el conjunto de indicadores, el índice, la tipología de perfiles, los mapas, los análisis de sensibilidad y el visualizador territorial. Otros componentes se reformularon. No se desarrolló una versión parroquial, no se incorporó una dimensión digital independiente, no se aplicaron modelos supervisados ni SHAP y el producto de visualización se mantuvo como prototipo interactivo. Estas decisiones respondieron a la calidad real de las fuentes, al tamaño de la muestra y a la estructura del índice, y contribuyeron a evitar una complejidad técnica sin justificación analítica.

En síntesis, el proyecto muestra la viabilidad técnica de integrar ciencia de datos, sistemas de información geográfica y aprendizaje automático no supervisado para construir un índice territorial reproducible e interpretable. El índice no ofrece una verdad definitiva sobre la Amazonía ni reemplaza el diagnóstico local. Su aporte consiste en ordenar información, revelar patrones, hacer explícitas las decisiones metodológicas y ofrecer una base para procesos institucionales de priorización más transparentes.

### 3.2. Limitaciones

La primera limitación técnica se relaciona con la heterogeneidad de las fuentes. Los datos proceden de censos, registros administrativos, sistemas geoespaciales y plataformas

especializadas que utilizan períodos, formatos, escalas, definiciones y niveles de actualización diferentes. Aunque el proceso ETL permitió homologar los productos, la base final no representa una medición simultánea de todas las dimensiones. Algunas diferencias entre cantones pueden corresponder parcialmente a la fecha de actualización o a las capacidades institucionales de registro.

La modificación de la división político-administrativa constituye una limitación específica. Los datos anteriores a la creación de Sevilla Don Bosco no permitieron separar con precisión su territorio y población de los registros de Morona. La exclusión de ambos cantones del modelo evitó imputaciones arbitrarias, pero restringió la cobertura del índice. Esta situación también afectó tasas cuyo numerador y denominador dependían de configuraciones territoriales diferentes.

La escala cantonal introduce el problema de agregación espacial. Los cantones amazónicos presentan grandes extensiones y una importante heterogeneidad interna. Una puntuación única puede ocultar diferencias entre cabeceras, parroquias, comunidades ribereñas, territorios indígenas y zonas de acceso limitado. Esta limitación no puede solucionarse redistribuyendo artificialmente valores cantonales. Una versión parroquial requeriría fuentes producidas originalmente en ese nivel y con denominadores compatibles.

El tamaño de la muestra analítica constituye una restricción para las técnicas multivariadas. El modelado incluyó 40 observaciones y 13 indicadores. Aunque se aplicaron procedimientos de diagnóstico, reducción y estabilidad, este número limita la generalización de la estructura de componentes y agrupamientos. Los resultados del PCA y del clustering deben entenderse como exploratorios y específicos del conjunto cantonal estudiado, no como parámetros extrapolables a otros territorios.

El estudio no dispuso de una variable externa o clasificación oficial de vulnerabilidad que permitiera validar el índice mediante un criterio independiente. Las evaluaciones

realizadas fueron principalmente internas: consistencia aritmética, métricas de clustering, estabilidad por submuestreo, análisis de sensibilidad, comparación de ponderaciones y coherencia espacial. Estas pruebas fortalecen la confiabilidad metodológica, pero no demuestran por sí mismas validez externa o capacidad predictiva.

Los algoritmos utilizados incorporan supuestos que condicionan los resultados. K-means y el método de Ward dependen de distancias euclidianas, de la escala de las variables y del número de grupos seleccionado. K-means tiende a favorecer agrupamientos compactos, mientras que el clustering jerárquico puede ser sensible a decisiones tempranas de fusión. DBSCAN tuvo una utilidad exploratoria limitada por el tamaño reducido de la muestra y por la sensibilidad de sus parámetros. La comparación de algoritmos y las pruebas de estabilidad redujeron estos riesgos, pero no los eliminan.

La normalización Min-Max también introduce dependencia respecto de los valores extremos. Al transformar el mínimo y máximo observados en los límites de la escala, un cantón extremo puede modificar la dispersión relativa de los demás. El subíndice sanitario mostró un efecto de techo: doce cantones alcanzaron 100 puntos, lo que redujo la capacidad para distinguir diferencias internas. Por tanto, el ranking sanitario debe interpretarse mediante grupos de empate y no como una ordenación estricta.

