

## ARTICLE

# Influencia de la heterogeneidad 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**

This study explores the incorporation of spatial heterogeneity in modeling landslide susceptibility in the complex terrain of the Colombian Andes. The data used included a dataset of 13,777 landslides recorded between 1970 and 2023, derived from high-resolution optical imagery, as well as precipitation data from the CHIRPS dataset. Additional spatial information, such as land cover, was derived using Sentinel-2 data processed in Google Earth Engine. These diverse geospatial data sources were integrated to provide a comprehensive understanding of the factors influencing landslide susceptibility. By applying Geographically Weighted Regression (GWR) and its extension, Multiband Geographically Weighted Regression (MGWR), we demonstrate significant improvements in the capacity of models to capture local variability and enhance predictive accuracy. GWR and MGWR allow for the identification of spatially differentiated patterns and specific scales of influence for each predictor, leading to a more precise representation of landslide susceptibility and highlighting the necessity of adapting models to the unique geographic context of each region. However, the study also underscores the need for careful bandwidth selection to avoid overfitting and ensure model stability. Despite the advantages of GWR and MGWR, limitations such as high computational complexity and local collinearity were identified, which can impact the interpretation and reliability of the results.

**Keywords:** landslide; susceptibility; spatial heterogeneity

## 1. Introducción

En las últimas décadas, el empleo de modelos estadísticos o de aprendizaje de máquinas (*machine learning*) han tenido una relevancia creciente en la evaluación de la susceptibilidad a movimientos en masa (Aristizábal, García, and Martínez 2015; Tehrani et al. 2022; Korup and Stolle 2014). Estos enfoques metodológicos se han consolidado como herramientas muy útiles para construir mapas de susceptibilidad, permitiendo identificar zonas más propensas por sus condiciones naturales a experimentar movimientos en masa. Los mapas de susceptibilidad por movimientos en masa son utilizados para el ordenamiento territorial, estudios ambientales, así como para la gestión del riesgo de desastres, entre otros (Guzzetti et al. 1999; Soeters and Van Westen 1996; Corominas et al. 2014).

Los modelos estadísticos expresan la correlación entre una variable respuesta (también denominada variable dependiente, resultado o de salida) y un conjunto de variables predictoras (covariables, variables independientes o regresores) (Dai et al. 2001). Estos métodos suponen que las condiciones geológicas, topográficas y climáticas que llevaron a la ocurrencia de movimientos en masa en el pasado proporcionan la información necesaria para establecer las ubicaciones y condiciones de futuros movimientos en masa sobre las laderas. Por lo que exigen contar con un inventario de movimientos en masa (Dai et al. 2001; Brabb 1984; Soeters and Van Westen 1996).

Una de las principales limitaciones de los modelos tradicionales para la evaluación de la susceptibilidad a movimientos en masa es la suposición de homogeneidad espacial (Lombardo et al. 2020).

Los coeficientes que entregan estos modelos representan una valor medio. Es decir, estos modelos asumen que la influencia de las variables predictoras con la variable respuesta es constante en todo el área de estudio, implicando que el terreno se considera homogéneo en términos de su respuesta a los movimientos en masa. Esta simplificación permite un modelado más sencillo, pero no captura la complejidad inherente a los sistemas naturales, donde la distribución espacial e influencia de los factores condicionantes en la ocurrencia de movimientos en masa, como la geología, las formas del terreno, el clima o la vegetación, pueden variar de manera significativa.

La heterogeneidad espacial es una característica fundamental de los datos geoespaciales y suponen importantes desafíos para los modelos estadísticos convencionales (Anselin and Griffith 1988; Cressie 2015; LeSage and Pace 2009). La heterogeneidad espacial se refiere a la variabilidad de las condiciones del terreno a lo largo del espacio (Anselin 1990). Por ejemplo, las propiedades geomorfológicas de una ladera pueden diferir sustancialmente de las de otra dentro de una misma cuenca, afectando la susceptibilidad de cada zona a los movimientos en masa. Ignorar esta heterogeneidad del espacio puede llevar a una subestimación o sobreestimación de la susceptibilidad en ciertas zonas del área de estudio.

Para abordar estas limitaciones, se han desarrollado modelos que permiten la variación espacial de los coeficientes de las variables predictoras, permitiendo capturar la heterogeneidad inherente al terreno (Rey, Arribas-Bel, and Wolf 2023). Entre estos enfoques se encuentran los modelos multiniveles (Wong and Mason 1985) (también conocidos como jerárquicos, y en algunos casos modelos mixtos o regímenes espaciales) y los modelos de regresión espacial ponderada (GWR, por sus siglas en inglés) (Brunsdon, Fotheringham, and Martin E Charlton 1996). Los modelos multiniveles permiten modelar la variabilidad espacial mediante la introducción de estructuras jerárquicas, asignando un coeficiente e intercepto para cada nivel o región incorporada en el modelo. Lo cual resulta particularmente útil cuando los datos presentan una organización natural en grupos o niveles (por ejemplo, cuencas y subcuenca). En contraste, los modelos GWR son menos complejos computacionalmente, ya que no requieren definir niveles jerárquicos. Los modelos GWR permiten que los coeficientes de las variables predictoras varíen para cada observación, proporcionando una caracterización más detallada de la influencia de las variables en función de su localización geográfica. Ambos enfoques presentan ventajas significativas frente a los modelos tradicionales, al capturar la variabilidad espacial, mejorando la precisión en la evaluación de la susceptibilidad a movimientos en masa.

En este trabajo, se explora la influencia de la heterogeneidad espacial en la evaluación de la susceptibilidad por movimientos en masa. Para ello, se utilizan cuencas hidrográficas como unidades de mapeo, una elección que permite ajustarse de manera más adecuada a las condiciones naturales del territorio. Además, en lugar de emplear una variable dependiente binaria, se utilizó la frecuencia de movimientos en masa, lo cual proporciona una mejor comprensión del fenómeno y permite explorar la variabilidad espacial de la influencia de las variables predictoras en el área de estudio. El objetivo principal de este trabajo no se encuentra tanto en la producción de un mapa de susceptibilidad, sino en evaluar la heterogeneidad espacial de las covariables y su influencia en la evaluación de la susceptibilidad por movimientos en masa. Este enfoque pretende aportar en la comprensión de la heterogeneidad espacial de los factores condicionantes, lo cual es crucial para la implementación de modelos de susceptibilidad que logren capturar de forma adecuada el fenómeno en cuestión.

