

UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Ingeniería



Desarrollo de mapas de predicción de riesgo a deslizamientos del departamento de San Marcos de Guatemala, empleando sistemas de información geográfica (GIS) y modelos de aprendizaje automático

Trabajo de graduación en modalidad de tesis presentado por Ana Alicia Keilhauer Varona para optar al grado académico de Licenciado en Ingeniería Civil Arquitectónica

Guatemala,
2025

UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Ingeniería




Desarrollo de mapas de predicción de riesgo a deslizamientos del departamento de San Marcos de Guatemala, empleando sistemas de información geográfica (GIS) y modelos de aprendizaje automático


Trabajo de graduación en modalidad de tesis presentado por Ana Alicia Keilhauer Varona para optar al grado académico de Licenciado en Ingeniería Civil Arquitectónica

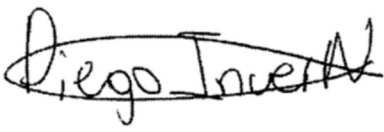
Guatemala,
2025

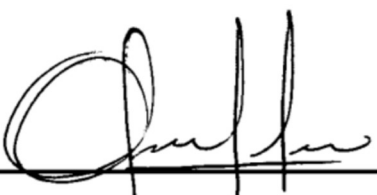
Vo. Bo.:

(f) 
Ing. Danilo Rodríguez

Tribunal ex.:

(f) 
Ing. Danilo Rodríguez

(f) 
Ing. Diego Incer

(f) 
Ing. Otoniel Echeverría

Fecha de examinación: 17 de junio, 2025.

ÍNDICE GENERAL

ÍNDICE DE FIGURAS	XI
ÍNDICE DE TABLAS	XIII
RESUMEN	XV
I. INTRODUCCIÓN	1
II. ANTECEDENTES	3
III. JUSTIFICACIÓN	5
IV. OBJETIVOS	7
A. Objetivo General	7
B. Objetivos Específicos.....	7
V. MARCO TEÓRICO.....	9
A. Deslizamientos	9
1. Concepto.....	9
2. Categorías.....	11
3. Partes de un deslizamiento	16
4. Estados de un deslizamiento.....	17
2. Causas principales de deslizamiento	18
3. Metodologías Mora Vahrson de determinación de áreas susceptibles a deslizamientos	21
4. Medidas de mitigación de deslizamientos.....	24
1. Legislación relacionada a deslizamientos o desastres naturales.....	28
B. Características geográficas del área de estudio – San Marcos, Guatemala	29
1. Descripción general.....	29
2. Fisiografía.....	31
3. Características meteorológicas	31
4. Características geológicas	34
C. Sistemas de Información Geográfica (SIG)	35
1. Definición.....	35
2. Usos de los sistemas de información geográfica.....	36

3.	Modelos de datos especiales.....	38
4.	Software de SIG	40
5.	Archivos compatibles con SIG.....	40
6.	Sistemas de Coordenadas de Referencia (SRC).....	41
7.	Aplicaciones directas de los sistemas de información geográfica.....	42
8.	Georreferenciación:	44
9.	Modelo de los datos:.....	44
D.	Evaluación y monitoreo de deslizamientos:	45
1.	Evaluación de riesgos:	45
b.	Impacto socioeconómico:.....	45
2.	Modelos de predicción de deslizamientos:.....	46
3.	Inventarios de deslizamientos:	47
E.	Fundamentos del aprendizaje automático:	48
1.	Machine Learning:	48
2.	Aprendizaje supervisado:	49
3.	Aprendizaje no supervisado:	50
4.	Random Forest:	51
5.	Algoritmo - XGBoost:.....	52
6.	Comparación general de algoritmos basados en árboles:.....	54
7.	Evaluación de modelos de clasificación:.....	55
VI.	METODOLOGÍA.....	59
A.	Delimitación del área de estudio:	59
B.	Recolección y preparación del inventario de deslizamientos:	59
C.	Recolección y preparación de información de datos de entrenamiento:	61
1.	Recolectar información geográfica de Guatemala:	61
D.	Preparación de información en capas de SIG (QGIS)	63
E.	Generación de puntos de no deslizamiento	64
F.	Generación de inventario total.....	68
G.	Extracción de valores para entrenamiento del modelo	68
H.	Construcción de la base de datos completa.....	70
I.	Generación de puntos de estudio.....	70

