

ARTICLE

Dependencia espacial en la evaluación de la susceptibilidad por movimientos en masa

Edier Aristizábal

Departamento de Geociencias y Medio Ambiente, Universidad Nacional de Colombia, Medellín, Colombia

Abstract

Landslides are critical geomorphological processes that substantially reshape the landscape through the downslope movement of soil and rock, often triggered by factors such as rainfall, earthquake, or anthropogenic interventions. These processes pose significant hazards to infrastructure, human safety, and socioeconomic stability. Conventional statistical models frequently fail to adequately capture the spatial nature of landslide susceptibility assessment, often leading to biased or misleading outcomes due to the assumption of independence among observations, which ignores inherent spatial heterogeneity and spatial dependence. This study addresses these limitations by employing Spatial Autoregressive Models (SAR), which explicitly account for spatial dependence through the integration of neighborhood matrices. The dataset comprises catchments from the Colombian Andes, incorporating morphometric predictors at both local and regional scales, including slope, hypsometry, basin area, and annual precipitation. We built a neighborhood matrix based on distance criteria, recognizing that geoenvironmental factors influencing landslides often extend beyond direct boundaries, requiring a broader understanding of spatial interactions. Our findings demonstrate that incorporating spatial dependence significantly enhances both the predictive accuracy and interpretative power of the models when compared to conventional approaches. Analysis using Moran's Index revealed that basin slope and precipitation exhibit strong spatial dependence, forming clusters of similar values, which underscores the necessity of accounting for spatial effects. The Spatial Durbin Error Model (SDEM) outperformed alternative models by providing higher adjusted R^2 values and optimizing the balance between model complexity and fit, as measured by the Akaike Information Criterion (AIC). By explicitly integrating spatial structure, this study provides a more robust and realistic assessment of landslide susceptibility, which is crucial for understanding and managing these hazards in regions like the Colombian Andes.

Keywords: landslide; susceptibility; spatial dependence

1. Introducción

Los movimientos en masa son procesos geomorfológicos transformadores del paisaje que corresponden al desprendimiento y desplazamiento de materiales, como suelo y roca, ladera abajo por efecto de la gravedad (Soeters and Van Westen 1996a; Brabb 1984). Estos eventos pueden ser desencadenados por factores como precipitación, sismos o intervención antrópica (Corominas et al. 2014; Fell et al. 2008). Dado lo anterior, la susceptibilidad por movimientos en masa se define como la probabilidad de ocurrencia de estos procesos en un área específica, lo cual es fundamental para identificar zonas potencialmente críticas y para tomar decisiones informadas en la gestión del territorio, la planificación urbana y la mitigación de desastres (Brabb 1984; Corominas et al. 2014).

La evaluación de la susceptibilidad por movimientos en masa es, en esencia, un problema espacial que involucra un amplio número de variables geoambientales (Reichenbach et al. 2018). Los movimientos en masa están directamente influenciados por el entorno donde ocurren, lo cual implica una estrecha relación con las características del terreno, así como con las áreas circundantes (Soeters and Van

Westen 1996b). Es por esta razón, que los inventarios de movimientos en masa, así como de variables morfométricas del terreno, condiciones geológicas y ambientales, permiten inferir las condiciones bajo las cuales se desencadenaron los eventos en el pasado, para predecir la ocurrencia de futuros movimientos en masa, bajo el supuesto que procesos similares podrían ocurrir en el futuro bajo condiciones semejantes (Dai *et al.* 2001; Soeters and Van Westen 1996a).

La identificación, caracterización y mapeo de estas variables geoambientales, que condicionan la ocurrencia de movimientos en masa, permiten incorporar en los modelos de susceptibilidad los patrones espaciales en la distribución de la ocurrencia de movimientos en masa (Reichenbach *et al.* 2018); lo que corresponde a un mapa de susceptibilidad. Una de las técnicas para establecer dichos patrones espaciales es a través de los métodos estadísticos, también denominados métodos basados en datos o recientemente aprendizaje de máquinas (*machine learning*) (Korup and Stolle 2014). Estos métodos se caracterizan por su capacidad de procesar grandes cantidades de datos y encontrar, mediante técnicas estadísticas, patrones que relacionan la variable respuesta (la ocurrencia o no de movimientos en masa) y las variables geoambientales incorporadas en el análisis como predictores.

Sin embargo, además de las variables geoambientales observables e incorporadas en los modelos de susceptibilidad, existen una serie de variables latentes, es decir que no pueden ser observadas o medidas directamente, pero que influye en el comportamiento de otras variables del modelo (Lombardo, Bakka, *et al.* 2019). Algunas de ellas asociadas a las características espaciales de las variables. Este componente latente en algunas ocasiones está incorporado dentro de las variables predictoras, pero no es considerado o permitido en los modelos estadísticos clásicos (Anselin and Griffith 1988; Cressie 1988; J. P. LeSage and Pace 2011). En otros casos, estos factores simplemente se omiten.

Para atender esta problemática es necesario inicialmente comprender la estructura espacial de los datos. La estructura espacial se puede descomponer en dos componentes principales: la heterogeneidad espacial y la dependencia espacial (Anselin, Bera, *et al.* 1996; S. Rey, Arribas-Bel, and Wolf 2023). La heterogeneidad espacial se refiere a la variación en las propiedades de los datos en diferentes áreas del espacio, lo que implica que las relaciones entre las variables no necesariamente sean homogéneas en toda la región de estudio (S. Rey, Arribas-Bel, and Wolf 2023). Por ejemplo, en una región montañosa como los Andes, las características geológicas y climáticas pueden variar significativamente de un lugar a otro, lo que puede generar que una variable tenga una influencia variable a lo largo de la zona de estudio en la ocurrencia de movimientos en masa. La dependencia espacial, por otro lado, se refiere a la tendencia de valores similares a agruparse espacialmente (Anselin and Griffith 1988). Esto implica que el valor de una variable en un punto del espacio está correlacionado con los valores de la misma variable en puntos cercanos, un fenómeno también conocido como autocorrelación. Esta dependencia puede originarse debido a procesos naturales que presentan continuidad espacial, como la geología y la topografía, o debido a factores ambientales como la vegetación y el clima (Samia *et al.* 2019; Lombardo, Opitz, *et al.* 2020), que tienden a presentar patrones espaciales similares en áreas cercanas.

Cuando la heterogeneidad y la dependencia espacial no se incorporan adecuadamente en los modelos, la estructura espacial queda remanente en los residuales, lo cual se traduce en una falta de ajuste adecuada del modelo (Anselin 2022; Cressie 2015). Este remanente espacial en los residuos es una violación de las suposiciones de independencia que subyacen a los modelos clásicos, ya que los modelos clásicos suelen asumir que las observaciones son independientes entre sí (sin dependencia espacial) y que la relación entre las variables es constante en toda la región de estudio (sin heterogeneidad espacial) (J. P. LeSage and Pace 2011; Banerjee *et al.* 2008). Esto afecta la validez de las inferencias estadísticas y resulta en modelos con predicciones sesgadas (Cressie 1988).

Para incorporar la dependencia espacial en los modelos, se puede utilizar los llamados Modelos Autoregresivos Simultáneos (Whittle 1954), los cuales permiten modelar la dependencia espacial

a través de la inclusión de una matriz de vecindad. Esta matriz define la estructura de los vecinos de cada unidad de mapeo, ya sea en función de la distancia entre unidades o a partir de relaciones de contigüidad (Anselin, Bera, et al. 1996; J. P. Elhorst et al. 2014).

Los modelos SAR son ampliamente utilizados en econometría y geografía, así como en otros campos que requieren un análisis espacial riguroso, como la epidemiología y la planificación urbana (J. P. Elhorst 2022; S. Rey, Arribas-Bel, and Wolf 2023; Jaya and Chadidjah 2021; Ver Hoef et al. 2018). Sin embargo, la aplicación de algunos de estos modelos requiere de técnicas especiales, ya que la estructura de dependencia espacial puede introducir complicaciones que no están presentes en los modelos estadísticos clásicos (Anselin 1990).

En este artículo presentamos el concepto de dependencia espacial, su influencia en la evaluación de la susceptibilidad por movimientos en masa y modelos estadísticos que nos permiten incorporar este componente espacial en la modelación. Para ello, se utiliza un conjunto de datos de movimientos en masa y cuencas ubicadas en los Andes colombianos. A través de este estudio, se evalúa la estructura espacial presente en los datos y se aplican modelos autoregresivos que permiten incorporar esta estructura de manera explícita. La incorporación de la dependencia espacial en los modelos de susceptibilidad no solo nos permite mejorar la capacidad predictiva de los modelos, sino que también ofrece una interpretación más realista de los factores que influyen en la ocurrencia de movimientos en masa.

2. Modelos de Autoregresión Espacial

Los modelos de autoregresión espacial son un tipo de modelo estadístico utilizado para capturar y modelar la dependencia espacial entre unidades de análisis geográficas (Whittle 1954; Getis 2009; Stakhovych and Bijmolt 2009; Ripley 1988). En lugar de asumir que las observaciones son independientes entre sí, estos modelos reconocen que las unidades cercanas en el espacio tienden a presentar valores similares (Anselin and Griffith 1988).

La aplicabilidad de los modelos de autoregresión espacial en los estudios de susceptibilidad a movimientos en masa radica en la naturaleza inherentemente espacial de estos fenómenos. Los movimientos en masa tienden a estar influenciados por las condiciones de las áreas adyacentes, como el relieve del terreno, condiciones climáticas, la cobertura del suelo y las características geológicas (Samia et al. 2019). Estos modelos son ampliamente utilizados en análisis espaciales debido a su capacidad para incluir explícitamente la estructura de dependencia entre las unidades vecinas a través de matrices de vecindad (J. P. LeSage and Pace 2011; Kelejian and Prucha 2007). Esta matriz introduce el componente espacial, permitiendo que el valor de la variable respuesta en cada unidad de análisis se vea influido por los valores de las unidades vecinas. Esto permite modelar explícitamente el efecto que tienen los vecinos en la variable de interés.