## 2. Modelo de regresión espacial ponderada

Los Modelos de Regresión Espacial Ponderados (GWR, por sus siglas en inglés) constituyen una herramienta estadística avanzada desarrollada para analizar la heterogeneidad espacial de relaciones entre una variable respuesta y un conjunto de variables predictoras en un contexto geográfico (Brunsdon, Fotheringham, and Martin E Charlton 1996; Fotheringham, Brunsdon, and Martin Charlton

2000; Fotheringham, Brunsdon, and Charlton 2009). A diferencia de los modelos de regresión tradicional, que asumen que la relación entre las variables es constante en todo el espacio de estudio, los modelos GWR permiten que los coeficientes varíen de manera local. Esto se logra ajustando un modelo de regresión independiente en cada ubicación geográfica, utilizando datos cercanos para ponderar las estimaciones (Brunsdon, Fotheringham, and Martin E Charlton 1996). Esta metodología es particularmente valiosa para el análisis de susceptibilidad a movimientos en masa, ya que permite representar cómo las diferentes condiciones del terreno afectan la susceptibilidad de manera no uniforme a lo largo de un área.

La formulación matemática de un modelo de regresión espacial ponderada es de la siguiente forma (Brunsdon, Fotheringham, and Martin E Charlton 1996):

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i)x_{ik} + \epsilon_i$$

donde:  $y_i$  es el valor de la variable respuesta en la ubicación  $i$ ,  $(u_i, v_i)$  las coordenadas geográficas de la ubicación  $i$ ,  $\beta_0(u_i, v_i)$  el término independiente local en la ubicación  $i$ ,  $\beta_k(u_i, v_i)$  el coeficiente local asociado con la variable predictora  $x_{ik}$  en la ubicación  $i$ .  $x_{ik}$  es el valor de la variable predictora  $k$  para la ubicación  $i$ , y  $\epsilon_i$  el término de error en la ubicación  $i$ .

Los modelos GWR son útiles en situaciones donde la influencia de las variables predictoras cambia espacialmente (Fotheringham, Brunsdon, and Charlton 2009), es decir. En estudios de movimientos en masa, estas relaciones pueden depender de la topografía local, las características del suelo, la vegetación, o incluso de los patrones de precipitación que varían en el espacio. Por ejemplo, la influencia de la pendiente o la cobertura vegetal sobre la susceptibilidad a movimientos en masa puede ser distinta en regiones con diferentes características geológicas o climáticas. La aplicación de GWR permite capturar esta variabilidad y proporciona un análisis más detallado y específico para cada localidad del área de estudio.

El aspecto clave de la GWR es que los coeficientes varían dependiendo de la ubicación geográfica , lo cual se logra a través del uso de un kernel de ponderación que da más peso a las observaciones cercanas y menos peso a las observaciones lejanas (Fotheringham, Brunsdon, and Martin Charlton 2000). En la práctica, este kernel suele ser una función gaussiana o una función adaptativa, que ajusta el peso según la distancia desde el punto central.

Una de las principales ventajas de los modelos GWR es la facilidad con la que los resultados de GWR pueden ser interpretados y visualizados en un contexto espacial, facilitando la identificación de patrones locales de susceptibilidad y ayudando a la toma de decisiones. Los mapas de coeficientes permiten identificar áreas donde ciertas variables tienen una mayor o menor influencia, aportando información útil para el manejo territorial y la planificación.

Sin embargo, los modelos GWR también presentan limitaciones importantes. Una de ellas es la multicolinealidad local, ya que, debido a la naturaleza local de las regresiones, es posible que en algunas áreas las variables predictoras estén altamente correlacionadas, lo que dificulta la interpretación de los coeficientes (Fotheringham, Brunsdon, and Martin Charlton 2000; Brunsdon, Fotheringham, and Martin E Charlton 1996). Además, los modelos GWR suelen ser computacionalmente más exigentes que los modelos de regresión globales o clásicos, especialmente cuando se trabaja con grandes volúmenes de datos espaciales.

Otra limitación a destacar es la selección del ancho de banda (BW) del kernel de ponderación (Guo, Ma, and Zhang 2008). El BW determina cuántas observaciones se consideran en el cálculo de los coeficientes locales, y su elección tiene un impacto en los resultados del modelo. Un BW demasiado

amplio puede resultar en un modelo más similar a una regresión global, mientras que un BW muy estrecho puede llevar a resultados sobre-ajustados y poco generalizables. La selección adecuada del BW suele requerir un proceso iterativo y cuidadoso.

### 3. Área de estudio y datos

El área de estudio se localiza al norte del paralelo 5°N y tiene una extensión aproximada de 50,000 km<sup>2</sup> (Fig. 1). Incluye las Cordilleras Occidental y Central de los Andes colombianos, las cuales se encuentran separadas por el cañón del río Cauca, una prominente depresión geográfica. Los límites naturales del área de estudio están definidos por el río Magdalena al este y el río Atrato al oeste. Dentro de este territorio se han delimitado 533 microcuencas. Aproximadamente el 73% de estas microcuencas tienen una extensión menor a 100 km<sup>2</sup>, con un área promedio de 48 km<sup>2</sup>. Esta configuración hidrográfica refleja la compleja interacción entre procesos geomorfológicos y las variadas condiciones topográficas y climáticas que caracterizan a los Andes colombianos.

Las variables predictoras seleccionadas para el área de estudio comprenden distintos aspectos geomorfológicos, climáticos y geológicos. Estas incluyen el área bajo la curva hipsométrica de cada microcuenca, elevación media, relieve local medio y precipitación media anual. Estas variables fueron seleccionadas en base a su relevancia para los procesos de inestabilidad de laderas, proporcionando un conjunto integral para la modelación de la susceptibilidad a deslizamientos.

Este estudio integra un exhaustivo inventario de 13,777 deslizamientos de tierra, incluyendo eventos superficiales y profundos, ocurridos entre 1970 y 2023 (Fig.1). Estos deslizamientos fueron identificados mediante la detección visual de imágenes ópticas de alta resolución (menor a 1 m) en color real, obtenidas de Google Earth™, lo que garantiza una precisa localización y clasificación de los eventos.

Los parámetros topográficos fueron derivados utilizando un Modelo Digital de Elevación (DEM) generado a partir de datos del radar de apertura sintética (SAR) en banda L del satélite ALOS-PALSAR, con una resolución espacial de 12.5 m por píxel Logan and Karr 2014. Estos parámetros permitieron caracterizar con gran detalle la morfometría del terreno, incluyendo la elevación, pendiente y relieve local, factores críticos para la evaluación de la susceptibilidad a deslizamientos.