La selección de indicadores estuvo condicionada por su disponibilidad y calidad. Algunos fenómenos complejos fueron representados mediante aproximaciones. Los establecimientos de salud por población y superficie miden disponibilidad potencial, pero no calidad, personal, equipamiento, horarios ni acceso efectivo. La densidad vial no incorpora movilidad fluvial, estado de las vías o tiempos reales de desplazamiento. La capacidad institucional se representó mediante un único índice operativo. Estas variables son útiles, pero no cubren la totalidad de sus dimensiones conceptuales.

La ponderación equilibrada responde a una decisión normativa, mientras que la ponderación PCA responde a la estructura de varianza y correlación. Ninguna de las dos es neutral. La primera otorga igual importancia a las dimensiones; la segunda puede sobrerrepresentar aquellas con mayor variabilidad estadística. La divergencia observada entre ambos rankings confirma que la priorización depende parcialmente del sistema de pesos, razón por la cual la versión PCA se mantuvo como sensibilidad y no como índice definitivo.

El análisis espacial depende de la matriz de vecindad utilizada y de los límites administrativos. Los indicadores de Moran y LISA permiten identificar asociaciones, pero no explican los mecanismos causales que las producen. Cantones no contiguos pueden compartir redes de movilidad, servicios o ecosistemas, mientras territorios vecinos pueden mantener relaciones débiles. Los patrones espaciales deben utilizarse como señales para una investigación posterior, no como evidencia causal.

El diseño transversal limita el análisis temporal. El estudio no permite determinar si la vulnerabilidad aumentó o disminuyó ni establecer relaciones de causa y efecto entre dimensiones. La incorporación de series ambientales no convierte al índice global en un modelo longitudinal, porque las demás variables corresponden a cortes específicos. Tampoco se construyó un modelo predictivo de vulnerabilidad futura.

La reproducibilidad depende de la conservación de los archivos, las versiones de las librerías y los entornos de ejecución. Google Colab facilita el acceso al procesamiento, pero sus dependencias pueden cambiar. La ausencia futura de una fuente, una actualización de formato o una modificación en una biblioteca podría afectar la ejecución. Los cuadernos, manifiestos, paquetes de respaldo y diccionarios reducen este riesgo, pero será necesario mantenerlos.

Finalmente, el mapa interactivo es un prototipo analítico y no un sistema institucional desplegado. No incluye todavía administración de usuarios, actualización automática,

monitoreo de versiones, filtros avanzados, descarga controlada, indicadores temporales ni mecanismos de gobernanza. Su utilidad actual consiste en demostrar la viabilidad de la consulta territorial, no en operar como plataforma definitiva.

### 3.3. Recomendaciones

Se recomienda mantener el data mart como producto vivo y versionado. Cada actualización debería registrar la fecha de descarga, el período de la fuente, su responsable institucional, los cambios de estructura y las transformaciones aplicadas. Los archivos originales deben conservarse sin sobreescritura y los productos derivados deben vincularse con manifiestos de ejecución.

Es necesario desarrollar un protocolo periódico de control de calidad. Este protocolo debería verificar llaves territoriales, duplicados, cobertura, valores faltantes, cambios de dominio, coherencia de denominadores, geometrías, variaciones anómalas y diferencias respecto de la versión anterior. La actualización no debería limitarse a reemplazar archivos, sino documentar el efecto de los cambios sobre los indicadores y el ranking.

En futuras versiones conviene fortalecer la dimensión sanitaria. Pueden explorarse indicadores de personal disponible, nivel de atención, camas, equipamiento, horarios, derivaciones, distancias o tiempos de acceso. También sería conveniente revisar la normalización de esta dimensión para reducir el efecto de techo y aumentar su capacidad discriminante sin eliminar diferencias territoriales reales.

La accesibilidad debería ampliarse mediante información multimodal. La densidad vial puede complementarse con tiempos de desplazamiento, estado de las vías, rutas fluviales, transporte disponible, estacionalidad y distancia a centros de servicios. En la Amazonía, la ausencia de carreteras no equivale necesariamente a incomunicación, y la presencia de una vía no garantiza acceso efectivo.

La incorporación de conectividad digital, capacidad fiscal, exposición climática, conflictos socioambientales y calidad efectiva de los servicios puede fortalecer el índice. Sin embargo, las nuevas variables deben ingresar únicamente cuando exista una definición conceptual clara, cobertura suficiente, llave territorial verificable y una interpretación no ambigua.