J.	Generación de modelo – Random Forest:	73
K.	Generación de modelo – XGBoost	75
L.	Presentación de resultados	76
VII.	RESULTADOS Y DISCUSIÓN	77
A.	Presentación de resultados – Mapa de predicción elaborado con Random Forest: ...	77
B.	Presentación de resultados – Mapa de predicción elaborado con XGBoost.	80
C.	Comparación de mapas de susceptibilidad.	82
D.	Importancia de variables predictoras	85
E.	Análisis crítico de los resultados	86
F.	Implicaciones prácticas	87
G.	Comparación con mapas institucionales	89
VIII.	CONCLUSIONES	93
IX.	RECOMENDACIONES	95
X.	BIBLIOGRAFÍA	97
XI.	APÉNDICES	103
A.	Script para obtención de variables predictoras desde Google Earth Engine	103
B.	Código para generación de mapa de susceptibilidad usando Random Forest	107
C.	Código para generación de mapa de susceptibilidad usando XGBoost	110
XII.	GLOSARIO	115

ÍNDICE DE FIGURAS

Figura 1. Deslizamiento rotacional.....	11
Figura 2. Esquema de deslizamiento rotacional.....	11
Figura 3. Deslizamiento traslacional.....	12
Figura 4. Esquema deslizamiento traslacional.....	12
Figura 5. Flujo de lodo.....	13
Figura 6. Esquema flujo de tierra.....	13
Figura 7. Caída de piedras en km. 259 RN12 Sur, La Castalia, San Marcos, Guatemala.....	14
Figura 8. Esquema de desprendimientos o caídas.....	14
Figura 9. Volcamiento típico de suelo pulverulento.....	15
Figura 10. Esquema de volcamientos.....	15
Figura 11. Componentes de un deslizamiento.....	16
Figura 12. Estados de los deslizamientos.....	18
Figura 13. Drenes subhorizontales o californianos.....	25
Figura 14. Cunetas como sistema de drenaje para contención de taludes.....	25
Figura 15. Talud protegido con soil nailing recubierto con geotextil.....	26
Figura 16. Muro de soil nailing recubierto con shotcrete (concreto lanzado).....	26
Figura 17. Muro recubierto con geotextil.....	27
Figura 18. Muro recubierto con concreto lanzado.....	27
Figura 19. Muro de contención de concreto armado.....	28
Figura 20. Mapa de relieve del departamento de San Marcos, Guatemala.....	30
Figura 21. Mapa de regiones climáticas de Guatemala.....	32
Figura 22. Mapa de registro de precipitación anual de Guatemala.....	33
Figura 23. Mapa geotectónico del país de Guatemala.....	34
Figura 24. Localización condicionada usando parámetros de elevación con ArcGIS.....	37
Figura 25. Uso de GIS para análisis en tendencias de deforestación en Brasil.....	37
Figura 26. Análisis de redes y enrutamiento vehicular en GIS.....	38
Figura 27. Esquema de la modelización de datos en SIG.....	45
Figura 28. Diagrama de flujo de aprendizaje supervisado.....	49
Figura 29. Diagrama de flujo de aprendizaje no supervisado.....	50
Figura 30. Matriz de confusión de modelos de aprendizaje automáticos.....	56
Figura 31. Generación de buffer de puntos de deslizamientos.....	64
Figura 32. Generación de corte de limite departamental y buffer.....	65
Figura 33. Generación de puntos aleatorios dentro de polígono con resta de Buffer.....	66
Figura 34. Generación de fechas aleatorias para puntos de no deslizamiento.....	67
Figura 35. Generación de puntos regulares a una distancia de 500m entre cada uno en extensión del terreno de San Marcos.....	71
Figura 36. Visualización de puntos regulares a lo largo de la extensión.....	71
Figura 37. Visualización de puntos regulares cortados a lo largo del polígono delimitante.....	72
Figura 38. Mapa de probabilidad de deslizamientos empleando Random Forest.....	77
Figura 39. Gráfico de ROC-AUC para modelo de Random Forest.....	78