Sin embargo, los modelos de autoregresión espacial también presentan algunas limitaciones (J. P. LeSage and Pace 2011; Anselin 2022). La estimación de los parámetros en los modelos de autoregresión espacial puede ser compleja, especialmente cuando se trabaja con grandes conjuntos de datos y matrices de vecindad densas. Adicionalmente, la definición de la matriz de vecindad es fundamental, y una elección incorrecta puede llevar a resultados sesgados o poco representativos.

Las matrices de vecindad son una pieza clave en los modelos de autoregresión espacial (Earnest et al. 2007). Una matriz de vecindad define qué unidades se consideran "vecinas" y cómo se relacionan entre sí. Cada elemento de la matriz representa la fuerza o presencia de una relación espacial entre dos unidades. Existen diferentes tipos de matrices de vecindad (Earnest et al. 2007; Getis 2009; Stakhovych and Bijmolt 2009):

- 1. Matriz de Vecindad Basada en la Contigüidad:** Este tipo de matriz define como vecinos a

aquellas unidades que comparten un límite común. Es ideal para el análisis en zonas divididas en celdas regulares o regiones administrativas. 114
115

2. Matriz de Vecindad Basada en Distancia: En este tipo de matriz, se consideran vecinos aquellas unidades que se encuentran dentro de un radio específico de distancia. La vecindad se define en términos de la cercanía física. 116
117
118

3. Matriz de Vecindad K-Veinos más Cercanos: Esta matriz define como vecinos a un número fijo de las unidades más cercanas a cada unidad de análisis. Asegura que cada unidad tenga el mismo número de vecinos, lo cual puede ser útil para evitar problemas de unidades aisladas sin vecinos. 119
120
121

La ecuación básica de un modelo de autoregresión simultaneo se puede representar de la siguiente manera (Cliff and Ord 1981; Kelejian and Prucha 2007; J. LeSage and Pace 2009): 122
123

$$Y = \rho WY + X\beta + \epsilon \quad (1)$$

donde el término ρ introduce el componente espacial, permitiendo que el valor de la variable respuesta Y en cada unidad de análisis esté influido por los valores de las unidades vecinas, ponderados por la matriz de vecindad (W). β es el vector de coeficientes de regresión asociados a las variables predictoras (X). Estos coeficientes reflejan la influencia de cada variable en la variable respuesta Y . (ϵ) es el término de error aleatorio distribuido de manera independiente e idéntica con media cero, que recoge la variabilidad no explicada por el modelo. 124
125
126
127
128
129

En los modelos de autoregresión simultaneo (SAR) de forma estricta la dependencia espacial está asociada a la variable respuesta Y , lo que implica una relación simultánea que involucra todas las observaciones. Esto requiere una operación como la inversa de $(1 - \rho W)$ para despejar los valores de Y . Sin embargo, la mayoría de los autores amplían este grupo a modelos espaciales con tres tipos de efectos de dependencia: (i) autocorrelación endógena en la variable respuesta (SAR), (ii) autocorrelación exógena en los predictores y (iii) autocorrelación en los términos de error. El modelo espacial general (GNS, por sus siglas en inglés), también conocido como modelo espacial SARAR(1,1) o modelo espacial de Cliff-Ord, abarca estos tres tipos de dependencias de la siguiente forma (J. LeSage and Pace 2009; Kelejian and Prucha 2007; Cliff and Ord 1981): 130
131
132
133
134
135
136
137
138

$$y = X\beta + WXy + \rho Wy + \lambda W\mu + \epsilon \quad (2)$$

donde λ es el coeficiente en una estructura autoregresiva espacial para el error μ . Es decir, que el término Wy captura los efectos de interacción endógena entre la variable respuesta, WXy modela los efectos de interacción exógena entre los predictores; mientras que el término $W\mu$ se ocupa de los efectos de interacción entre los residuales. 139
140
141
142

De acuerdo como se incluyan o eliminan los términos se pueden obtener diferentes especificaciones (P. Elhorst and Vega 2013). La más común es el modelo con Autoregresión Espacial (SAR), donde $\lambda = 0$, y por lo tanto solo incluye la variable dependiente con autoregresión espacial, ρWy . En el caso donde $\rho = 0$, el modelo de Error Espacial (SEM) retiene los términos de error $\lambda W\mu$, mientras que el modelo con Autoregresión Espacial en X (SLX) comprende la autoregresión espacial de los predictores, WXy . Otras variantes corresponden al modelo Combinado Autoregresivo de Kelejian-Prucha (SAC), que incluye autocorrelación en ρWy y $\lambda W\mu$; el modelo de Durbin Espacial (SDM), que combina una variable dependiente autoregresiva ρWy y predictores con autoregresión espacial WXy ; y el modelo de Error de Durbin Espacial (SDEM), que incluye un término de error espacial $\lambda W\mu$ combinado con predictores con autoregresión espacial WXy (J. P. Elhorst 2022; Anselin 2022). 143
144
145
146
147
148
149
150
151
152

Un primer paso esencial en el uso de modelos espaciales es estimar cualquier posible estructura espacial en los datos de entrada o en el residuo de los modelos. Dos pruebas comunes para detectar efectos espaciales son el gráfico de dispersión de Moran (Anselin and Griffith 1988) y la prueba del Multiplicador de Lagrange (LM) (Anselin, Bera, et al. 1996). El gráfico de dispersión de Moran ilustra la relación entre cada observación y el promedio de sus vecinos, y ayuda a identificar conglomerados locales y valores atípicos espaciales. El gráfico categoriza las asociaciones espaciales en cuatro tipos: valores altos rodeados de vecinos con valores altos (Q1), valores bajos rodeados de valores bajos (Q3), Bajo-Alto (Q2) y Alto-Bajo (Q3), que indican conglomerados o valores atípicos relativos a la media general. El estadístico I de Moran es la pendiente de un ajuste lineal al gráfico de dispersión, y mide la dependencia espacial general:

$$MI = n \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n \sum_{j=1}^n w_{ij} \sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

donde n es el número de observaciones, y_i es el valor estandarizado de la variable respuesta en la ubicación i . El Índice de Moran (MI) varía de -1 a $+1$, donde cero indica ausencia de autocorrelación espacial, $+1$ indica una agrupación perfecta de valores similares, y -1 indica dispersión.

La prueba del Multiplicador de Lagrange (LM) se utiliza para decidir sobre un modelo de regresión espacial adecuado y evaluar la autocorrelación espacial tanto en la variable respuesta como en sus residuales (Anselin and Griffith 1988). La prueba LM compara modelos restringidos (mínimos cuadrados ordinarios sin términos espaciales) con modelos no restringidos (que incluyen términos espaciales). El término espacial puede ser una variable dependiente autoregresiva espacialmente o un término de error autoregresivo espacialmente. El estadístico LM se calcula como la diferencia entre los logaritmos de verosimilitud de los modelos restringido y no restringido, siguiendo una distribución χ^2 con un grado de libertad bajo la hipótesis nula de no dependencia espacial. Si el estadístico de la prueba excede el valor crítico, se rechaza la hipótesis nula de independencia espacial.

3. Área de estudio y datos

El área de estudio se encuentra al norte de 5°N y abarca las Cordilleras Occidental y Central de los Andes colombianos, separadas por el cañón del Cauca. La región está delimitada por el río Atrato al oeste y el río Magdalena al este. Cubriendo aproximadamente $50,000 \text{ km}^2$, el área se dividió en 533 subcuencas dentro de las cuencas del Atrato (25% de las subcuencas), Cauca (50%) y Magdalena (25%). Aproximadamente el 73% de estas subcuencas drenan áreas menores a 100 km^2 , con un área mediana de 48 km^2 .

Para el estudio se utilizó un catálogo de 13,777 movimientos en masa que ocurrieron entre 1970 y 2023, los cuales fueron detectados manualmente usando imágenes ópticas en color verdadero de alta resolución ($<1 \text{ m}$) de Google Earth™. Como variable dependiente se utilizó la frecuencia de movimientos en masa, y como variables predictoras la elevación media (E) de la cuenca, el relieve local medio (H), la precipitación anual media (P) y el área de la cuenca (A). Todos los parámetros del terreno se calcularon utilizando el modelo digital de elevación (DEM) del Radar de Apertura Sintética en Banda L del Satélite de Observación Avanzada de la Tierra - Arreglo en Fase (ALOS-PALSAR), con una resolución de píxel de 12.5m Logan and Karr 2014. los datos de precipitación se obtuvieron del Grupo CHIRPS (por sus siglas en inglés Climate Hazard Infrared Precipitation Sensors) Funk, Peterson, and Landsfeld 2015, versión 2.0, para datos de precipitación de resolución de 5 km desde 1981 hasta 2023.