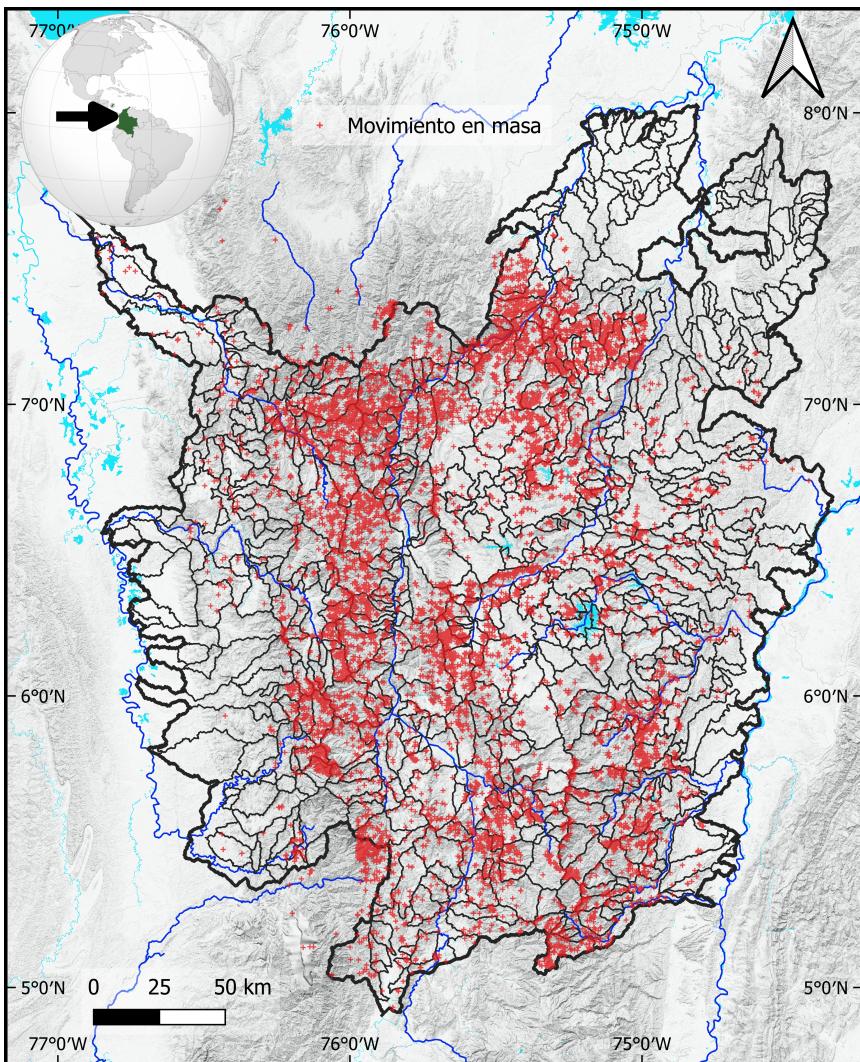
Los datos de precipitación fueron derivados a partir de los datos proporcionados por CHIRPS (extitClimate Hazard Group InfraRed Precipitation with Station Data), versión 2.0, que presenta una resolución espacial de 5 km y cubre el periodo entre 1981 y 2023 Funk, Peterson, and Landsfeld 2015. Estos datos fueron procesados mediante la plataforma Google Earth Engine (GEE), empleando el lenguaje Javascript para la automatización y replicabilidad de los cálculos.

Para la implementacion de los modelos se utilizó la librería mgwr en Python desarrollada por Oshan et al. 2019.

### 4. Resultados

En la Tabla 3 presentamos los resultados del modelo de regresión global aplicado para evaluar el impacto de las variables predictoras clave en la susceptibilidad a movimientos en masa. El intercepto y las variables predictoras, elevación y relieve, son estadísticamente significativas, como lo indican sus valores p, mientras que Lluvia no muestra una influencia significativa al nivel de significancia dado (0.05).

Inicialmente para cada variable se implementó un modelo GWR univariado utilizando como predictor cada variable y utilizando diferentes BW en términos de distancia. La Figura 2 presenta la variación espacial de los coeficientes locales para la variable de elevación obtenidos bajo tres diferentes valores de BW: 10, 20 y 200 km. En este caso, el número de observaciones (en este caso cuencas)



**Figure 1.** Mapa de localización del área de estudio en las Cordilleras Occidental y Central de los Andes colombianos. Los puntos rojos representan los movimientos en masa registrados entre 1970 y 2023. Se muestran los principales ríos y la delimitación de las 533 subcuencas.

con los cuales se modela y estima los coeficientes en cada punto son variables, ya que corresponden al número de cuencas que se encuentran dentro del rango de distancia establecido. A medida que aumenta el ancho de banda, se observa que la influencia de la elevación en la susceptibilidad a movimientos en masa presenta transiciones más suaves, lo que indica una menor variabilidad local. Esto sugiere que un mayor ancho de banda captura tendencias espaciales más amplias, tiendiendo a promediar los efectos localizados, mientras que un menor ancho de banda permite capturar una mayor heterogeneidad espacial y relaciones más localizadas. En particular, la alta variabilidad de los coeficientes estimados con el ancho de banda más pequeño (10 km) implica una significativa influencia espacial local, la cual se atenúa al incrementar el ancho de banda.

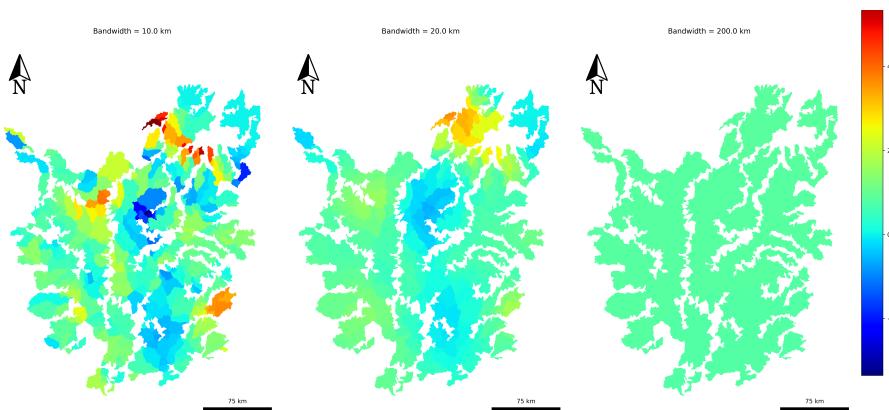
La Figura 3 muestra la variación espacial de los coeficientes locales para la variable de relieve local de la cuenca. Los valores extremos de los coeficientes estimados para el BW=10 km sugieren que en

158  
159  
160  
161  
162  
163  
164  
165  
166

167  
168

**Table 1.** Resumen de los resultados del modelo de regresión global, mostrando los coeficientes estimados (Est.), errores estándar (SE) y valores p para cada variable predictora. Además, se presentan los diagnósticos del modelo —incluyendo la verosimilitud logarítmica, el AIC y el  $R^2$  ajustado para evaluar el ajuste del modelo.