Figura 40. Mapa de predicción de deslizamientos empleando XGBoost.....	80
Figura 41. Gráfico de ROC-AUC para modelo de XGBoost.	81
Figura 42. Histograma de importancias por variable - Modelo Random Forest.	85
Figura 43. Histograma de importancias por variable - Modelo XGBoost.....	85
Figura 44. Mapa de vulnerabilidades principales del departamento de San Marcos, Guatemala	89
Figura 45. Mapa de amenazas de la República de Guatemala (enfocado a San Marcos).	89
Figura 46. Mapa de susceptibilidad a deslizamientos del territorio de Guatemala	90
Figura 47. Mapa de probabilidad de deslizamientos empleando Random Forest.	91
Figura 48. Mapa de predicción de deslizamientos empleando XGBoost.....	91

ÍNDICE DE TABLAS

Tabla 1. Clasificación por litología y valor SL.....	22
Tabla 2. Valores relativos de relieve (Rr), su clasificación y valores resultantes del factor de pendiente (Sr).	22
Tabla 3. Clasificación de valores promedio mensuales de precipitación.	23
Tabla 4. Determinación del factor de intensidad sísmica como detonante para deslizamientos, empleando valores de intensidad de 100 años y la escala modificada de Mercalli.....	23
Tabla 5. Índice de intensidad de precipitación resultante de la clasificación máxima de precipitaciones diarias para un periodo de retorno de 100 años. En caso de datos pluviométricos bajos se considera la columna 2.	23
Tabla 6. Regiones fisiográficas departamento de San Marcos	31
Tabla 7. Ejemplos de Sistemas de Coordenadas de Referencia (SCR).	42
Tabla 8. Software o programas empleados para este trabajo.....	60
Tabla 9. Archivos utilizados para la construcción del inventario de deslizamientos.	61
Tabla 10. Fuentes de información utilizadas en el análisis.....	61
Tabla 11. Atributos en mapas de Series de Suelos de Guatemala.	62
Tabla 12. Atributos taxonómicos y de calidad del suelo de Guatemala	62
Tabla 13. Librerías de Python utilizadas y su función principal para modelo Random Forest. .	74
Tabla 14. Librerías/componentes de Python utilizadas y su función principal para modelo XGBoost.....	75
Tabla 15. Matriz de confusión - Modelo Random Forest.....	78
Tabla 16. Reporte de clasificación - Modelo Random Forest.	78
Tabla 17. Matriz de confusión de modelo XGBoost.	81
Tabla 18. Reporte de clasificación de modelo XGBoost.....	81