Para evaluar la dependencia espacial usamos la librería PySAL (S. J. Rey and Anselin 2007) en Python y los modelos fueron implementados en *spatialreg* (Roger Bivand 2022) en lenguaje R. Para evaluar

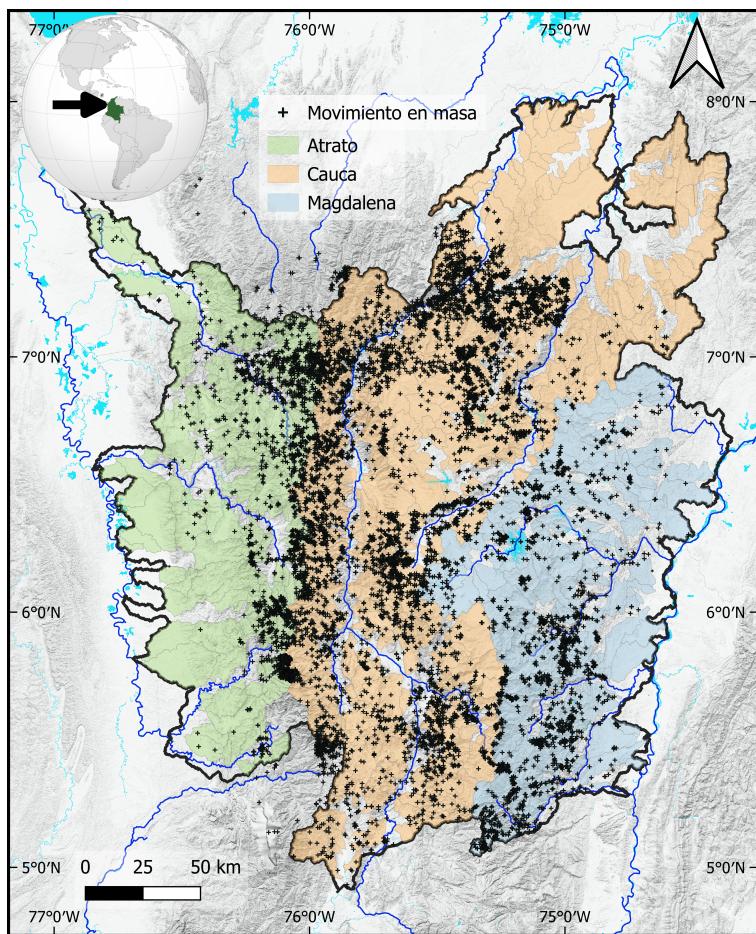


Figure 1. Localización del área de estudio en la región andina de Colombia, mostrando la división en las cuencas del Atrato (verde), Cauca (naranja) y Magdalena (azul). Los deslizamientos registrados entre 1970 y 2023 se representan mediante cruces negras.

la bondad de ajuste y comparar diferentes modelos, utilizamos el Criterio de Información de Akaike (AIC), donde el modelo con el AIC más bajo logra el mejor equilibrio entre ajuste y complejidad.

4. Resultados

La Tabla 1 presentada el resumen de los resultados obtenidos para el modelo clásico sin considerar la dependencia espacial, proporcionando una referencia inicial para evaluar la efectividad de los modelos posteriores que incorporan la dependencia espacial. Las variables Área, Hipso y Pendiente presentan coeficientes positivos y significativos, lo que indica una relación positiva con la frecuencia de movimientos en masa. Por otro lado, la variable Lluvia muestra un coeficiente negativo significativo, lo cual sugiere que un aumento en la precipitación promedio anual está correlacionado con una menor ocurrencia de deslizamientos. Adicionalmente, la tabla incluye tres pruebas clave para evaluar la estructura espacial de los residuales: el índice de Moran, y las pruebas del Multiplicador de Lagrange para los errores (Error) y la autoregresión espacial de la variable respuesta (Lag). El valor del índice de Moran para los residuales es 0.39, lo cual indica una autocorrelación espacial positiva en los residuos del modelo, sugiriendo que existe un patrón espacial que el modelo base no ha logrado

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209
210
211
212

capturar adecuadamente. Los valores elevados en las pruebas LM para Error (316.04) y Lag (276.76) respaldan esta observación, indicando la presencia de una dependencia espacial significativa que debe ser modelada, con un valor mayor para el término del error, lo que sugiere la implementación de un modelo SEM.

Table 1. Resumen del modelo base sin dependencia espacial. Se muestran los coeficientes estimados, errores estándar, valores t y valores p para los parámetros del modelo. Además, se incluyen los resultados del índice de autocorrelación espacial de Moran para los residuales, así como las pruebas LM del Multiplicador de Lagrange para los errores y la autoregresión espacial.

Parámetro	Coeficiente	Std. Error	t-value	p-value
(Intercepto)	1.84	0.05	36.72	0.00
Área	0.62	0.05	12.33	0.00
Hipso	0.14	0.06	2.47	0.01
Pendiente	0.69	0.05	12.50	0.00
Lluvia	-0.29	0.05	-5.34	0.00
I. Moran (Residuales)	0.39			
LM Test (Error)	316.04			
LM Test (Lag)	276.76			

La Figura 2 presenta los resultados del modelo base. En el panel izquierdo se presentan diagramas de caja de los residuales agrupados por cuenca hidrográfica (Atrato, Cauca y Magdalena). Aunque los valores medianos se encuentran relativamente próximos a cero, se puede observar la presencia de diferencias en la variabilidad de los residuales entre las tres cuencas, lo cual sugiere que las características espaciales de cada cuenca podrían estar influyendo en el comportamiento de los residuales. En el panel derecho se presenta la distribución espacial de los valores estimados por el modelo para cada subcuenca. Estos resultados indican que el modelo base, al no incorporar la dependencia espacial, podría estar omitiendo componentes importantes que explican la variación en la susceptibilidad a movimientos en masa.

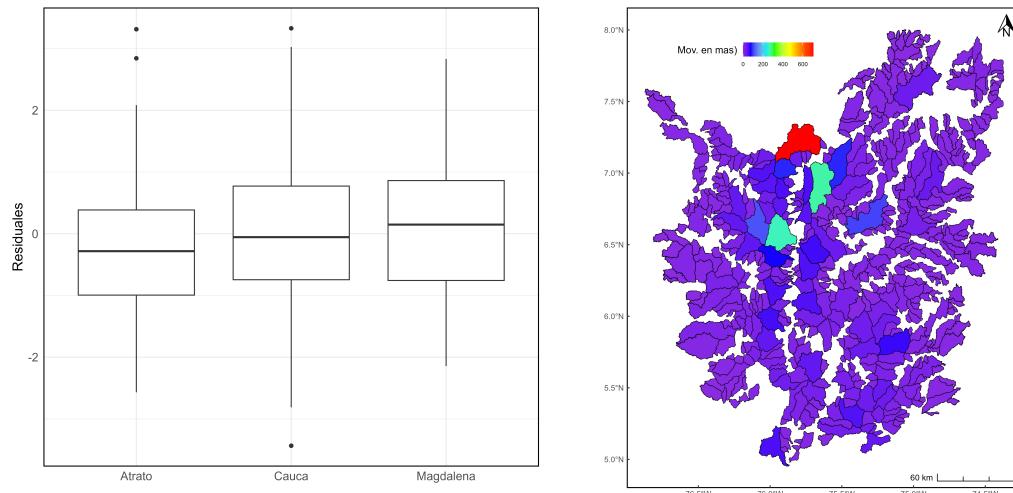


Figure 2. Resultados del modelo base sin dependencia espacial. El panel izquierdo presenta diagramas de caja de los residuales agrupados por cuenca hidrográfica (Atrato, Cauca y Magdalena), indicando diferencias en la variabilidad entre cuencas. El panel derecho muestra la distribución espacial de los valores estimados por el modelo en cada subcuenca, utilizando un esquema de colores donde los tonos rojos indican valores estimados más altos de deslizamientos y los tonos azules y verdes representan valores más bajos.

El primer paso para incluir la dependencia espacial es establecer la relación entre las observaciones mediante la matriz de vecindad. La Figura 3 presenta un ejemplo de la matriz de vecindad generada para el área de estudio, utilizando el criterio de distancia con un umbral de 20 km. El patrón de conectividad resultante resalta áreas con mayor densidad de conexiones, especialmente en la parte central de la región de estudio, donde el área de las cuencas es más pequeña y existe mayor proximidad entre unidades. Estas áreas con mayor número de vecinos indican una interacción espacial más fuerte.

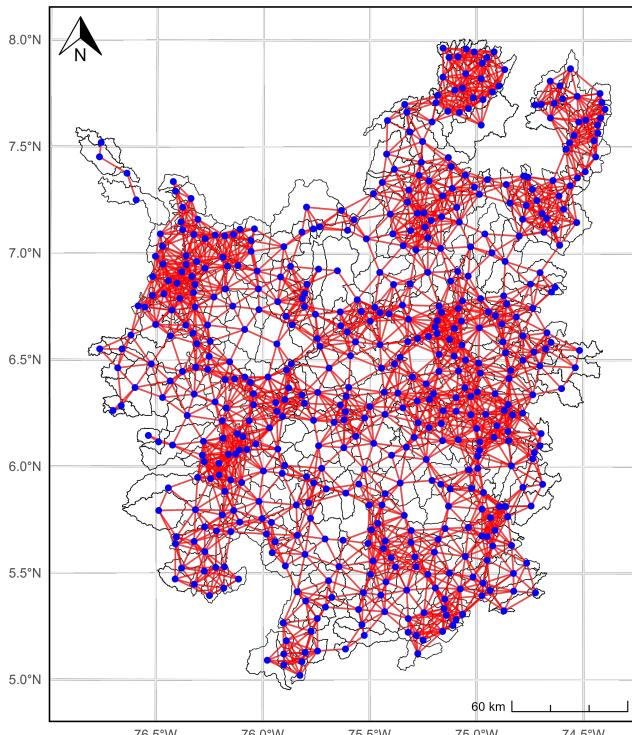


Figure 3. Ejemplo de la matriz de vecindad del área de estudio, basada en el criterio de distancia con valor de 20 km. Los puntos azules representan los centroides de las subcuencas y las líneas rojas indican las conexiones entre subcuencas consideradas vecinas dentro del umbral de 20 km.

La Figura 4 muestra los resultados del análisis del índice de Moran para la variable área de la cuenca, con la matriz de vecindad previamente obtenida. La línea negra indica el ajuste lineal que se corresponde con el cálculo del índice de Moran, el cual tiene un valor de 0.1. Este valor positivo, aunque relativamente bajo, indica una ligera autocorrelación espacial, sugiriendo que las áreas de las subcuencas están moderadamente correlacionadas con las áreas de sus subcuencas vecinas.