Una futura escala parroquial debe desarrollarse solo con datos producidos originalmente a ese nivel. No se recomienda distribuir valores cantonales mediante proporciones de población o superficie. Antes de construir un índice parroquial será necesario evaluar la cobertura de todas las dimensiones y la estabilidad de las tasas para unidades con poblaciones pequeñas.

Desde el punto de vista del modelado, conviene comparar transformaciones y normalizaciones alternativas, así como procedimientos robustos frente a extremos. También pueden evaluarse métodos de agrupamiento probabilístico, clustering difuso o enfoques con restricciones espaciales. Estas alternativas deben considerarse como análisis de sensibilidad y no como una búsqueda de mayor complejidad algorítmica.

La validación externa debe constituir la siguiente fase del proyecto. Se recomienda organizar talleres, entrevistas o mesas técnicas con gobiernos locales, jurisdicciones eclesiales, organizaciones amazónicas, especialistas sectoriales y representantes comunitarios. El objetivo no sería solicitar una aprobación general del índice, sino contrastar los perfiles, identificar variables ausentes, reconocer factores protectores y revisar interpretaciones que los datos secundarios no permiten observar.

Para su uso institucional se recomienda adoptar un protocolo de priorización que combine al menos cinco componentes: puntuación global, subíndice crítico, perfil territorial, sensibilidad metodológica y validación cualitativa. Ninguna asignación de recursos debería

fundamentarse únicamente en la posición del ranking. Los casos con sensibilidad alta requieren una revisión adicional antes de ser utilizados para decisiones operativas.

El mapa interactivo puede evolucionar hacia un tablero territorial con filtros por dimensión, provincia, perfil y sensibilidad; comparación entre cantones; ficha metodológica; metadatos; descarga de tablas; consulta de indicadores originales; y visualización de la incertidumbre. También debe mantener las geometrías validadas y evitar simplificaciones que produzcan errores topológicos.

Se recomienda establecer reglas de gobernanza para las variables interculturales y los datos vinculados con pueblos indígenas. La incorporación de información adicional debería considerar beneficios colectivos, autoridad sobre los datos, responsabilidad institucional y prevención de usos estigmatizantes. Los productos públicos deben mantener un nivel de agregación que no exponga información sensible.

Finalmente, una futura investigación puede transformar el índice en un sistema de seguimiento temporal. Para ello se requerirían versiones comparables de los indicadores, reglas estables de normalización y documentación de cambios territoriales. Un enfoque longitudinal permitiría analizar tendencias, pero debería diferenciar los cambios reales de aquellos producidos por actualizaciones de fuentes o modificaciones metodológicas.

# Referencias

Abarca-Alvarez, F. J., Reinoso-Bellido, R., & Campos-Sánchez, F. S. (2019). Decision model for predicting social vulnerability using artificial intelligence. *ISPRS International Journal of Geo-Information, 8*(12), Artículo 575. https://doi.org/10.3390/ijgi8120575

Adger, W. N. (2006). Vulnerability. *Global Environmental Change*, *16*(3), 268–281. https://doi.org/10.1016/j.gloenvcha.2006.02.006

Agencia de Aseguramiento de la Calidad de los Servicios de Salud y Medicina Prepagada. (2026). *Licenciamientos a establecimientos de salud* [Conjunto de datos]. Portal de Datos Abiertos Ecuador. https://www.datosabiertos.gob.ec/dataset/licenciamientos-a-establecimientos-de-salud

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis, 27*(2), 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Banco Mundial. (2021). *Informe sobre el desarrollo mundial 2021: Datos para una vida mejor*. https://www.worldbank.org/en/publication/wdr2021

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics, 3*(1), 1–27. https://doi.org/10.1080/03610927408827101

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodríguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J., & Hudson, M. (2020). The CARE principles for Indigenous data governance. *Data Science Journal, 19*(1), Artículo 43. https://doi.org/10.5334/dsj-2020-043

Comisión Económica para América Latina y el Caribe. (2022). *Desigualdad territorial*. https://igualdad.cepal.org/sites/default/files/2022-02/12.%20CC_Desigualdad%20territorial_es_0.pdf

Cutter, S. L., Boruff, B. J., & Shirley, W. L. (2003). Social vulnerability to environmental hazards. *Social Science Quarterly, 84*(2), 242–261. https://doi.org/10.1111/1540-6237.8402002