RESUMEN

Debido a condiciones geográficas, geomorfología, litológicas y el relieve topográfico del país de Guatemala, existe una alta vulnerabilidad a eventos de deslizamientos y derrumbes de tierra. Los deslizamientos de tierra implican el movimiento de un volumen de suelo sobre una pendiente, y estos eventos tienden a afectar a la población de Guatemala al resultar en daños económicos, fatalidades y daños a infraestructura. Se ha estudiado el efecto de distintos factores como el tipo de suelo, drenaje, precipitación, viento, entre otros para elaborar mapas de predicción de riesgo de deslizamientos, que se emplean como herramientas para análisis de vulnerabilidad socioeconómica, planificación territorial y prevención o gestión de desastres naturales. Se elaboró un programa utilizando sistemas de información geográfica y aprendizaje automático (machine learning) como trabajo de graduación que buscaba encontrar la probabilidad de deslizamientos en el departamento de San Marcos, Guatemala. Tomando en cuenta los factores mencionados previamente, se puede predecir el riesgo de deslizamientos para un área o punto determinado, lo que permite utilizar dicha información para fines de planeamiento territorial y gestión de desastres. Debido a que hay muchos factores influyentes para los deslizamientos de suelo, este mapa de amenaza por derrumbe se desarrolla buscando determinar la importancia o peso para cada parámetro. Empleando como referencia mapas de predicción de deslizamientos o de vulnerabilidad a desastres naturales elaborados por instituciones gubernamentales, se realizó una comparación visual entre los mapas. Esto con el fin de comparar los criterios y métodos de cálculo utilizados. Finalmente, se tiene previsto el uso de dicho mapa para el manejo de desastres naturales y en la planificación territorial, quedando a disposición de instituciones e individuos gubernamentales.

ABSTRACT

Due to the geographic conditions, geomorphology, lithology, and topographic relief of Guatemala, there is a high vulnerability to landslides and earth movements. Landslides involve the movement of a mass of soil down a slope, and these events tend to impact the population of Guatemala by causing economic losses, fatalities, and damage to infrastructure. The effect of various factors such as soil type, drainage, precipitation, wind, among others, has been studied to create landslide risk prediction maps, which are used as tools for socioeconomic vulnerability analysis, land-use planning, and the prevention or management of natural disasters. As part of a graduation project, a program was developed using geographic information systems (GIS) and machine learning to estimate the probability of landslides in the department of San Marcos, Guatemala. By taking into consideration the previously mentioned factors, it is possible to predict landslide risk for a specific area or point, which allows the resulting information to be used for territorial planning and disaster management purposes. Since many different factors influence landslides, this landslide hazard map was developed to determine the relative importance or weight of each parameter. Using landslide prediction or natural disaster vulnerability maps created by government institutions as references, a visual comparison was made between the maps in order to assess the criteria and calculation methods used. Finally, the use of this map is intended to support natural disaster management and land-use planning, making it available to government institutions and individuals.

I. INTRODUCCIÓN

Los deslizamientos de tierra constituyen una de las amenazas geológicas de mayor incidencia en regiones montañosas con altos índices de precipitación, como el departamento de San Marcos, Guatemala. Estos eventos pueden generar impactos sociales, económicos y ambientales severos, comprometiendo viviendas, infraestructura vial y actividades productivas. Su ocurrencia está fuertemente influenciada por factores como la pendiente del terreno, la intensidad de la lluvia, el tipo de suelo y la cobertura vegetal (Fell et al., 2008). En este contexto, los Sistemas de Información Geográfica (SIG) y las técnicas de análisis espacial se han consolidado como herramientas fundamentales para el estudio de la susceptibilidad a deslizamientos, al permitir la integración y análisis de múltiples variables espaciales (Van Westen et al., 2006).

A pesar del creciente interés por aplicar modelos predictivos para mapear zonas susceptibles a deslizamientos, la mayoría de las investigaciones en Guatemala se han desarrollado a escalas locales o con enfoques cualitativos. Esto limita su aplicabilidad para la planificación y gestión del territorio a nivel departamental. Además, existe una limitada incorporación de metodologías basadas en inteligencia artificial y aprendizaje automático (machine learning), las cuales, han demostrado ser eficaces en diversos contextos geográficos para clasificar zonas de riesgo con altos niveles de precisión (Pham et al., 2018; Reichenbach et al., 2018). Esta brecha evidencia la necesidad de propuestas técnicas que aborden el problema desde una perspectiva integral, utilizando herramientas modernas y datos geoespaciales actualizados.