La Figura 5 presenta el análisis del índice de Moran para la variable Curva Hipsométrica. El índice de Moran es 0.39, un valor que indica una autocorrelación espacial moderada. Los puntos en el cuadrante Q1 (alto-alto) y Q3 (bajo-bajo) destacan la existencia de subcuencas que tienden a agruparse con características hipsométricas similares.

La Figura 6 presenta el análisis del índice de Moran para la variable pendiente media de las subcuencas. Este patrón sugiere una clara tendencia espacial, con áreas de alta pendiente agrupadas en ciertas partes de la región de estudio. El índice de Moran obtenido es 0.81, un valor elevado que indica una fuerte autocorrelación espacial.

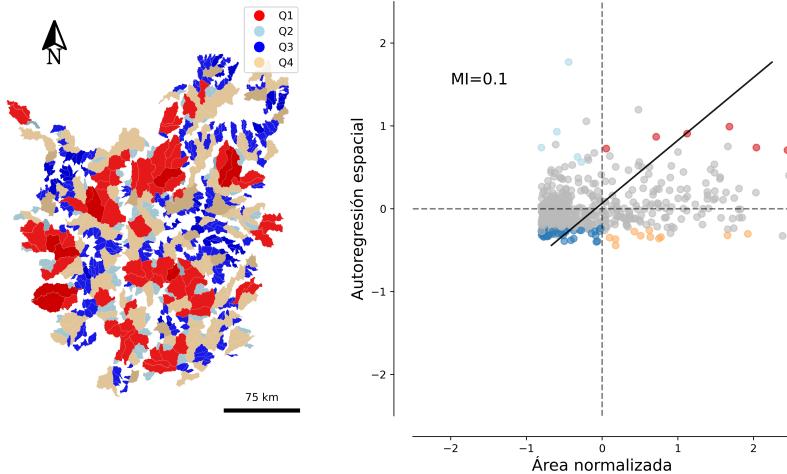


Figure 4. Análisis del índice de Moran para la variable Área de la cuenca con una matriz de vecindad con una distancia de 20 km. El panel derecho muestra el mapa de las subcuenas clasificadas en cuatro cuadrantes del diagrama de dispersión de Moran: Q1 (alto-alto, rojo), Q2 (bajo-alto, azul claro), Q3 (bajo-bajo, azul) y Q4 (alto-bajo, rosado), los colores en gris representan observaciones que no son estadísticamente significativas. El panel derecho presenta el diagrama de dispersión de Moran, donde la línea negra indica el índice de Moran ($MI = 0.1$), sugiriendo una ligera autocorrelación espacial para el área de las subcuenas.

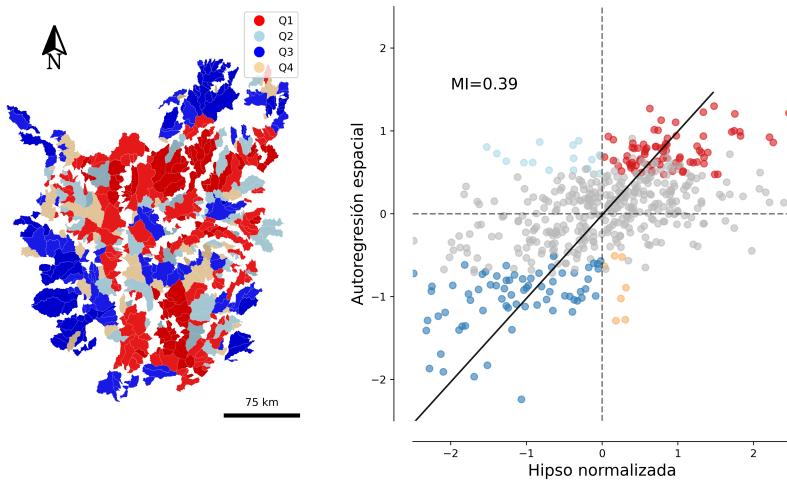


Figure 5. Análisis del índice de Moran para la variable Curva Hipsométrica con una matriz de vecindad con una distancia de 20 km.

La Figura 7 presenta el análisis del índice de Moran para la variable precipitación media anual de las subcuenas. Este patrón espacial es muy evidente en la región de estudio, lo cual indica una tendencia clara en la distribución de la precipitación. El índice de Moran (0.91) sugiere una autocorrelación espacial muy fuerte. Los puntos del cuadrante Q1 están claramente separados, lo cual muestra una agrupación fuerte de áreas con alta precipitación, y los del cuadrante Q3 muestran una relación

242
243
244
245
246

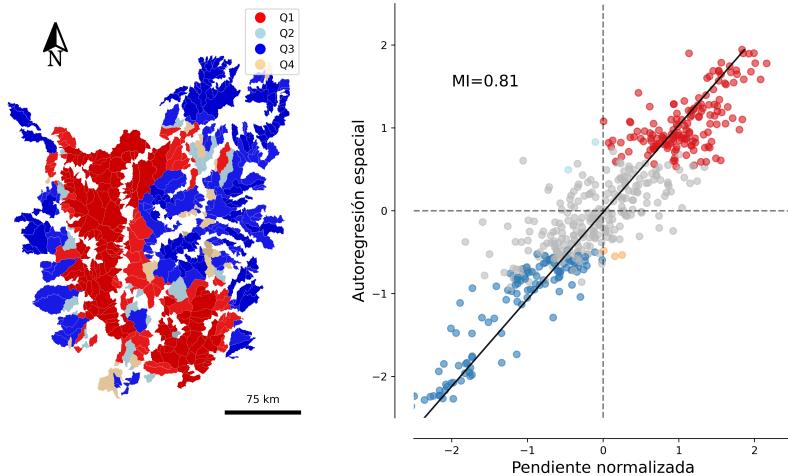


Figure 6. Análisis del índice de Moran para la variable pendiente media de la subcuenca con una matriz de vecindad con una distancia de 20 km.

similar para áreas con baja precipitación.

247

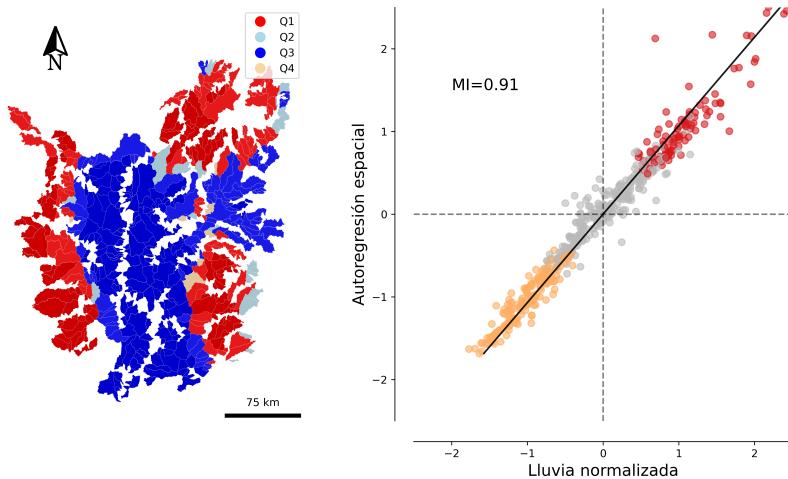


Figure 7. Análisis del índice de Moran para la variable precipitación media anual con una matriz de vecindad con una distancia de 20 km.

La Figura 8 presenta los diagramas de dispersión de Moran para la variable pendiente media, evaluados con diferentes matrices de vecindad basadas en distancias de 5, 10, 20, 50 y 100 km. Cada gráfico muestra cómo varía la autocorrelación espacial de la pendiente en función del criterio de distancia utilizado para definir los vecinos. A medida que aumenta la distancia, se observa una disminución gradual en el valor del índice de Moran (MI), que pasa de 0.95 en la distancia de 5 km a 0.37 en la distancia de 100 km. Esta tendencia indica que, para distancias menores, existe una fuerte autocorrelación espacial, lo cual significa que las subcuencas con pendientes similares tienden a estar muy

248

249

250

251

252

253

254

próximas. Conforme la distancia aumenta, la autocorrelación se debilita, sugiriendo que la influencia de la pendiente media disminuye a medida que se consideran subcuenas más distantes entre sí.

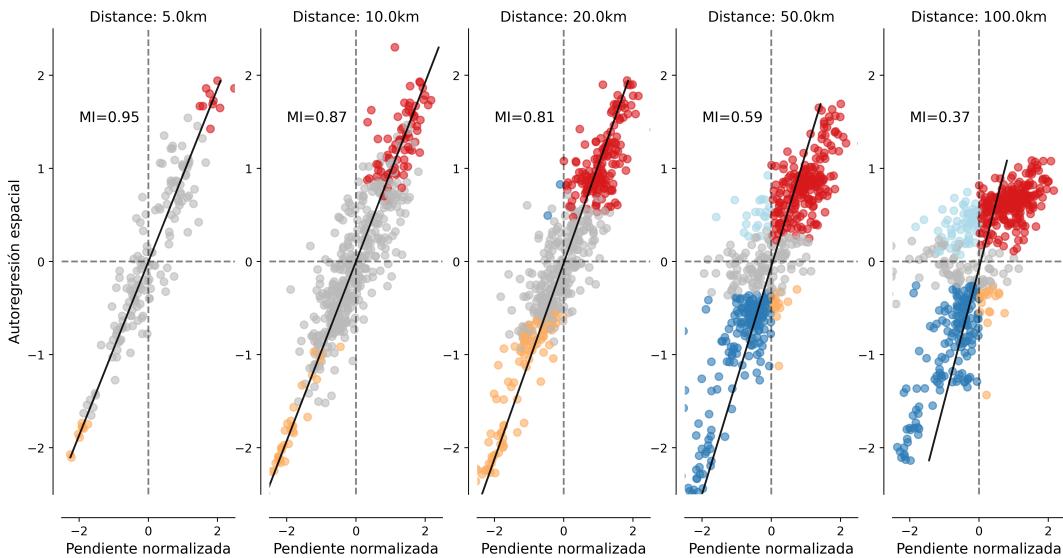


Figure 8. Diagramas de dispersión de Moran para la variable pendiente media, utilizando matrices de vecindad basadas en diferentes distancias (5, 10, 20, 50 y 100 km).