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-1*(2), 224–227. https://doi.org/10.1109/TPAMI.1979.4766909

Fernandez, M. A., Bucaram, S. J., & Renteria, W. (2015). Assessing local vulnerability to climate change in Ecuador. *SpringerPlus, 4*, Artículo 738. https://doi.org/10.1186/s40064-015-1536-z

Food and Agriculture Organization of the United Nations. (2020). *Global forest resources assessment 2020: Main report*. https://doi.org/10.4060/ca9825en

Gudynas, E. (2011). Buen vivir: Germinando alternativas al desarrollo. *América Latina en Movimiento,* (462). https://www.gudynas.com/publicaciones/articulos/GudynasBuenVivirGerminandoALAI11.pdf

Harvey, D. (2006). *Spaces of global capitalism: Towards a theory of uneven geographical development*. Verso.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179–185. https://doi.org/10.1007/BF02289447

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification, 2*, 193–218. https://doi.org/10.1007/BF01908075

International Organization for Standardization. (2008). *ISO/IEC 25012:2008 software engineering—Software product quality requirements and evaluation (SQuaRE)—Data quality model*.

Instituto Nacional de Estadística y Censos. (2024). *VIII Censo de Población y VII de Vivienda 2022* [Base de datos y documentación técnica]. Archivo Nacional de Datos y Metadatos Estadísticos. https://anda.inec.gob.ec/anda5/index.php/catalog/1085

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A, 374*(2065), Artículo 20150202. https://doi.org/10.1098/rsta.2015.0202

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika, 23*, 187–200. https://doi.org/10.1007/BF02289233

Kimball, R., & Caserta, J. (2004). *The data warehouse ETL toolkit: Practical techniques for extracting, cleaning, conforming, and delivering data*. Wiley.

Kimball, R., & Ross, M. (2013). *The data warehouse toolkit: The definitive guide to dimensional modeling* (3rd ed.). Wiley.

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. SAGE Publications.

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2015). *Geographic information science and systems* (4th ed.). Wiley.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. En L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281–297). University of California Press.

Mansur, A. V., Brondízio, E. S., Roy, S., Hetrick, S., Vogt, N. D., & Newton, A. (2016). An assessment of urban vulnerability in the Amazon Delta and Estuary: A multi-criterion index of flood exposure, socio-economic conditions and infrastructure. *Sustainability Science, 11*(4), 625–643. https://doi.org/10.1007/s11625-016-0355-7

MapBiomas Ecuador. (2024). *Colección 3.0 de mapas anuales de cobertura y uso del suelo en la Amazonía entre 1985 y 2024* [Base de datos geoespacial]. https://ecuador.mapbiomas.org/#

Massey, D. (1994). *Space, place and gender*. Polity Press.

Ministerio de Educación. (2025). *Registro de Matrícula - MINEDUC* [Conjunto de datos]. Portal de Datos Abiertos Ecuador. https://www.datosabiertos.gob.ec/dataset/registro-de-matricula-mineduc

Ministerio de Salud Pública. (2026). *GeoSalud 3.5.1* [Geovisualizador]. https://geosalud.msp.gob.ec/geovisualizador/

OpenAI. (2026). *ChatGPT* [Modelo de lenguaje de gran escala]. https://chatgpt.com/

Organisation for Economic Co-operation and Development, & European Commission, Joint Research Centre. (2008). *Handbook on constructing composite indicators: Methodology and user guide*. OECD Publishing. https://doi.org/10.1787/9789264043466-en

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830. http://jmlr.org/papers/v12/pedregosa11a.html

Peng, R. D. (2011). Reproducible research in computational science. *Science, 334*(6060), 1226–1227. https://doi.org/10.1126/science.1213847

Programa de las Naciones Unidas para el Desarrollo. (2020). *Informe sobre desarrollo humano 2020: La próxima frontera: El desarrollo humano y el Antropoceno*. https://hdr.undp.org/content/human-development-report-2020

Provost, F., & Fawcett, T. (2013). *Data science for business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media.