Ante esta situación, el presente estudio tiene como objetivo desarrollar un modelo de predicción de susceptibilidad a deslizamientos en el departamento de San Marcos, Guatemala, utilizando algoritmos de aprendizaje supervisado como Random Forest y XGBoost. El modelo se construye a partir de un inventario de deslizamientos, variables condicionantes como pendiente, altitud, precipitación acumulada, tipo de suelo, uso del suelo y geología, y se apoya en plataformas como Google Earth Engine y Quantum GIS (QGIS) para el procesamiento de datos. Se espera que los resultados generen un insumo técnico confiable para fortalecer los procesos de planificación territorial, prevención de desastres y gestión del riesgo en áreas vulnerables.

II. ANTECEDENTES

En Guatemala, los deslizamientos de tierra representan uno de los fenómenos naturales más frecuentes y devastadores, particularmente durante la temporada de lluvias. La complejidad geológica, la topografía escarpada y los patrones de precipitación intensos convierten al país en un territorio altamente susceptible a estos eventos. En investigaciones previas, como la tesis desarrollada por S.A. Orozco en el año 2022, se abordó este problema mediante la aplicación de modelos de aprendizaje automático para la generación de mapas de susceptibilidad a deslizamientos a nivel nacional. Dicha investigación resaltó la importancia de integrar tanto factores permanentes del terreno (elevación, pendiente, tipo de suelo, litología) como factores desencadenantes como la precipitación acumulada, con el fin de alimentar modelos como Decision Tree, Random Forest, XGBoost, Support Vector Machine y Regresión Logística.

III. JUSTIFICACIÓN

Los derrumbes son eventos con un alto grado de impacto socioeconómico, que provocan la pérdida de patrimonio de personas, empresas e instituciones, así como efectos indirectos en los ingresos individuales, aumentos en el gasto público e inflación. También afectan gravemente la dinámica del comercio local, al interrumpir rutas de acceso, reducir la movilidad de consumidores y limitar la operación de pequeños negocios, lo que debilita las economías familiares y comunitarias. Los mapas de predicción de deslizamientos se generan empleando el procesamiento de datos ingresados, utilizando varios parámetros influyentes para predecir y evaluar el riesgo. Estos mapas proveerán información importante, resaltando áreas y regiones de alta susceptibilidad, que proporcionan información importante para la planificación urbana, el manejo y prevención de emergencias. Utilizando sistemas de información geográfica (SIG), se pueden integrar datos socioeconómicos y desarrollar estudios de vulnerabilidad social o riesgo, tomando en cuenta factores como densidad de población, nivel socioeconómico, acceso a servicios e infraestructura pública.

Los sistemas de información geográfica permiten desarrollar bases de datos georreferenciadas, que se pueden utilizar para evaluar y analizar la vulnerabilidad a deslizamientos de tierra en áreas geográficas específicas. Utilizando bases de datos de instituciones como el MAGA (Ministerio de Agricultura, Ganado y Alimentos), el INSIVUMEH (Instituto Nacional de Sismología, Vulcanología, Meteorología e Hidrología) y CONRED (Coordinadora Nacional para la Reducción de Desastres) se pueden integrar con varios procesos de análisis de datos para generar mapas analíticos. Estos datos incluyen: topografía del suelo, características geográficas, propiedades mecánicas del suelo, datos meteorológicos, vegetación, información geológica y más.

Empleando los datos y las herramientas especificadas previamente, se puede realizar un análisis de susceptibilidad de deslizamientos, para identificar áreas con alta probabilidad de la ocurrencia de eventos y determinar la distribución espacial de deslizamientos de acuerdo con las predicciones realizadas. Este análisis se puede realizar utilizando métodos de aprendizaje automático (machine learning), como árboles de decisión mediante Random Forest o XGBoost. Empleando factores permanentes como las características del suelo, pendiente, tipo de suelo, presencia de agua, estratigrafía, y factores desencadenantes meteorológicos como lluvia y viento, se puede determinar las áreas en las que exista mayor susceptibilidad a deslizamientos para el departamento de San Marcos.