La Figura ?? presenta el cambio en el índice de Moran para cada variable predictora. Cada gráfico evalúa el índice de Moran utilizando matrices de vecindad basadas en distancias que van desde 5 km hasta 200 km, aumentando cada 5 km. La figura señala cómo cambia la autocorrelación espacial de cada variable a medida que se expande el radio de influencia entre subcuenas. Además, cada punto está etiquetado con el porcentaje de observaciones (subcuenas) que muestran resultados estadísticamente significativos, lo cual proporciona una visión adicional sobre la consistencia de la autocorrelación espacial a diferentes escalas.

Para el área de la cuenca (superior izquierda) se observa una disminución rápida del índice de Moran conforme aumenta la distancia. El porcentaje de subcuenas con resultados estadísticamente significativos es relativamente bajo, con un máximo del 17% en distancias cortas y disminuyendo rápidamente, lo cual indica que la autocorrelación espacial del área es moderada y de alcance local. Para la variable curva hipsométrica (superior derecha) el índice de Moran comienza en valores más elevados (alrededor de 0.6) y disminuye gradualmente hasta 0.2. La autocorrelación espacial para esta variable se mantiene significativa en una mayor proporción de subcuenas (hasta 41% en distancias cortas), lo cual sugiere que la distribución de la hipsometría tiene un patrón más robusto y generalizado en el área de estudio. Para la precipitación media anual (inferior derecha) se observa inicialmente un índice de Moran cercano a 1, lo cual indica una fuerte autocorrelación espacial que se diluye lentamente a medida que aumenta la distancia, pero aún mantiene valores relativamente altos (> 0.5) incluso a mayores distancias. El porcentaje de subcuenas con resultados significativos es también muy alto (> 60%), lo cual refleja la homogeneidad espacial de la precipitación en el área de estudio. Finalmente, para la pendiente media (inferior izquierda) el índice de Moran para la pendiente también comienza en valores altos, alrededor de 0.95, y disminuye conforme aumenta la distancia, aunque sigue siendo significativo hasta distancias intermedias. El porcentaje de subcuenas con resultados significativos es alto en distancias cortas (hasta 53%) y luego disminuye gradualmente, indicando que la pendiente tiene una fuerte correlación espacial local que se debilita lentamente.

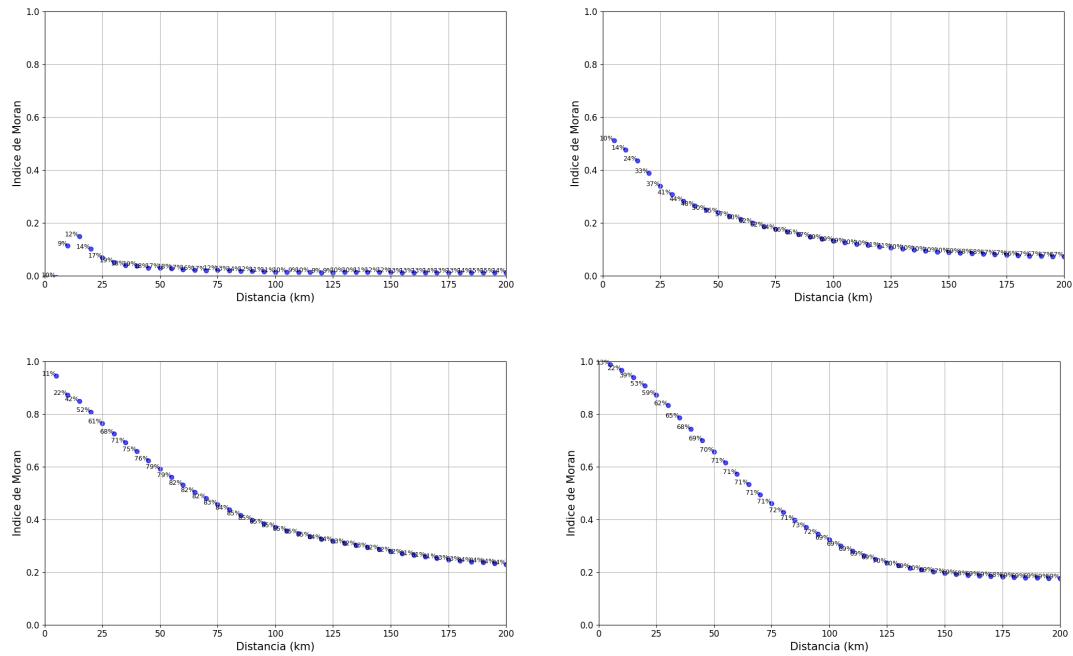


Figure 9. Cambios en el índice de Moran para las variables Área de la Cuenca (superior derecha), Curva Hipsométrica (superior izquierda), Precipitación Media Anual (inferior derecha) y Pendiente Media (inferior izquierda), utilizando matrices de vecindad basadas en distancias de 5 km a 200 km, con incrementos de 5 km. Cada punto muestra el porcentaje de subcuenas con resultados estadísticamente significativos.

La Figura ?? presenta el índice de Moran utilizando como matriz de vecindad el criterio de los vecinos más cercanos, con valores de K variando entre 1 y 200. El índice de Moran para el área de la cuenca comienza en valores bajos (aproximadamente 0.2) cuando se consideran pocos vecinos (K = 1), y disminuye rápidamente hasta alcanzar casi cero para valores de K superiores. El porcentaje de subcuenas con resultados significativos también es bajo, alcanzando un máximo de 17%. La autocorrelación espacial para la variable hipsométrica muestra valores iniciales del índice de Moran más altos (cerca de 0.6) cuando se consideran pocos vecinos, lo cual decrece gradualmente hasta alcanzar valores cercanos a cero para K mayores a 100. El porcentaje de significancia es mayor en comparación con el área, alcanzando hasta un 42%. El índice de Moran para la precipitación es inicialmente muy alto, cercano a 1, disminuyendo de forma sostenida conforme aumenta K. Para la pendiente media, el índice de Moran también comienza con valores elevados (alrededor de 0.95), y se reduce de manera constante a medida que se incrementa K. El porcentaje de subcuenas con significancia estadística es alto (más del 80%) para valores pequeños de K y disminuye progresivamente.

La Tabla 2 presenta los resultados de los modelos espaciales que incluyen la dependencia espacial en la modelación de la ocurrencia de movimientos en masa: SAR (*Spatial Autoregressive Model*), SEM (*Spatial Error Model*), SAC (*Spatial Autoregressive Combined Model*), SLX (*Spatial Lagged X Model*), SDEM (*Spatial Durbin Error Model*), SDM (*Spatial Durbin Model*), y GNS (*General Nesting Spatial Model*). La variable curva hipsométrica tiene un coeficiente positivo y significativo en todos los modelos, lo que sugiere que tiene un efecto robusto en la ocurrencia de deslizamientos. Por otro lado, la variable Lluvia presenta un efecto negativo en la mayoría de los modelos, aunque en algunos casos sin alcanzar la significancia. Además, se presentan los coeficientes asociados a las dependencias espaciales, como ρ (autocorrelación espacial en la variable dependiente) y λ (autocorrelación en los errores). Los valores de ρ y λ son elevados en los modelos SAR y SEM, respectivamente, indicando

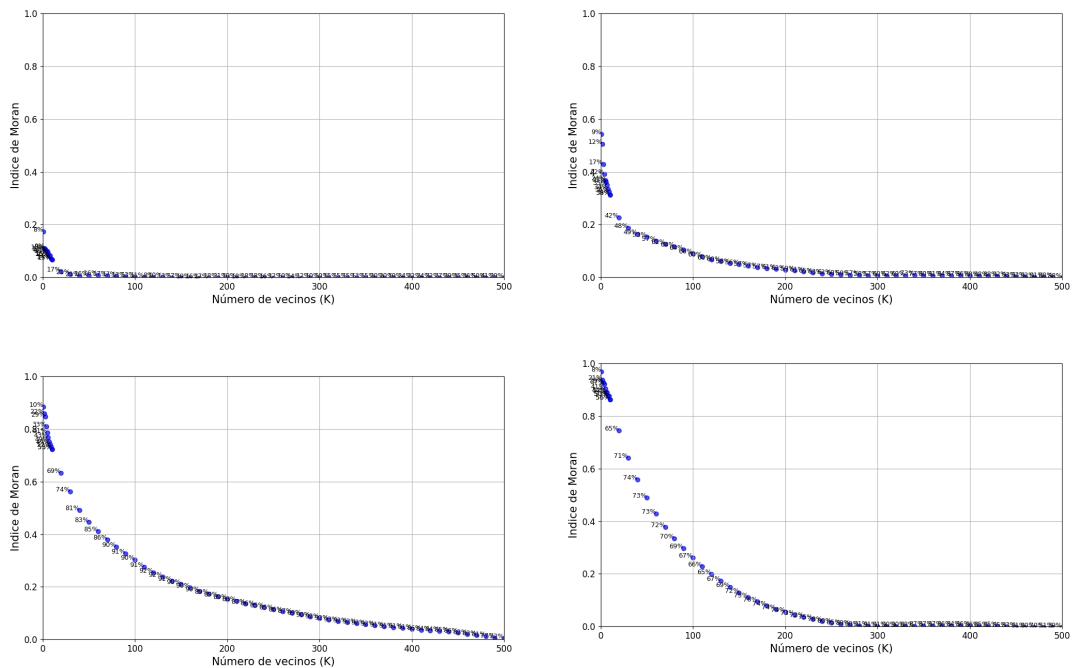


Figure 10. Cambios en el índice de Moran para las variables Área de la Cuenca, Curva Hipsométrica, Precipitación Media Anual y Pendiente Media, utilizando matrices de vecinos más cercanos (KNN) con valores de K entre 1 y 10 (incrementos de 1) y de 10 hasta 200 (incrementos de 5). Cada punto muestra el valor del índice de Moran y el porcentaje de subcuenca con resultados estadísticamente significativos.

que la dependencia espacial está capturando efectos importantes no modelados por los predictores. También se incluye el parámetro Wx para las variables explicativas en los modelos que consideran interacciones espaciales con los predictores (SLX, SDEM, SDM, y GNS). El modelo SAC tiene el menor valor de AIC (1455), lo cual sugiere que este modelo ofrece el mejor equilibrio entre ajuste y complejidad, seguido de cerca por el SDEM (1460) con un R^2 ligeramente mayor.