Variable	Est.	SE	p-value
Intercepto	1.84	0.054	0.000
Elevación	0.537	0.079	0.000
Relieve	0.524	0.065	0.000
Lluvia	-0.047	0.069	0.493
Log-likelihood	-858.936		
AIC	1725.87		
Adj. R2	0.371		



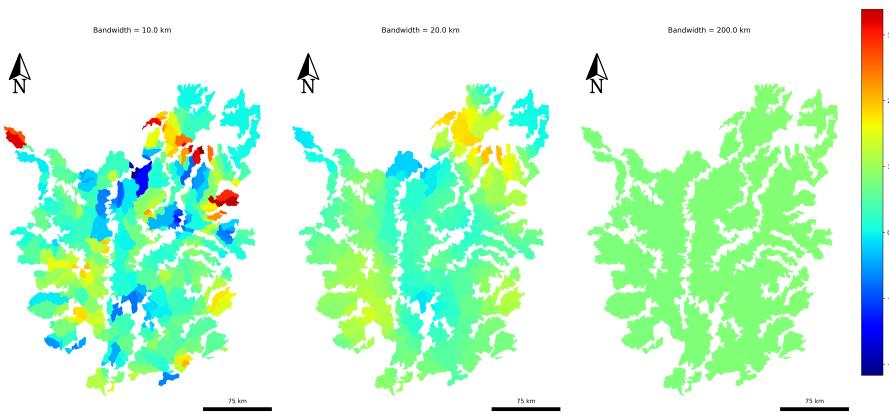
**Figure 2.** Distribución espacial de los coeficientes locales para la variable de elevación obtenida a partir del modelo GWR con tres diferentes anchos de banda (BW): 10 km, 20 km y 200 km. Los gradientes de color representan la magnitud y la dirección de la influencia de la elevación en la susceptibilidad a deslizamientos, donde los colores más fríos indican una influencia menor o negativa y los colores más cálidos indican una influencia positiva más fuerte.

certas áreas el relieve tiene una influencia positiva o negativa mucho más intensa que en otras, lo que se pierde a medida que se amplía la escala del análisis. Esto se refleja en la homogeneización del patrón al utilizar un BW=200 km, donde se observa una suavización considerable de los efectos del relieve.

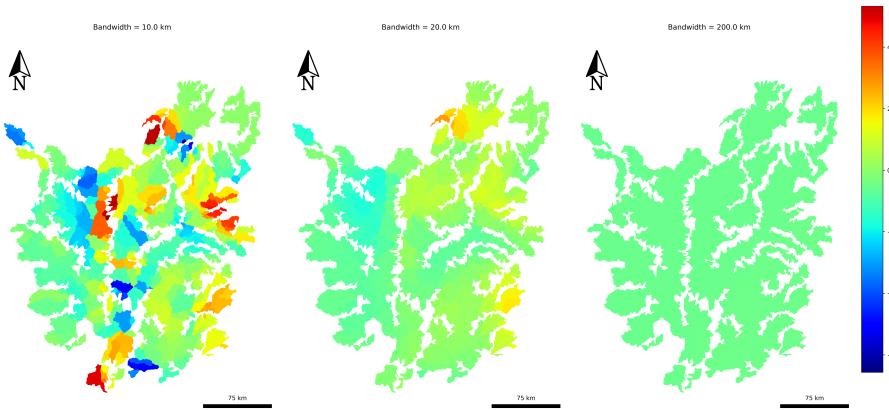
La Figura 4 presenta la variabilidad espacial de los coeficientes estimados para la variable de lluvia media anual. En el BW=10 km, se observa una heterogeneidad considerable en los coeficientes locales, lo que sugiere que la influencia de la precipitación varía notablemente a pequeña escala dentro de la cuenca. Se identifican áreas con coeficientes significativamente positivos (colores cálidos) y negativos (colores fríos), lo que indica que el impacto de la lluvia puede intensificar o reducir la susceptibilidad a movimientos en masa dependiendo de la ubicación específica. Sin embargo, a medida que se incrementa el BW a 20 km y posteriormente a 200 km, los coeficientes se suavizan, perdiéndose las diferencias locales y prevaleciendo un patrón más homogéneo que representa el efecto promedio de la lluvia a gran escala.

La Tabla 3 presenta los resultados diagnósticos de los modelos GWR univariados implementados utilizando como criterio del BW la distancia (10, 20, 200 km). En términos de Log-Likelihood, los valores más altos indican un mejor ajuste del modelo. Se observa que a menor ancho de banda (10 km), los valores de Log-Likelihood tienden a ser más altos, lo cual indica que los modelos son capaces

169  
170  
171  
172173  
174  
175  
176  
177  
178  
179  
180  
181182  
183  
184  
185



**Figure 3.** Distribución espacial de los coeficientes locales para la variable de relieve obtenida mediante el modelo GWR con tres BW distintos: 10 km, 20 km y 200 km.



**Figure 4.** Distribución espacial de los coeficientes locales para la variable de lluvia media anual estimados mediante el modelo GWR para tres diferentes BW: 10 km, 20 km y 200 km.

de capturar mejor las variaciones locales en la relación entre las variables predictoras y la respuesta. Sin embargo, conforme aumenta el ancho de banda a 200 km, se produce una reducción significativa en el valor de Log-Likelihood, lo cual refleja una pérdida de precisión en la captura de variaciones locales. El AIC (Criterio de Información de Akaike) sigue una tendencia similar, mostrando menores valores para anchos de banda más pequeños. Esto sugiere que, a pesar de la mayor complejidad del modelo al considerar las variaciones locales, el ajuste logrado con anchos de banda menores es estadísticamente favorable. En contraste, un ancho de banda de 200 km produce mayores valores de AIC, lo cual indica un modelo menos eficiente para describir las variaciones espaciales. En cuanto al  $R^2$  ajustado, se observa que los valores más altos se encuentran con el ancho de banda de 10 km, especialmente para la elevación (0.646). Esto sugiere que, con anchos de banda más pequeños, el modelo es capaz de explicar mejor la variabilidad en la susceptibilidad a deslizamientos. A medida que el ancho de banda aumenta, el valor de  $R^2$  ajustado disminuye significativamente, lo cual refleja una menor capacidad del modelo para capturar las dependencias espaciales precisas y específicas.

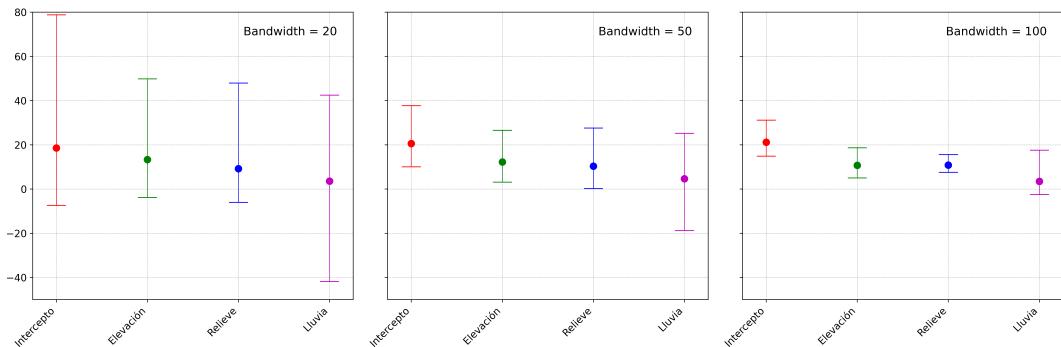
La Figura 5 muestra los resultados de los modelos de GWR multivariado, donde se incluyeron las tres variables predictoras (elevación, relieve y lluvia media anual) y como Bw se utilizó el número

186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200

**Table 2.** Comparación de resultados diagnósticos para modelos de Regresión Geográficamente Ponderada (GWR). La tabla muestra el Log-Likelihood, AIC y R2 ajustado para cada una de las variables predictoras (elevación, relieve y lluvia media anual) a diferentes anchos de banda (10, 20 y 200 km).