IV. OBJETIVOS

A. Objetivo general

Generar mapas de predicción de deslizamientos a partir de modelos de aprendizaje supervisados utilizando bases de datos en GIS y Machine Learning.

B. Objetivos específicos

- Recolectar y organizar información geoespacial, ambiental y meteorológica relevante para su posterior análisis en sistemas de información geográfica (SIG).
- Aplicar técnicas de aprendizaje automatizado por medio de árboles de decisión.
- Analizar de forma visual los mapas generados en comparación con mapas de susceptibilidad a deslizamientos elaborados por institutos nacionales.
- Evaluar la importancia relativa de las variables geoespaciales empleadas en el modelo, con el propósito de identificar los factores que más influyen en la susceptibilidad a deslizamientos en el área de estudio.

V. MARCO TEÓRICO

A. Deslizamientos

1. Concepto

Los deslizamientos de tierra, también conocidos como movimientos en masa o movimientos de ladera, son procesos gravitacionales en los que una porción de terreno se desplaza ladera abajo. En términos generales, se trata de “un proceso de degradación natural y uno de los factores más importantes de construcción del paisaje” (González Bolaños et al., 2023), aunque a su vez constituyen un peligro geológico significativo. Una definición clásica los describe como “el movimiento de los materiales térreos de un talud, que resultan de un desplazamiento hacia abajo y hacia afuera de suelos, rocas y vegetación, bajo la influencia de la gravedad” (Ministerio de Medio Ambiente y Recursos Naturales (MARN), 2004). En otras palabras, un deslizamiento ocurre cuando una masa de suelo y/o roca pierde su estabilidad y se desplaza pendiente abajo por efecto de la gravedad (Cruden & Varnes, 1996), proceso que puede ser rápido y catastrófico o lento y casi imperceptible dependiendo de las condiciones.

Desde el punto de vista geotécnico, la ocurrencia de un deslizamiento depende del equilibrio entre las fuerzas resistentes del material de la ladera y las fuerzas desencadenantes que tienden a desestabilizarla. La gravedad por sí sola “puede mover los materiales térreos sólo cuando es capaz de vencer la resistencia del material que le impide moverse” (MARN, 2004), por lo que cualquier factor que disminuya dicha resistencia o incremente las fuerzas actuantes puede propiciar una falla. Los factores condicionantes (o pasivos) son las características intrínsecas del terreno que influyen en su estabilidad: la morfología (pendiente o inclinación del talud), la geología (litología, estructuras y propiedades mecánicas de rocas y suelos) y la hidrología interna (contenido de agua, nivel freático) son ejemplos de factores que determinan la resistencia y cohesión de la ladera (de Medio Ambiente y Recursos Naturales (MARN), 2004). Estos factores condicionantes configuran la predisposición natural de una ladera al deslizamiento, pero generalmente actúan de forma lenta o permanente en el tiempo. Cuando el balance entre resistencia y carga se ve comprometido más allá de un umbral crítico –por cambios en estos factores internos o por agentes externos– la ladera entra en estado de inestabilidad y puede desencadenarse el movimiento del material pendiente abajo (Chacón Montero, 2003).

Existen asimismo factores detonantes o desencadenantes (factores activos o externos) que inician directamente el deslizamiento al alterar bruscamente el equilibrio de la ladera. El detonante más común es la lluvia intensa y prolongada, cuya infiltración en el terreno eleva la presión de poros, aumenta el peso del suelo y reduce la fricción interna, disminuyendo drásticamente la resistencia al corte del material. De hecho, “la acción de la lluvia constituye el agente activador más regular” de deslizamientos en muchas regiones montañosas (Chacón Montero, 2003). Numerosos desastres por deslizamientos han sido asociados a eventos de lluvia extrema; por ejemplo, lluvias excepcionales pueden generar decenas de miles de deslizamientos