Red Amazónica de Información Socioambiental Georreferenciada. (2020). *Amazonía bajo presión 2020*. https://www.raisg.org/es/publicacion/amazonia-bajo-presion-2020/

Red Amazónica de Información Socioambiental Georreferenciada. (s.f.). *Mapas y datos cartográficos sobre Amazonía* [Base de datos geoespacial]. Recuperado el 29 de junio de 2026, de https://www.raisg.org/es/mapas/

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Rule, A., Birmingham, A., Zuniga, C., Altintas, I., Huang, S.-C., Knight, R., Moshiri, N., Nguyen, M. H., Rosenthal, S. B., Pérez, F., & Rose, P. W. (2019). Ten simple rules for reproducible research in Jupyter notebooks. *PLOS Computational Biology, 15*(7), e1007007. https://doi.org/10.1371/journal.pcbi.1007007

Saltelli, A. (2007). Composite indicators between analysis and advocacy. *Social Indicators Research, 81*, 65–77. https://doi.org/10.1007/s11205-006-0024-9

United Nations Office for Disaster Risk Reduction. (2022). *Global assessment report on disaster risk reduction 2022: Our world at risk: Transforming governance for a resilient future*. United Nations. https://www.undrr.org/gar2022

Walsh, C. (2009). *Interculturalidad, Estado, sociedad: Luchas (de)coloniales de nuestra época*. Universidad Andina Simón Bolívar; Abya-Yala.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data, 3*, Artículo 160018. https://doi.org/10.1038/sdata.2016.18

Yeo, I.-K., & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika, 87*(4), 954–959. https://doi.org/10.1093/biomet/87.4.954

# Anexos

Los materiales complementarios del Proyecto de Titulación se encuentran disponibles en un repositorio digital organizado de acuerdo con la arquitectura del data mart territorial desarrollado durante la investigación. Los anexos contienen documentación, matrices, bases derivadas, resultados analíticos, productos geoespaciales y cuadernos de código necesarios para examinar la trazabilidad y reproducibilidad del estudio. Los archivos originales y derivados conservan la estructura y el versionamiento utilizados durante el procesamiento.

**Repositorio digital del proyecto:** data_mart_territorial

## Anexo A. Inventario y arquitectura del data mart territorial

**Descripción:**

Este anexo documenta la organización general del data mart territorial construido para el Proyecto de Titulación. Presenta la estructura de carpetas, la función de cada componente y la relación entre los datos originales, productos procesados, base integrada, indicadores, resultados de modelado, visualizaciones y código. Su finalidad es facilitar la navegación del repositorio y reconstruir el flujo completo desde las fuentes originales hasta los productos analíticos finales.

**Enlace al repositorio:** data_mart_territorial

**Contenido principal:**

| Componente | Contenido |
|---|---|
| **00_documentacion** | README general, bitácora de procesamiento, diccionario maestro y matriz de archivos |
| **01_datos_crudos** | Fuentes originales conservadas sin sobrescritura |
| **02_datos_procesados** | Productos depurados, homologados y agregados |

| | |
|---|---|
| **03_base_integrada** | Base maestra cantonal y productos complementarios |
| **04_indicadores** | Catálogo, evaluación, matriz principal y sensibilidad |
| **05_modelado** | PCA, clustering, índice, validación e interpretabilidad |
| **06_visualizaciones** | Gráficos, cartografía y productos finales |
| **07_codigo** | Cuadernos ejecutados y código de reproducción |

## Anexo B. Matriz de fuentes y archivos definitivos

**Descripción:**

Este anexo contiene la matriz detallada de fuentes y archivos utilizados durante la construcción del data mart. Documenta la procedencia institucional, período de referencia, nivel territorial, formato, llave de integración, cobertura, controles aplicados, ubicación dentro del repositorio y decisión metodológica adoptada para cada producto. La matriz permite identificar cuáles archivos se utilizaron en el modelo principal, cuáles se conservaron para análisis de sensibilidad o contextualización y cuáles fueron excluidos.

**Enlace al repositorio:** matriz_archivos_definitivos_v1.xlsx

**Contenido principal:** matriz de los 35 productos procesados definitivos mencionados en la metodología, junto con su documentación de procedencia y uso.

## Anexo C. Diccionario maestro de variables

**Descripción:**

Este anexo contiene el diccionario maestro retrospectivo elaborado a partir de las fuentes integradas. Documenta las variables canónicas del data mart, incluyendo su nombre, definición, fuente, período, nivel territorial, unidad de medida, tipo de variable, tratamiento aplicado, función analítica, restricciones y nivel de confianza documental. Su finalidad es

preservar la trazabilidad semántica de los datos y facilitar futuras actualizaciones o reutilizaciones.