Variable	BW (km)	Log-Likelihood	AIC	Adjusted R2
Elevación	10	-638.548	1525.507	0.646
Relieve	10	-639.657	1556.960	0.632
Lluvia	10	-655.705	1547.364	0.628
Elevación	20	-762.526	1608.335	0.530
Relieve	20	-774.700	1639.763	0.504
Lluvia	20	-769.946	1618.306	0.519
Elevación	200	-888.531	1784.140	0.297
Relieve	200	-892.630	1792.289	0.286
Lluvia	200	-942.725	1892.442	0.136

de cuencas, en lugar de una distancia. En este caso el número de cuencas, con los cuales se modela cada observación, es constante de acuerdo con el número seleccionado. En este caso se seleccionó 20, 50 y 100 cuencas como BW. Por le contrario la distancia es variable, ya que las cuencas vecinas se encuentran a distancias diferentes. El gráfico ilustra cómo los valores de los coeficientes varían dependiendo del BW utilizado. Con BW=20 cuencas, los coeficientes muestran mayor variabilidad, lo cual sugiere que este BW capta mejor la heterogeneidad espacial. En contraste, a medida que el BW aumenta a 50 y 100 cuencas, los coeficientes tienden a estabilizarse, lo cual indica una menor variabilidad espacial y una mayor suavización de los efectos locales. La elevación y el relieve presentan una influencia más estable y positiva en la susceptibilidad a deslizamientos, mientras que la lluvia muestra una mayor variabilidad dependiendo del BW. A menor BW, el modelo es capaz de capturar mejor las relaciones locales entre las variables, mientras que, con mayores BW, las relaciones tienden a promediarse en un mayor número de cuencas, lo cual puede suavizar las variaciones locales.



**Figure 5.** Resultados de los modelos de Regresión Geográficamente Ponderada (GWR) considerando simultáneamente tres variables predictoras: elevación, relieve y lluvia media anual. Las subfiguras representan tres diferentes BW: 20, 50 y 100 cuencas. Los puntos representan los valores medios estimados de los coeficientes y las barras de error indican los valores mínimos y máximos.

La Tabla 3 muestra los estadísticos diagnósticos correspondientes a los modelos GWR multivariados utilizando como criterio el número de cuencas (20, 50 y 100 cuencas vecinas). Se puede observar que a medida que el ancho de banda aumenta, el valor del log-verosimilitud disminuye de -764.793 (para un ancho de banda de 20 cuencas) a -822 (para un ancho de banda de 100 cuencas), lo cual indica que el ajuste del modelo se reduce ligeramente con mayores BW. Este comportamiento también se refleja

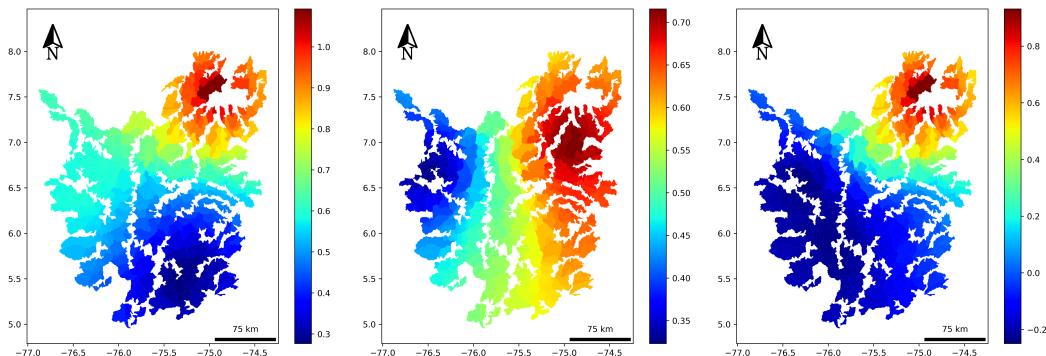
en los valores del Criterio de Información de Akaike (AIC), que incrementan con el aumento del BW, desde 1611 hasta 1668, sugiriendo una menor eficiencia del modelo al suavizar más la variabilidad espacial. El  $R^2$  ajustado también disminuye con el aumento del BW, pasando de 0.527 a 0.444, lo cual indica que la capacidad explicativa del modelo tiende a reducirse cuando se aumenta el área de influencia de cada punto en la ponderación geográfica. Este patrón resalta la importancia de un adecuado balance entre el BW y la captura de la heterogeneidad espacial; un menor BW tiende a capturar mejor los patrones locales, mientras que un mayor BW puede ocultar variaciones relevantes al promediar sobre un área mayor.

**Table 3.** Comparación de resultados diagnósticos para modelos de Regresión Geográficamente Ponderada (GWR) combinando las variables elevación, relieve y lluvia media anual. Los resultados se muestran para tres diferentes BW: 20, 50 y 100 cuencas.

Bandwidth	Log-Likelihood	AIC	Adjusted R2
20	-764.793	1611.273	0.527
50	-799.253	1637.195	0.483
100	-822.270	1668.303	0.444

Para mejorar el modelo, la librería MGWR en Python ofrece una función que selecciona el BW para cada variable que optimiza el modelo. La Figura 6 presenta la variación espacial de los coeficientes estimados para cada una de las tres variables predictoras seleccionadas — elevación, relieve y lluvia media anual — utilizando el BW óptimo recomendado por la librería *mgwr* con 92, 522 y 524 cuencas respectivamente. La figura resalta la capacidad del enfoque GWR para capturar la heterogeneidad espacial de los efectos de cada predictor sobre la susceptibilidad a deslizamientos, mostrando que cada variable tiene un impacto diferente dependiendo de la localización específica dentro del área de estudio. Los BW óptimos utilizados permiten obtener el mejor equilibrio entre el detalle espacial y la estabilidad de las estimaciones, maximizando la precisión local de cada variable. En el primer panel, correspondiente a la elevación (92 cuencas vecinas), se observa una variación considerable en los coeficientes a lo largo del área de estudio. Se identifican zonas con valores positivos significativos (en rojo), lo cual indica una relación positiva entre la elevación y la susceptibilidad, mientras que en otras zonas la relación se debilita o se vuelve negativa. El segundo panel, correspondiente al relieve (522 cuencas vecinas), muestra cómo los coeficientes cambian espacialmente, indicando la importancia del relieve en el área de estudio. Al igual que en la elevación, se puede apreciar una distribución espacial heterogénea, con zonas donde el relieve tiene una mayor influencia en la susceptibilidad a deslizamientos. Por último, el tercer panel corresponde a la lluvia media anual (524 cuencas vecinas), donde se observa que las zonas con mayor intensidad de color (positivo o negativo) reflejan una relación espacialmente diferenciada entre la precipitación y la susceptibilidad. En este caso, se observan áreas en las que la precipitación tiene un mayor efecto en la susceptibilidad, mientras que en otras la influencia es menos significativa.