en una sola región (MARN, 2004). Otro factor detonante importante son los sismos: los terremotos sacuden el terreno y pueden provocar deslizamientos de diversos tipos al inducir esfuerzos dinámicos que exceden la resistencia del suelo o roca (MARN, 2004). Ejemplos históricos incluyen los miles de deslizamientos gatillados por el terremoto de Guatemala de 1976 (M 7.6) o el colapso de la ladera de Las Colinas durante el terremoto de El Salvador en 2001 (MARN, 2004). Otros detonantes posibles abarcan la actividad volcánica (erupciones que sacuden o debilitan las laderas) y las intervenciones antrópicas abruptas, como cortes excesivos de taludes, deforestación o sobre carga por construcciones, las cuales pueden desencadenar inestabilidad al modificar rápidamente las condiciones del terreno (González Bolaños et al., 2023).

En la literatura especializada se han propuesto diversas clasificaciones de los deslizamientos de tierra, atendiendo tanto al tipo de movimiento como al material involucrado. Una clasificación ampliamente aceptada internacionalmente es la propuesta por D. J. Varnes, posteriormente refinada por Cruden & Varnes, distingue cinco modos principales de movimiento: caída (desprendimiento casi libre de fragmentos de roca o suelo), vuelco (giro o volteo de un bloque alrededor de un punto inferior), deslizamiento propiamente dicho (desplazamiento coherente de una masa a lo largo de una superficie de falla, ya sea rotacional o traslacional), expansión lateral (fracturamiento y extensión lateral gradual de terrenos llanos, a menudo ligado a licuación de estratos subyacentes) y flujo (movimiento veloz de material altamente deformado que fluye como líquido) (Cruden & Varnes, 1996; Varnes, 1978). A estos se añade el tipo compuesto o complejo, que combina dos o más modos en un mismo evento (por ejemplo, una caída que se transforma en flujo al acumularse los escombros).

En contextos hispanohablantes es común emplear una clasificación simplificada en tres grandes grupos: derrumbes o caídas de rocas/suelos, deslizamientos (en sentido estricto, rotacionales o traslacionales) y flujos de lodo o detritos (MARN, 2004). Cada tipo de deslizamiento presenta características morfológicas propias (por ejemplo, las caídas generan un talud escarpado o cicatriz en la cabecera, mientras los deslizamientos dejan superficies de rotura y los flujos forman depósitos lobulados en la base) y puede requerir diferentes enfoques de estabilización. Reconocer la clasificación y los procesos específicos de cada deslizamiento es fundamental en geotecnia y geología para evaluar adecuadamente el riesgo asociado y diseñar medidas de mitigación efectivas en proyectos de ingeniería civil (Chacón Montero, 2003).

2. Categorías

a. Deslizamientos rotacionales

Ocurren cuando la superficie o falla de ruptura es curva o cóncava. Son causados por la saturación de agua en el suelo, que presenta un aumento en la presión de poro del suelo, aumentando el peso del suelo, y reduciendo propiedades del suelo resistentes a deslizamientos, como cohesión y habilidad cementante entre partículas de suelo. Este tipo de deslizamiento ocurre principalmente en suelos cohesivos uniformes o macizos rocosos con muchas fracturas y ausencia de estructura. Dependiendo de las propiedades del suelo, los deslizamientos rotacionales pueden ser simples, múltiples o sucesivos (Gijón Ferrér, 1988).



Figura 1. Deslizamiento rotacional.
(Fuente: Universidad Católica de Chile).