La Tabla 3 presenta los resultados del modelo SDEM considerando matrices de vecindad basadas en distancias de 20, 40, 60, 80, y 100 km. A medida que aumenta la distancia de vecindad, se observa una reducción en la magnitud y significancia de varios parámetros. Por ejemplo, el intercepto disminuye de un valor significativo en 20 km (1.82) a valores no significativos a partir de los 60 km, sugiriendo que la capacidad del modelo para capturar patrones consistentes disminuye con el incremento de la distancia. La variable área de la cuenca mantiene un coeficiente significativo en todas las distancias, lo cual destaca su relevancia en la explicación de la ocurrencia de deslizamientos. La variable hipso muestra un coeficiente positivo que pierde significancia conforme aumenta la distancia, mientras que la pendiente presenta un coeficiente positivo y significativo en todas las distancias, lo cual indica que la pendiente sigue siendo un factor importante de acuerdo con el alcance espacial considerado. La variable lluvia, por otro lado, muestra un comportamiento mixto, siendo significativa solo en la distancia de 60 km y pasando a no ser significativa a mayores distancias. En cuanto a los efectos espaciales (Wx), los coeficientes correspondientes a las variables predictoras muestran una alta variabilidad. El coeficiente de la pendiente autoregresión espacial (Wx – Pendiente) es significativo para la distancia de 20 km, lo que sugiere que la influencia de la pendiente en las subcuenca vecinas se percibe a esta escala. El coeficiente de Nagelkerke R^2 disminuye conforme aumenta la distancia, lo cual indica una reducción en la capacidad explicativa del modelo con el incremento de

305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325

Table 2. Resultados de diferentes modelos espaciales aplicados para la susceptibilidad por movimientos en masa. Se presentan los coeficientes estimados (con errores estándar entre paréntesis y significado estadístico en negrilla) para las variables Área, Hipso, Pendiente y Lluvia, así como los coeficientes espaciales ρ y λ . También se presenta el ajuste del modelo a través del coeficiente de Nagelkerke R^2 y el AIC.

Parámetro	SAR	SEM	SAC	SLX	SDEM	SDM	GNS
Intercepto	0.651 (0.08)	1.819 (0.14)	1.129 (0.23)	1.859 (0.05)	1.823 (0.14)	0.549 (0.08)	1.558 (0.52)
Área	0.105 (0.05)	0.106 (0.05)	0.106 (0.05)	0.101 (0.06)	0.105 (0.05)	0.103 (0.05)	0.106 (0.05)
Hipso	0.556 (0.04)	0.588 (0.04)	0.594 (0.04)	0.592 (0.05)	0.603 (0.04)	0.587 (0.04)	0.601 (0.04)
Pendiente	0.295 (0.05)	0.534 (0.08)	0.452 (0.09)	0.477 (0.13)	0.415 (0.10)	0.411 (0.10)	0.407 (0.10)
Lluvia	-0.106 (0.05)	-0.380 (0.11)	-0.211 (0.09)	-0.636 (0.30)	-0.452 (0.27)	-0.445 (0.24)	-0.452 (0.27)
Wx-Área				0.185 (0.12)	0.066 (0.16)	-0.012 (0.10)	0.060 (0.15)
Wx-Hipso					0.290 (0.14)	0.209 (0.16)	-0.280 (0.12)
Wx-Pendiente					0.238 (0.15)	0.387 (0.17)	-0.170 (0.13)
Wx-Lluvia					0.433 (0.32)	0.173 (0.33)	0.380 (0.26)
ρ	0.657		0.388			0.702	0.148
λ		0.716	0.464		0.708		0.636
Nagelkerke R^2	0.624	0.627	0.631		0.632	0.631	0.633
AIC	1464.141	1460.692	1455.857	1642.684	1460.333	1461.73	1461.973

la vecindad. El AIC también aumenta progresivamente a medida que aumenta la distancia, lo que refleja la reducción del ajuste del modelo. Se observa que para distancias mayores a 20 km, los valores de MI aumentan y sus p -valores se tornan significativos ($p < 0.05$), sugiriendo la presencia de autocorrelación residual que no está siendo capturada por el modelo.

La Figura 11 presenta los resultados del modelo SDEM con los residuales del modelo agrupados por cuencas hidrográficas y la distribución espacial de los valores estimados para cada subcuenca en el área de estudio. En los diagramas de caja, se observa que la mediana de los residuales en las tres cuencas se encuentra cerca de cero, lo cual indica que el modelo no presenta un sesgo significativo.

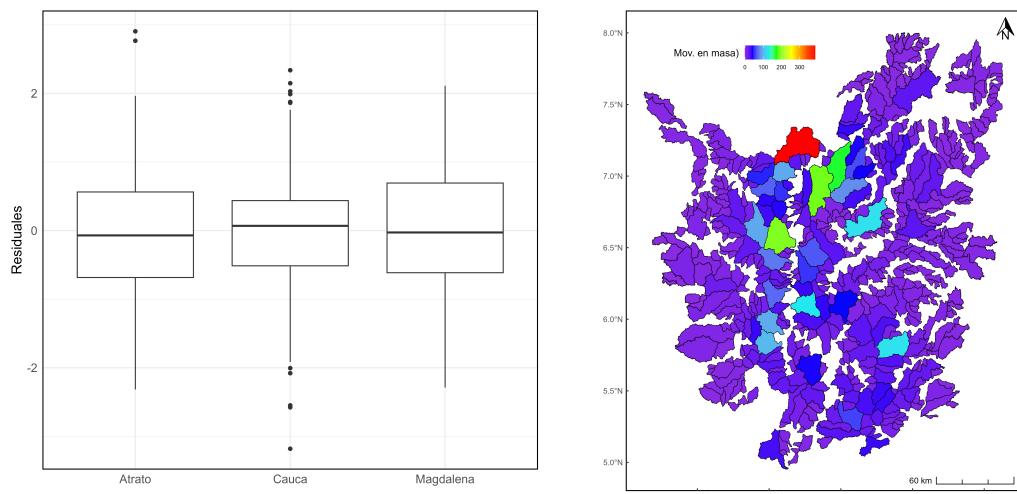


Figure 11. Resultados del modelo SDEM. El panel izquierdo muestra diagramas de caja de los residuales agrupados por cuenca hidrográfica (Atrato, Cauca y Magdalena), indicando la variabilidad y presencia de valores atípicos. El panel derecho presenta la distribución espacial de los valores estimados de deslizamientos por subcuenca, con un esquema de colores que varía de tonos bajos (morado) a altos (rojo).

Table 3. Resultados del modelo SDEM para diferentes distancias de vecindad. Se presentan los coeficientes estimados (con errores estándar entre paréntesis) para las variables predictoras, junto con los coeficientes espaciales (Wx) y el coeficiente de autocorrelación en los errores (λ). También se muestran el coeficiente de Nagelkerke R^2 , el AIC, el índice de Moran (MI) y su p -valor

Parámetro	20km	40km	60km	80km	100km
Intercepción	1.82 (0.14)	1.74 (0.32)	1.66 (0.54)	1.42 (0.87)	1.29 (1.03)
Área	0.60 (0.04)	0.57 (0.04)	0.57 (0.04)	0.57 (0.05)	0.57 (0.05)
Hipso	0.11 (0.05)	0.09 (0.05)	0.09 (0.05)	0.09 (0.05)	0.08 (0.05)
Pendiente	0.42 (0.10)	0.43 (0.10)	0.53 (0.09)	0.56 (0.09)	0.51 (0.09)
Lluvia	-0.45 (0.27)	-0.24 (0.20)	-0.40 (0.16)	-0.21 (0.17)	0.00 (0.17)
Wx -Área	0.21 (0.16)	-0.44 (0.38)	0.43 (0.53)	0.96 (0.71)	1.29 (0.86)
Wx -Hipso	0.07 (0.16)	-0.04 (0.32)	-0.02 (0.42)	0.04 (0.50)	0.44 (0.54)
Wx -Pendiente	0.39 (0.17)	0.50 (0.27)	0.41 (0.34)	0.34 (0.42)	0.41 (0.45)
Wx -Lluvia	0.17 (0.33)	-0.06 (0.36)	0.33 (0.36)	-0.05 (0.47)	-0.73 (0.63)
λ	0.708	0.871	0.919	0.949	0.957
Nagelkerke R^2	0.632	0.616	0.59	0.574	0.568
AIC	1460.333	1482.916	1518.336	1537.402	1545.834
MI	0.006	0.026	0.031	0.04	0.041
p -value Moran	0.361	0.015	0.001	0.001	0.001

5. Discusión

Durante muchas décadas, se ha reconocido el componente espacial de los datos, especialmente en disciplinas relacionadas con la geografía, econometría y epidemiología (Anselin 2022; S. Rey, Arribas-Bel, and Wolf 2023). Sin embargo, en la práctica, la incorporación efectiva de este componente en los modelos estadísticos ha sido limitada. Solo recientemente, con los avances acelerados en la captura de datos espaciales, mediante el sensoramiento remoto y el Internet de las cosas (IoT), así como con el aumento de la capacidad de almacenamiento y procesamiento de los ordenadores, se han desarrollado técnicas que permiten abordar de forma más adecuada la espacialidad de los datos. Estos avances han facilitado la implementación de modelos geoespaciales más robustos, como los modelos de autoregresión espacial (Wall 2004), los modelos multiniveles (Lee and Nelder 1996) o los procesos Gaussianos (Vasudevan et al. 2009), que permiten incorporar la estructura espacial inherente a los fenómenos naturales.