La Tabla 4 presenta los resultados diagnósticos del modelo de Regresión Geográficamente Ponderada (GWR) que utiliza anchos de banda óptimos para cada variable explicativa. En el caso del intercepto, el coeficiente medio es 2.020, con una desviación estándar de 0.233, indicando variabilidad considerable. Para la variable Elevación, el BW óptimo fue de 92 cuencas, resultando en un coeficiente medio de 0.700, con valores que oscilan entre 0.067 y 1.312. En cuanto al Relieve (522 cuencas vecinas) y Lluvia (524 cuencas vecinas), los coeficientes muestran efectos más reducidos, siendo incluso negativo en el caso de Lluvia, lo que sugiere una relación inversa con la susceptibilidad. Los valores del logaritmo de verosimilitud (-741.056), el AIC (1566.360) y el  $R^2$  ajustado (0.567) indican un ajuste adecuado del modelo, reflejando la importancia de incorporar BW diferenciados para capturar mejor la heterogeneidad espacial.



**Figure 6.** Variación espacial de los coeficientes estimados para las variables elevación, relieve y lluvia media anual utilizando el ancho de banda óptimo (44, 524 y 129 cuencas vecinas respectivamente) en los modelos de Regresión Geográficamente Ponderada (GWR).

**Table 4.** Resultados diagnósticos del modelo de Regresión Geográficamente Ponderada (GWR) utilizando BW óptimos para cada variable. Se presentan los coeficientes medios, mínimos y máximos, así como la desviación estándar de los coeficientes para el intercepción y las covariaciones elevación, relieve y lluvia.

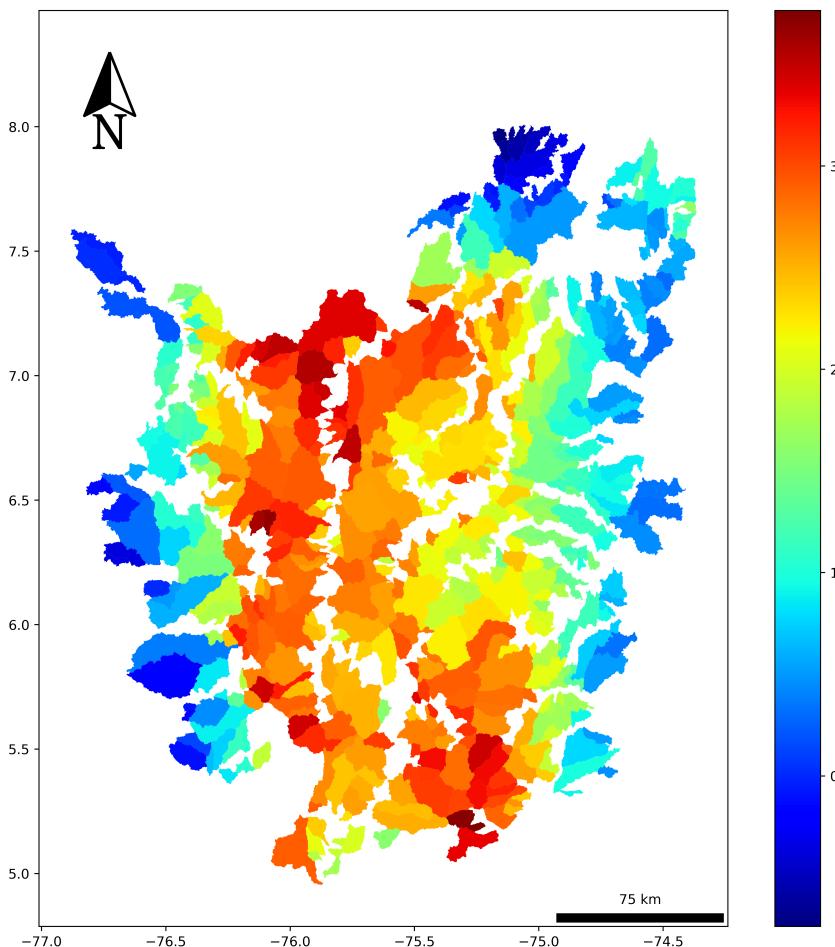
Variables	BW (cuencas))	Coeficiente medio	Coef. Min.	Coef. Max.	STD
Intercepción	44	2.020	0.896	4.004	0.233
Elevación	92	0.700	0.067	1.312	0.187
Relieve	522	0.279	0.213	0.340	0.084
Lluvia	524	-0.030	-0.048	0.001	0.089
Log-Likelihood	-741.056				
AIC	1566.360				
Adjusted $R^2$	0.567				

La Figura 7 presenta la distribución espacial de los valores estimados por el modelo GWR con el BW óptimo para la frecuencia de deslizamientos por cuenca. Cada área en el mapa representa la estimación de la frecuencia de deslizamientos, donde se observa una clara variabilidad espacial de la susceptibilidad a deslizamientos dentro del área de estudio. Los colores indican la magnitud de la densidad estimada, con tonos cálidos (rojos) representando áreas con mayores densidades de deslizamientos y tonos fríos (azules) indicando menores densidades.

## 5. Discusión

La heterogeneidad espacial constituye un reto esencial al modelar la susceptibilidad a movimientos en masa debido a la gran variabilidad de las condiciones geológicas, topográficas y climáticas en el área de estudio. Esta variabilidad influye en los mecanismos de inestabilidad y, en consecuencia, requiere la incorporación de modelos que puedan capturar esta diversidad de manera precisa. Los modelos multiniveles han demostrado ser una herramienta útil para abordar esta heterogeneidad, ya que permiten definir distintas escalas de análisis que reflejan la variabilidad en los coeficientes de los predictores (Lee and Nelder 1996). Una de las debilidades de los modelos multiniveles es la necesidad de definir previamente las regiones geográficas para las cuales se espera que los coeficientes varíen, lo cual puede limitar la capacidad del modelo para representar la heterogeneidad espacial de manera precisa, si estas regiones no se definen adecuadamente.

Sin embargo, la definición de estas regiones implica la necesidad de tener un conocimiento a priori



**Figure 7.** Distribución espacial de los valores estimados por el modelo de Regresión Geográficamente Ponderada (GWR) para la frecuencia de deslizamientos por cuenca, utilizando el ancho de banda óptimo. Los colores reflejan la densidad estimada, donde los tonos cálidos indican una mayor susceptibilidad a los deslizamientos y los tonos fríos una menor.

sobre la heterogeneidad de las variables para determinar qué áreas deberían ser modeladas de manera independiente. Esto resulta particularmente desafiantre en regiones donde la información detallada es limitada o la región es compleja. En contraste, la Regresión Geográficamente Ponderada (GWR) ofrece una ventaja significativa al no requerir la división previa del área de estudio en regiones fijas. En lugar de esto, GWR permite que los coeficientes varíen localmente, estimándolos con base en un ancho de banda (BW) que puede ser definido en términos del número de unidades de mapeo vecinas o un área fija (Brunsdon, Fotheringham, and Martin E Charlton 1996). Esta flexibilidad proporciona una representación más precisa de la heterogeneidad espacial, sin las limitaciones impuestas por la necesidad de segmentación previa.