**Enlace al repositorio:** Diccionario_maestro_variables_v1.xlsx

**Contenido principal:** diccionario maestro de 924 variables canónicas, así como la documentación específica asociada con la base maestra cantonal.

### Anexo D. Catálogo y evaluación de indicadores

**Descripción:**

Este anexo presenta el catálogo completo de indicadores construidos a partir de la base maestra cantonal y los resultados de su evaluación conceptual, documental y estadística. Incluye fórmulas, orientación respecto de la vulnerabilidad, cobertura, valores faltantes, variabilidad, presencia de valores extremos, correlaciones y decisión final de inclusión.

**Enlace al repositorio:** catalogo_indicadores_candidatos_v1.csv

**Contenido principal:**

- 34 indicadores candidatos;
- 13 indicadores seleccionados para el modelo principal;
- 14 indicadores reservados para análisis de sensibilidad;
- variables contextuales o de estratificación;
- indicadores pendientes o excluidos;
- matriz definitiva de modelado.

### Anexo E. Resultados detallados del análisis de componentes principales

**Descripción:**

Este anexo contiene los resultados completos del preprocesamiento y del análisis de componentes principales utilizado como procedimiento exploratorio y como referencia para los análisis de sensibilidad. Incluye las transformaciones aplicadas a los indicadores, matrices

estandarizadas, criterios de retención, varianza explicada, cargas, comunalidades y resultados de las variantes PCA evaluadas.

**Enlace al repositorio:** 01_preprocesamiento_pca

**Visualizaciones complementarias:** 01_preprocesamiento_pca

**Contenido principal:** PCA de 13 indicadores, variante de 11 indicadores, análisis paralelo, criterio de Kaiser, gráficos de sedimentación, cargas factoriales, comunalidades, rotación Varimax y coordenadas de los cantones.

### Anexo F. Resultados detallados del clustering

**Descripción:**

Este anexo reúne los resultados del aprendizaje automático no supervisado utilizado para identificar perfiles territoriales. Documenta las representaciones de datos evaluadas, algoritmos y números de agrupamientos comparados, métricas de calidad interna, pruebas de estabilidad y asignación definitiva de los cantones a los tres perfiles territoriales.

**Enlace al repositorio:** 02_clustering

**Visualizaciones complementarias:** 02_clustering

**Contenido principal:** resultados de K-means, clustering jerárquico con enlace Ward y diagnóstico exploratorio mediante DBSCAN; silhouette score, Davies-Bouldin, Calinski-Harabasz, estabilidad mediante submuestreo e Índice de Rand ajustado; asignación cantonal y caracterización dimensional de los perfiles.

### Anexo G. Interpretabilidad y sensibilidad por cantón

**Descripción:**

Este anexo presenta los resultados detallados utilizados para explicar la composición del índice y evaluar la robustez de las puntuaciones y clasificaciones. Permite reconstruir la contribución de cada indicador y dimensión para cada cantón y examinar la variación de los resultados frente a cambios metodológicos plausibles.

**Enlace al repositorio:** 04_validacion_interpretabilidad

**Contenido principal:** contribuciones por indicador y dimensión, dimensión dominante, comparación entre ponderación equilibrada y ponderación PCA, exclusión sucesiva de dimensiones, simulaciones de perturbación de pesos, rangos de posición, probabilidades de categoría y clasificación de sensibilidad metodológica.

### Anexo H. Productos cartográficos y visualizaciones complementarias

**Descripción:**

Este anexo contiene los productos gráficos y geoespaciales generados para representar e interpretar los resultados del índice. Incluye las versiones de alta resolución de los mapas utilizados en el documento, cartografía complementaria y el mapa interactivo desarrollado como prototipo de consulta territorial.

**Enlace al repositorio:** 04_validacion_interpretabilidad

**Enlace a mapa interactivo del índice:**

mapa_interactivo_indice_vulnerabilidad_socioambiental_2026.html

**Contenido principal:** mapas del índice global, seis subíndices, perfiles territoriales, sensibilidad metodológica y asociaciones espaciales locales; gráficos complementarios; archivo HTML del mapa interactivo; y capas territoriales enriquecidas disponibles en formatos geoespaciales.