La omisión de la estructura espacial en modelos estadísticos clásicos puede generar resultados sesgados o no representativos, ya que dichos modelos generalmente asumen independencia entre las observaciones (Anselin and Griffith 1988; J. P. LeSage and Pace 2011). Esta estructura espacial en los datos viola supuestos básicos de los modelos de regresión clásica, como la independencia de los residuos, la homocedasticidad (varianza constante de los residuos), y la ausencia de multicolinealidad entre los predictores (Cressie 1988; Ripley 1988; Anselin and Griffith 1988). El impacto de la dependencia espacial en las pruebas estadísticas y en las medidas de ajuste puede ser engañoso debido a la estimación sesgada de la varianza del error, los niveles de significancia de las pruebas T y los valores de R^2 (Anselin 1990).

Tener una estructura espacial significa que los datos exhiben una correlación debido a su proximidad en el espacio, lo que resulta en patrones que deben ser adecuadamente modelados. La heterogeneidad espacial, que se refiere a la variación en las propiedades a lo largo del espacio, y la dependencia espacial, que implica que los valores en ubicaciones cercanas tienden a estar correlacionados, son las estructuras espaciales más comunes descritas en la literatura (Anselin and Griffith 1988; S. Rey,

Arribas-Bel, and Wolf 2023). La dependencia espacial es observada en procesos naturales; por ejemplo, en el caso de los movimientos en masa, las áreas con alta susceptibilidad tienden a influir en áreas vecinas debido a la continuidad de factores como la pendiente, la litología y la cobertura del suelo (Samia *et al.* 2019). Esto implica que un evento en una ubicación específica puede aumentar la probabilidad de ocurrencia en ubicaciones cercanas, generando patrones espaciales de correlación que los modelos deben incorporar.

La naturaleza espacial de los datos exige la implementación de modelos que puedan manejar explícitamente esta característica. En el presente estudio, utilizamos modelos de regresión espacial para abordar la dependencia espacial de manera adecuada. Los modelos SAR permiten capturar dicha dependencia a través de una matriz de vecindad, la cual puede definirse según criterios de contigüidad o distancia. Los resultados señalan cómo la elección del criterio de vecindad tiene un impacto directo en los resultados del modelo, ya que determina cómo se representan las interacciones espaciales y cómo se propaga la influencia entre las unidades de análisis. Por ejemplo, una matriz basada en la contigüidad puede ser importante cuando los bordes entre las unidades discretas utilizadas controlen o determine la interacción entre vecinos. Mientras que una matriz basada en la distancia puede tener en cuenta efectos que se extienden más allá de los límites inmediatos, lo cual puede ser crucial dependiendo del fenómeno estudiado. Para el caso de este estudio, hemos seleccionado trabajar con matrices basadas en distancia, ya que los parámetros morfométricos, como pendiente e hipsometría, así como la precipitación, se manifiestan en el terreno sin seguir un límite o borde establecido. El caso de la precipitación por ejemplo, aunque puede tener influencias del relieve, no presenta límites basados en contigüidad. Por lo tanto una aproximación basada en distancias puede representar de mejor forma el proceso espacial subyacente. Sin embargo, esto debe ser evaluado en cada caso de aplicación.

La inclusión explícita de la dependencia espacial no solo puede mitigar los problemas derivados de la asunción de independencia en los modelos estadísticos convencionales, sino que también aporta información adicional que mejora el desempeño y capacidad predictiva del modelo. En este estudio, la consideración de la dependencia espacial ha mejorado tanto la interpretabilidad como la precisión del modelo, dado que se incorpora la estructura espacial subyacente al fenómeno estudiado. Esto ha permitido una representación que se puede ajustar mejor a los patrones observados y una mejora significativa en la calidad de las predicciones, comparado con los modelos clásicos que no consideran este aspecto fundamental.

Para este estudio se emplearon datos de subcuencas localizadas en los Andes colombianos, junto con variables predictoras morfométricas que operan tanto a escala local como regional. Por ejemplo, se consideraron variables locales como la pendiente, y variables regionales como la precipitación media anual. La combinación de variables con efectos locales o efectos regionales es observada, y en este caso estimadas, para casos como la evaluación de la susceptibilidad por movimientos en masa, lo que contribuye a una comprensión más profunda de la influencia espacial y la estructura del fenómeno.

Los resultados del estudio evidenciaron la presencia de una estructura espacial en función de la distancia para las variables consideradas. En el caso del área de la cuenca, no se identificó una estructura espacial significativa, mientras que la hipsometría de la cuenca mostró una estructura espacial moderada. Por otro lado, las variables de pendiente y precipitación evidenciaron un fuerte componente espacial, con una marcada tendencia a la formación de clústeres de valores altos y bajos.

Para evaluar la dependencia espacial, se analizó el índice de Moran en función de la distancia para todas las variables, considerando la significancia estadística de las observaciones. Se encontró que, al emplear distancias cortas, la correlación espacial aumentaba, pero también lo hacía la varianza, lo cual disminuía la significancia estadística en muchas observaciones. Esto se debe a que un aumento en la varianza introduce una mayor incertidumbre en los estimadores del modelo, lo cual dificulta

la detección de patrones significativos y disminuye la capacidad para rechazar hipótesis nulas. A distancias mayores, la correlación espacial disminuía, al igual que la varianza.

Los resultados arrojaron que para la pendiente y la precipitación media anual a medida que aumenta la distancia se reduce la influencia entre las subcuenca, pero a diferentes tasas. Para la pendiente se observa una tasa de reducción menor comparada con la precipitación; mientras que para la precipitación media anual es mayor a distancias cortas pero se va reduciendo a medida que aumenta la distancia. En ambos casos se alcanzan valores alrededor de 0,2 para distancias mayores a 100 km.

También se implementó el criterio de los vecinos más cercanos para obtener las matrices de vecindad, los resultados indicaron que la matriz basada en distancia capturó de manera más adecuada la dependencia espacial de los datos. Las curvas de la variación del índice de Moran con el número de vecinos mostraron tendencias similares al criterio de distancia, pero con un área bajo la curva menor, indicando una menor capacidad explicativa en comparación con el criterio basado en la distancia.

En cuanto al modelo de regresión clásico, los resultados, junto con el índice de Moran y la prueba de Lagrange, evidenciaron la presencia de estructura espacial en los residuales del modelo. Por tanto, se implementaron modelos de autoregresión espacial. Las matrices espaciales en los modelos SAR permiten la incorporación de la dependencia en las variables exógenas (predictores) o variable endógena (el término de error o la variable respuesta). J. P. LeSage 2014 propone considerar solo dos modelos espaciales: el SDM y el SDEM. El SDM asume que los efectos de vecindad son globales por que la autoregresión es simultánea al estar en la variable respuesta, mientras que el SDEM se encuentra en el término del error por lo que se asume que son locales. Los autores recomiendan usar un modelo SEM cuando no haya una comprensión clara del origen del efecto de dependencia (J. P. Elhorst et al. 2014). Otra recomendación está en que si no existen razones teóricas que justifiquen un modelo específico, podría ser preferible confiar en la especificación SLX más simple, en lugar de adoptar el modelo más complejo SDM (Halleck Vega and J. P. Elhorst 2017; Halleck Vega and J. P. Elhorst 2015).

En este caso implementamos todos los modelos para comparar sus resultados. Para la construcción de los modelos se utilizó una distancia de 25 km como criterio de vecindad, dado que las variables con mayor dependencia espacial, como la pendiente y la precipitación, presentaban altos niveles de dependencia en este rango, y más del 60% de los datos mostraban significancia estadística. Esta elección optimizó el equilibrio entre la correlación espacial y la varianza, obteniendo resultados robustos y significativos. Aunque los resultados fueron satisfactorios, cabe mencionar que existen otros criterios para optimizar la matriz de vecindad que no se exploraron en este trabajo. Por ejemplo, el punto de inflexión de la curva de dependencia podría ser utilizado para establecer una distancia óptima en la matriz de vecindad. Este enfoque podría proporcionar una alternativa para representar la estructura espacial de los datos y mejorar la capacidad predictiva del modelo.

La prueba de Lagrange en los residuales del modelo clásico sugirió la incorporación de la dependencia en el término de error, lo cual fue posteriormente confirmado. En términos del coeficiente de determinación ajustado (R^2 ajustado) y el criterio de información de Akaike (AIC), los modelos con mejor desempeño fueron el SAC, SEM y el SDEM, con resultados similares, todos ellos comparten en común que el componente de autoregresión es incorporado en el error. Aunque el SDEM presentó un R^2 ajustado ligeramente superior.

La incorporación de la estructura espacial mejoró significativamente los resultados en comparación con el modelo clásico, eliminando la estructura espacial de los residuales, lo cual confirma la relevancia de considerar la dependencia espacial en el modelado de fenómenos naturales como los movimientos en masa. Además, el SDEM, que presentó el mejor desempeño, fue utilizado para construir distintas versiones del modelo, modificando la matriz de vecindad en función de la distancia. Se observó que, a medida que aumentaba la distancia para definir la vecindad, los valores de R^2 ajustado

tado disminuían, mientras que el índice de Moran en los residuales y el coeficiente de autoregresión del error (λ) aumentaban de manera significativa.