Los modelos GWR estiman una superficie suavizada que refleja cómo varía la influencia de cada parámetro sobre el área de estudio (Fotheringham, Brunsdon, and Charlton 2009). Estas superficies permiten identificar las zonas en las que las variables predictoras tienen una mayor o menor influencia sobre la susceptibilidad a los movimientos en masa, lo cual resulta esencial para la interpretación de los patrones espaciales de la amenaza. La parametrización de esta superficie suavizada depende

275  
276  
277  
278  
279  
280  
281  
282  
283

284  
285  
286  
287  
288

del tipo de kernel elegido, que determina cómo se ponderan las observaciones vecinas en cada modelo local. La elección del kernel adecuado es crucial para capturar de manera correcta la variabilidad espacial de los coeficientes (Guo, Ma, and Zhang 2008). Los kernels pueden ser adaptativos o fijos, y cada uno presenta ventajas y desventajas según el contexto de aplicación. Los kernels adaptativos permiten ajustar el número de vecinos a incluir en la estimación de cada punto, lo cual es útil para regiones con una distribución irregular de los datos. En cambio, los kernels fijos utilizan una distancia constante, lo cual puede ser más apropiado en regiones con una distribución de datos homogénea. Nuestros resultados corroboran que la elección adecuada del kernel afecta la suavización de la superficie resultante y, por ende, la capacidad del modelo para capturar patrones locales sin perder la generalización. Esto significa que se requiere del criterio y conocimiento de la zona para establecer y estimar la superficie óptima.

En el presente estudio se destaca cómo la incorporación de la variabilidad local en los modelos GWR mejora de manera significativa el ajuste del modelo. Esto se evidencia en el incremento del  $R^2$  ajustado y la mejora de otras métricas, tales como el AIC y el log-likelihood. Al permitir que los coeficientes varíen espacialmente, GWR capta mejor las relaciones entre las variables predictoras y la susceptibilidad a deslizamientos, lo cual es particularmente relevante en un contexto geográficamente heterogéneo como el de las zonas montañosas. En este estudio, se observa que el modelo global explica solo el 37% de la varianza del problema, mientras que al incluir la heterogeneidad espacial a través de los modelos GWR, se logra incrementar esta capacidad hasta un 57%. Este aumento significativo demuestra la importancia de considerar la variabilidad espacial para mejorar la precisión y el ajuste de los modelos de susceptibilidad a deslizamientos.

No obstante, la reducción excesiva del BW puede llevar a una disminución en el tamaño efectivo de los datos que se utilizan para estimar los coeficientes en cada punto. Por ejemplo, en este estudio se observó que, al reducir el BW a valores muy bajos, se produjo una disminución significativa en la estabilidad del modelo, lo que puso en evidencia el delicado balance entre la captura de variabilidad local y la estabilidad general del modelo. Cuando se reduce el BW, se considera un menor número de vecinos, lo cual aumenta el riesgo de sobreajuste del modelo. En este estudio, se identificó que una reducción excesiva del BW llevó a una disminución en la estabilidad del modelo, lo cual evidencia la importancia de establecer un umbral adecuado para minimizar el sobreajuste. El sobreajuste ocurre cuando el modelo se adapta de manera muy precisa a los datos locales, lo que reduce su capacidad para generalizar los resultados a nuevas áreas o condiciones. Esto también puede incrementar la varianza del modelo, haciéndolo más inestable. Por lo tanto, encontrar un equilibrio adecuado entre la captura de las variaciones locales y la estabilidad del modelo es un reto clave al implementar GWR.

Los modelos GWR no solo representan una herramienta poderosa para modelar la susceptibilidad a deslizamientos, sino también sirven como un análisis exploratorio para comprender la escala de variación espacial de cada variable antes de aplicar modelos más complejos, como los modelos jerárquicos o multiniveles.

Una extensión de GWR que resulta particularmente útil en este contexto es la Regresión Geográficamente Ponderada Multibanda (MGWR), la cual permite una optimización del ancho de banda para cada variable predictor (Fotheringham, Yang, and Kang 2017). Esto significa que se puede determinar la escala óptima de variación de cada predictor de manera independiente, proporcionando una visión más detallada y precisa sobre la escala de influencias de las diferentes variables en el área de estudio. Esta información puede ser utilizada para definir mejor los componentes de los modelos jerárquicos que se pueden aplicar posteriormente.

Aunque el enfoque tradicional de GWR se basa en un modelo lineal con distribución Gaussiana, existen esfuerzos por extender su aplicación a otros tipos de funciones de enlace, lo que podría expandir su uso en diferentes contextos de modelado. En particular, la extensión de GWR para

utilizar una función de enlace logística (Nkeki and Asikhia 2019) resulta de gran interés para modelar variables de respuesta binarias, como la ocurrencia o no de deslizamientos en ciertas zonas. Sin embargo, estos enfoques presentan ciertos desafíos. La aplicación de GWR con funciones de enlace no gaussianas a menudo enfrenta problemas de inestabilidad, particularmente debido a la naturaleza local de los coeficientes y la posible colinealidad de los predictores en ciertas áreas del espacio. Estos factores pueden dificultar la convergencia del modelo y producir resultados poco confiables si no se realiza una correcta selección de los parámetros y del BW. Por ello, es necesario realizar un análisis cuidadoso y considerar estos posibles problemas antes de aplicar GWR en contextos no lineales.

Pese a sus ventajas, GWR también presenta una serie de limitaciones que deben ser consideradas en el contexto de modelado de susceptibilidad a movimientos en masa. En primer lugar, la complejidad computacional asociada con la estimación local de los coeficientes puede ser considerablemente alta, especialmente en áreas grandes con muchas observaciones. A medida que el tamaño de los datos aumenta, el costo computacional también lo hace, lo cual limita la aplicabilidad de GWR en estudios de gran escala sin un acceso adecuado a recursos de cálculo. En este estudio, se observó que los requerimientos computacionales incrementaron significativamente al aplicar GWR, particularmente al manejar datos de alta resolución espacial y múltiples variables predictoras, lo cual resalta la necesidad de contar con recursos computacionales robustos para asegurar la eficiencia del modelado. Otro problema importante es la colinealidad local. A diferencia de los modelos globales, donde se puede evaluar la colinealidad de las variables de manera única para toda el área de estudio, en GWR la colinealidad puede variar en diferentes ubicaciones. Esto significa que una variable que no presenta colinealidad a escala global puede mostrar alta correlación con otras variables en ciertas áreas específicas. Esta colinealidad local afecta la interpretación de los resultados y puede llevar a estimaciones sesgadas o inconsistentes. Asimismo, la interpretación de los coeficientes suavizados en GWR suele ser más compleja en comparación con un modelo global. Mientras que los modelos globales ofrecen un único coeficiente que puede interpretarse como la relación promedio entre una variable predictor y la respuesta, los coeficientes locales en GWR reflejan esta relación de manera específica para cada ubicación. Aunque esto permite una interpretación más detallada, también aumenta la dificultad de generalizar los resultados a otras áreas o situaciones.