### Anexo I. Controles de calidad y auditoría geoespacial

**Descripción:**

Este anexo reúne la evidencia principal de los procedimientos de control aplicados durante la limpieza, homologación, integración y validación geoespacial de los datos. Los controles específicos de cada fuente se conservan en los respectivos módulos temáticos de `02_datos_procesados`, mientras que el anexo identifica los controles transversales

utilizados para validar la construcción de la base maestra cantonal y los productos territoriales finales.

**Enlace al repositorio**: data_mart_territorial

**Contenido principal:**

| Componente | Ubicación en el data mart | Función |
|---|---|---|
| **Controles temáticos de fuentes** | `02_datos_procesados/[módulo_temático]/controles/` | Verificación de estructura, cobertura, llaves, duplicados, faltantes y consistencia durante el procesamiento de cada fuente |
| **Control de integración cantonal** | `03_base_integrada/` — `control_integracion_cantonal_v1.xlsx` | Verificación de la integración de fuentes en la base maestra |
| **Manifiesto de integración** | `03_base_integrada/` — `manifiesto_integracion_cantonal_v1.json` | Registro de la ejecución, estructura y productos generados durante la integración |
| **Base territorial validada** | `03_base_integrada/` — `base_maestra_cantonal_v1.gpkg` | Geometría y atributos territoriales utilizados como referencia del análisis |
| **Auditoría geoespacial final** | `05_modelado/04_validacion_interpretabilidad/` | Validación de geometrías, cobertura, correspondencia territorial y controles topológicos |
| **Productos de control cartográfico** | `06_visualizaciones/04_validacion_interpretabilidad/` | Evidencia derivada de la validación de los productos espaciales finales |

Los controles temáticos no se reproducen individualmente en el presente anexo debido a su número y a que forman parte de la trazabilidad interna de cada fuente. Se conservan en las carpetas originales del data mart, de manera que cada producto procesado pueda examinarse conjuntamente con la evidencia de control que sustentó su aprobación.

## Anexo J. Código y reproducibilidad

**Descripción:**

Este anexo contiene los cuadernos Jupyter/Google Colab utilizados para ejecutar las diferentes etapas del procesamiento y modelado. Los notebooks conservan el código y las salidas de ejecución necesarias para revisar y reproducir la construcción de la base integrada, los indicadores, el análisis de componentes principales, el clustering, el índice y las pruebas de validación e interpretabilidad.

**Enlace al repositorio:** 07_codigo

**Contenido principal:**

| Componente | Función |
|---|---|
| **`00_reproducibilidad`** | Documentación técnica para reproducir el flujo analítico. Contiene `README_reproducibilidad.md`, `requirements.txt` e `inventario_entradas_salidas_reproducibilidad.csv`. |
| **`01_etl`** | Limpieza, homologación, transformación e integración de las fuentes; construcción y controles de la base maestra cantonal. |
| **`02_indicadores`** | Construcción, evaluación y selección de indicadores; generación de la matriz del modelo principal y de los indicadores de sensibilidad. |
| **`03_preprocesamiento_pca`** | Orientación y transformación de variables, estandarización, normalización, diagnósticos y análisis de componentes principales. |
| **`04_clustering`** | Comparación de algoritmos y configuraciones, evaluación de métricas, análisis de estabilidad y construcción de los perfiles territoriales. |
| **`05_indice_subindices`** | Construcción de los seis subíndices, cálculo del índice global, clasificación relativa y comparación de estrategias de ponderación. |
| **`06_validacion_interpretabilidad`** | Validación aritmética y territorial, descomposición de contribuciones, análisis de sensibilidad, análisis espacial y generación de productos finales de interpretación. |

## Anexo K. Declaración y bitácora de uso de inteligencia artificial generativa

**Descripción:**

Este anexo documenta el uso supervisado de inteligencia artificial generativa durante el Proyecto de Titulación. Registra la herramienta empleada, las finalidades de uso, los límites de su participación y los mecanismos de verificación aplicados por la autora. Incluye

una muestra representativa de las interacciones utilizadas para apoyar la estructuración y depuración de código, documentación de cuadernos y revisión técnica y editorial.

**Enlace al repositorio:** 00_uso_ia_generativa

**Contenido principal:** declaración metodológica de uso de ChatGPT; ejemplo representativo de interacción; delimitación de tareas realizadas por la herramienta y decisiones reservadas a la autora; procedimientos de ejecución y verificación humana.