6. Conclusiones

Los hallazgos de este estudio subrayan la importancia de integrar la dependencia espacial en la modelación de la susceptibilidad a movimientos en masa. Los modelos de regresión tradicionales, que generalmente asumen la independencia de las observaciones, no logran capturar adecuadamente la estructura espacial inherente a los procesos geomorfológicos como los movimientos en masa, lo que resulta en estimaciones sesgadas, precisión reducida y conclusiones potencialmente erróneas. Al incorporar la dependencia espacial, los modelos pueden representar de manera más precisa las dinámicas de los movimientos en masa, mejorando así la capacidad predictiva y la fiabilidad de los resultados.

Los modelos autoregresivos espaciales, en particular el Modelo de Error de Durbin Espacial (SDEM), demostraron en nuestra zona de estudio ser efectivos para abordar las interacciones espaciales que los modelos clásicos no logran capturar. El SDEM, que incorpora la dependencia espacial tanto en las variables predictoras como en los residuales, mostró un desempeño superior en varias métricas de evaluación, incluyendo el R^2 ajustado y el AIC. Al tener en cuenta los efectos de vecindad locales, el SDEM ofreció una comprensión más integral de cómo factores geoambientales como la pendiente, la precipitación y la hipsometría contribuyen a la susceptibilidad por movimientos en masa.

El estudio también destaca la importancia de seleccionar una matriz de vecindad adecuada. En este caso, una matriz de vecindad basada en la distancia capturó de manera efectiva las dinámicas espaciales de variables clave, como la pendiente y la precipitación anual. Nuestros datos sugieren que el uso de matrices basadas en la distancia proporciona una representación más flexible y precisa de las interacciones espaciales, que no necesariamente se ajustan a límites administrativos o naturales.

Variables como la pendiente y la precipitación mostraron una notable agrupación espacial, lo que hace necesario incluir estos componentes espaciales en los modelos predictivos para mejorar la precisión. El uso del Índice de Moran y las pruebas del Multiplicador de Lagrange validó aún más la necesidad de la modelación espacial, ya que estas pruebas identificaron una autocorrelación espacial sustancial en los residuales de los modelos de regresión tradicionales.

Acknowledgement

El presente trabajo se realizó con el apoyo de la Fundación Alexander von Humboldt (<https://www.humboldt-foundation.de/>) a través del programa *Georg Forster Research Fellowship* para investigadores.

7. Open Research

Data and code used for this study are fully available in Github (<https://github.com/edieraristizabal/ModelosAutoregresionEspacial>).

References

- Anselin, Luc (1990). "Spatial dependence and spatial structural instability in applied regression analysis". In: *Journal of Regional science* 30.2, pp. 185–207.
- (2022). "Spatial econometrics". In: *Handbook of spatial analysis in the social sciences*, pp. 101–122.
- Anselin, Luc, Anil K Bera, et al. (1996). "Simple diagnostic tests for spatial dependence". In: *Regional science and urban economics* 26.1, pp. 77–104.

- Anselin, Luc and Daniel A Griffith (1988). "Do spatial effects really matter in regression analysis?" In: *Papers in Regional Science* 65.1, pp. 11–34. 495
496
- Banerjee, Sudipto et al. (2008). "Gaussian predictive process models for large spatial data sets". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 70.4, pp. 825–848. 497
498
- Brabb, Earl E (1984). "Innovative approaches to landslide hazard and risk mapping." In. 499
- Cliff, Andrew David and J Keith Ord (1981). "Spatial processes: models & applications". In: *(No Title)*. 500
- Corominas, Jordi et al. (2014). "Recommendations for the quantitative analysis of landslide risk". In: *Bulletin of engineering geology and the environment* 73, pp. 209–263. 501
502
- Cressie, Noel (1988). "Spatial prediction and ordinary kriging". In: *Mathematical geology* 20, pp. 405–421. 503
504
- (2015). *Statistics for spatial data*. John Wiley & Sons. 505
- Dai, FC et al. (2001). "Assessment of landslide susceptibility on the natural terrain of Lantau Island, Hong Kong". In: *Environmental geology* 40, pp. 381–391. 506
- Earnest, Arul et al. (2007). "Evaluating the effect of neighbourhood weight matrices on smoothing properties of Conditional Autoregressive (CAR) models". In: *International journal of health geographics* 6, pp. 1–12. 507
508
509
- Elhorst, J Paul et al. (2014). *Spatial econometrics: from cross-sectional data to spatial panels*. Vol. 479. Springer. 510
511
512
- Elhorst, J Paul (2022). "The dynamic general nesting spatial econometric model for spatial panels with common factors: Further raising the bar". In: *Review of Regional Research* 42.3, pp. 249–267. 513
- Elhorst, Paul and Solmaria Halleck Vega (2013). "On spatial econometric models, spillover effects, and W". In. 514
515
- Fell, Robin et al. (2008). "Guidelines for landslide susceptibility, hazard and risk zoning for land use planning". In: *Engineering geology* 102.3-4, pp. 85–98. 516
517
- Funk, Chris, Pamela Peterson, and Martin Landsfeld (2015). "The Climate Hazard Group InfraRed Precipitation with Station Data (CHIRPS)". In: *Bulletin of the American Meteorological Society* 96, pp. 1401–1419. 518
519
520
- Getis, Arthur (2009). "Spatial weights matrices". In: *Geographical Analysis* 41.4, pp. 404–410. 521
- Halleck Vega, Solmaria and J Paul Elhorst (2015). "The SLX model". In: *Journal of Regional Science* 55.3, pp. 339–363. 522
523
- (2017). "Regional labour force participation across the European Union: a time–space recursive modelling approach with endogenous regressors". In: *Spatial Economic Analysis* 12.2-3, pp. 138–160. 524
525
526
- Jaya, IGMN and A Chadidjah (2021). "Spatial Autoregressive in Ecological Studies: A Comparison of the SAR and CAR Models." In: *Engineering Letters* 29.1. 527
- Kelejian, Harry H and Ingmar R Prucha (2007). "The relative efficiencies of various predictors in spatial econometric models containing spatial lags". In: *Regional Science and Urban Economics* 37.3, pp. 363–374. 528
529
530
- Korup, Oliver and Amelie Stolle (2014). "Landslide prediction from machine learning". In: *Geology today* 30.1, pp. 26–33. 531
532
- Lee, Youngjo and John A Nelder (1996). "Hierarchical generalized linear models". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 58.4, pp. 619–656. 533
534
- LeSage, James and Robert Kelley Pace (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC. 535
536
- LeSage, James P (2014). "Spatial econometric panel data model specification: A Bayesian approach". In: *Spatial statistics* 9, pp. 122–145. 537
538
- LeSage, James P and Robert Kelley Pace (2011). "Pitfalls in higher order model extensions of basic spatial regression methodology". In: *Review of Regional Studies* 41.1, pp. 13–26. 539
540
541
542

- Logan, Thomas L. and Channing L. Karr (2014). "Advanced Land Observing Satellite (ALOS) - Phased Array-Type L-Band Synthetic Aperture Radar (PALSAR)". In: *Journal of Geospatial Research* 7, pp. 102–119. 543
544
545
- Lombardo, Luigi, Haakon Bakka, et al. (2019). "Geostatistical modeling to capture seismic-shaking patterns from earthquake-induced landslides". In: *Journal of Geophysical Research: Earth Surface* 124.7, pp. 1958–1980. 546
547
548
- Lombardo, Luigi, Thomas Opitz, et al. (2020). "Space-time landslide predictive modelling". In: *Earth-Science Reviews* 209, p. 103318. 549
550
- Reichenbach, Paola et al. (2018). "A review of statistically-based landslide susceptibility models". In: *Earth-science reviews* 180, pp. 60–91. 551
552
- Rey, Sergio, Dani Arribas-Bel, and Levi John Wolf (2023). *Geographic data science with python*. Chapman and Hall/CRC. 553
554
- Rey, Sergio J. and Luc Anselin (2007). "PySAL: A Python Library of Spatial Analytical Methods". In: *The Review of Regional Studies* 37.1, pp. 5–27. 555
556
- Ripley, Brian D (1988). *Statistical inference for spatial processes*. Cambridge university press. 557
- Roger Bivand (2022). "R Packages for Analyzing Spatial Data: A Comparative Case Study with Areal Data". In: *Geographical Analysis* 54.3, pp. 488–518. doi: [10.1111/gean.12319](https://doi.org/10.1111/gean.12319). 558
559
- Samia, Jalal et al. (2019). "Dynamic path dependent landslide susceptibility modelling". In: *Natural Hazards and Earth System Sciences Discussions* 2019, pp. 1–20. 560
561
- Soeters, Robert and CJ Van Westen (1996a). "Slope instability recognition, analysis and zonation". In: *Landslides: investigation and mitigation* 247, pp. 129–177. 562
563
- (1996b). "Slope instability recognition, analysis and zonation". In: *Landslides: investigation and mitigation* 247, pp. 129–177. 564
565
- Stakhovych, Stanislav and Tammo HA Bijmolt (2009). "Specification of spatial models: A simulation study on weights matrices". In: *Papers in Regional Science* 88.2, pp. 389–409. 566
567
- Vasudevan, Shrihari et al. (2009). "Gaussian process modeling of large-scale terrain". In: *Journal of Field Robotics* 26.10, pp. 812–840. 568
569
- Ver Hoef, Jay M et al. (2018). "Spatial autoregressive models for statistical inference from ecological data". In: *Ecological Monographs* 88.1, pp. 36–59. 570
571
- Wall, Melanie M (2004). "A close look at the spatial structure implied by the CAR and SAR models". In: *Journal of statistical planning and inference* 121.2, pp. 311–324. 572
573
- Whittle, Peter (1954). "On stationary processes in the plane". In: *Biometrika*, pp. 434–449. 574