Finalmente, la posibilidad de sobreajuste es una desventaja importante en GWR, particularmente en áreas donde los datos son escasos. Cuando el modelo se ajusta demasiado a los datos locales, los resultados pierden su capacidad de ser aplicables en otros contextos. Esto hace que los modelos GWR sean sensibles a la distribución espacial de los datos y puede llevar a predicciones poco confiables en áreas donde las observaciones son limitadas o ausentes.

Esta heterogeneidad se puede dar a diferentes escalas de acuerdo con la variable en estudio. A nivel local, algunas variables pueden mostrar una variabilidad significativa que se relaciona estrechamente con las características del terreno, como la geomorfología y los procesos en áreas de pequeña escala. Por otro lado, otras variables pueden presentar patrones de variación más consistentes a un nivel regional. En nuestro análisis, los resultados indican que la variabilidad espacial de ciertas variables tiende a disminuir cuando se incrementa el ancho de banda a alrededor de 100 km o más. Este comportamiento sugiere que, a mayores escalas, la variabilidad se homogeniza, lo cual destaca la importancia de la variabilidad a nivel local y la fuerte influencia de las condiciones particulares de cada cuenca y de los terrenos en los Andes colombianos.

## 6. Conclusiones

Este estudio ha demostrado la importancia de incorporar la heterogeneidad espacial al modelar la susceptibilidad a movimientos en masa en regiones complejas, como los Andes colombianos. A través de la aplicación de la Regresión Geográficamente Ponderada (GWR), se ha evidenciado una

mejora significativa en la capacidad de los modelos para capturar la variabilidad local y mejorar la precisión de las predicciones. Estos enfoques permitieron identificar patrones espaciales diferenciados y escalas de influencia específicas para cada predictor, proporcionando una representación más precisa de la susceptibilidad a movimientos en masa y destacando la importancia de la adaptación de los modelos al contexto geográfico particular de cada región.

Uno de los principales hallazgos de este estudio es que la variabilidad espacial puede ser mejor representada cuando se permite que los coeficientes de los predictores varíen localmente, lo cual es particularmente relevante en áreas con alta complejidad geológica, climática y topográfica, como los Andes Colombianos. Los resultados mostraron que los modelos GWR superan considerablemente a los modelos globales tradicionales, con incrementos significativos en la varianza explicada y en las métricas de ajuste del modelo. Sin embargo, también se subraya la necesidad de un equilibrio adecuado al seleccionar el ancho de banda, para evitar problemas de sobreajuste y garantizar la estabilidad del modelo.

A pesar de las ventajas que ofrecen GWR también se identificaron limitaciones, como la alta complejidad computacional y la presencia de colinealidad local, que pueden afectar la interpretación de los coeficientes y la fiabilidad de los resultados. Para abordar estas limitaciones, es fundamental utilizar estrategias como la detección de colinealidad y la regularización, así como contar con recursos computacionales robustos.

En conclusión, la aplicación de modelos espaciales avanzados como GWR representa un avance significativo en la modelización de la susceptibilidad a movimientos en masa, proporcionando una herramienta poderosa para capturar la variabilidad espacial y mejorar la comprensión de los patrones de susceptibilidad, particularmente en entornos montañosos y complejos.

## 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/ModelosMGWR>).

## References

- Anselin, Luc (1990). "Spatial dependence and spatial structural instability in applied regression analysis". In: *Journal of Regional science* 30.2, pp. 185–207.
- 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.
- Aristizábal, E, E García, and C Martínez (2015). "Susceptibility assessment of shallow landslides triggered by rainfall in tropical basins and mountainous terrains". In: *Natural Hazards* 78.1, pp. 621–634.
- Brabb, Earl E (1984). "Innovative approaches to landslide hazard and risk mapping." In:
- Brunsdon, Chris, A Stewart Fotheringham, and Martin E Charlton (1996). "Geographically weighted regression: a method for exploring spatial nonstationarity". In: *Geographical analysis* 28.4, pp. 281–298.
- 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.
- Cressie, Noel (2015). *Statistics for spatial data*. John Wiley & Sons.

- 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. 425  
426
- Fotheringham, A Stewart, Chris Brunsdon, and Martin Charlton (2000). *Quantitative geography: perspectives on spatial data analysis*. Sage. 427  
428
- Fotheringham, A Stewart, Chris Brunsdon, and ME Charlton (2009). "Geographically weighted regression". In: *The Sage handbook of spatial analysis* 1, pp. 243–254. 429  
430
- Fotheringham, A Stewart, Wenbai Yang, and Wei Kang (2017). "Multiscale geographically weighted regression (MGWR)". In: *Annals of the American Association of Geographers* 107.6, pp. 1247–1265. 431  
432
- 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. 433  
434
- Guo, Luo, Zhihai Ma, and Lianjun Zhang (2008). "Comparison of bandwidth selection in application of geographically weighted regression: a case study". In: *Canadian Journal of Forest Research* 38.9, pp. 2526–2534. 435  
436
- Guzzetti, Fausto et al. (1999). "Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy". In: *Geomorphology* 31.1-4, pp. 181–216. 437  
438
- Korup, Oliver and Amelie Stolle (2014). "Landslide prediction from machine learning". In: *Geology today* 30.1, pp. 26–33. 439  
440
- 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. 441  
442
- LeSage, James and Robert Kelley Pace (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC. 443  
444
- 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. 445  
446
- Lombardo, Luigi et al. (2020). "Space-time landslide predictive modelling". In: *Earth-Science Reviews* 209, p. 103318. 447  
448
- Nkeki, Felix Ndidi and Monday Ohi Asikhia (2019). "Geographically weighted logistic regression approach to explore the spatial variability in travel behaviour and built environment interactions: Accounting simultaneously for demographic and socioeconomic characteristics". In: *Applied geography* 108, pp. 47–63. 449  
450
- Oshan, Taylor M et al. (2019). "mgwr: A Python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale". In: *ISPRS International Journal of Geo-Information* 8.6, p. 269. 451  
452
- Rey, Sergio, Dani Arribas-Bel, and Levi John Wolf (2023). *Geographic data science with python*. Chapman and Hall/CRC. 453  
454
- Soeters, Robert and CJ Van Westen (1996). "Slope instability recognition, analysis and zonation". In: *Landslides: investigation and mitigation* 247, pp. 129–177. 455  
456
- Tehrani, Faraz S et al. (2022). "Machine learning and landslide studies: recent advances and applications". In: *Natural Hazards* 114.2, pp. 1197–1245. 457  
458
- Wong, George Y and William M Mason (1985). "The hierarchical logistic regression model for multilevel analysis". In: *Journal of the American Statistical Association* 80.391, pp. 513–524. 459  
460