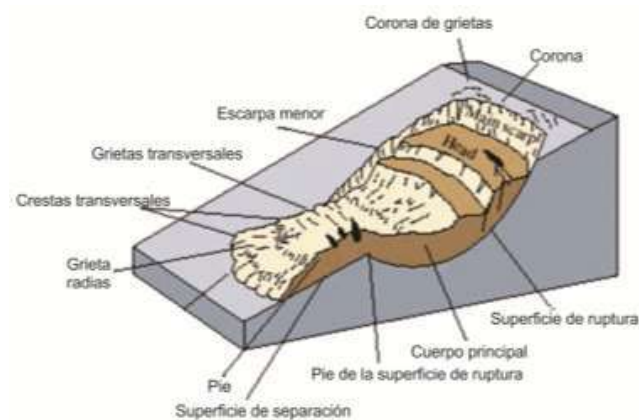


Figura 2. Esquema de deslizamiento rotacional.
(Fuente: OpenGeology, s.f.)

b. Deslizamientos traslacionales

Los deslizamientos traslacionales son movimientos de tierra en los que el suelo se desprende a favor de un plano de debilidad que tiene una dirección o pendiente paralela a la superficie del talud. Este tipo de deslizamiento es superficial debido a que ocurre en suelos que tienen una fricción mayor a medida que aumenta la profundidad, como en suelos cohesivos no uniformes, es decir, hay una intercalación de niveles con propiedades distintas (Tipos de Movimientos en Masa – Clasificaciones, 2007).



Figura 3. Deslizamiento traslacional.
(Fuente: Chunga, K. (2017))

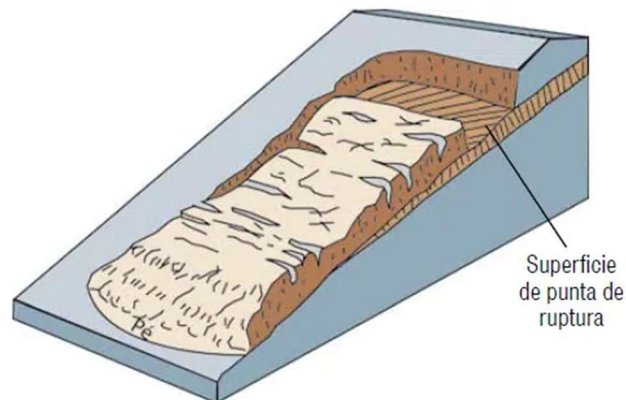


Figura 4. Esquema deslizamiento traslacional.
(Fuente: OpenGeology, s.f.)

c. Flujos

Los flujos son movimientos de masa de materiales sin cohesión, y tienden a ser susceptibles a perder considerable parte de su resistencia con movimientos. Debido a la falta de cohesión, este suelo se moviliza como un fluido lo que causa que no haya una superficie de falla o de ruptura definida (Gijón Ferrér, 1988).



Figura 5. Flujo de lodo.
(Fuente: Imagen de fuente libre, s.f.).

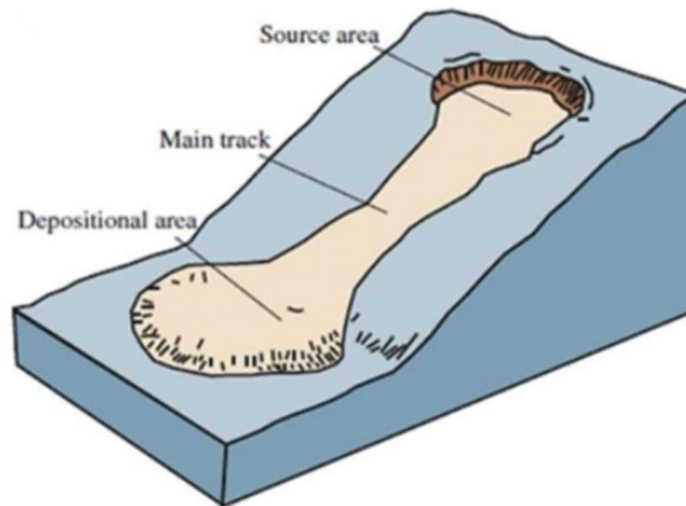


Figura 6. Esquema flujo de tierra.
(Fuente: OpenGeology, s.f.).

d. Caídas

Las caídas o desprendimientos ocurren cuando bloques de material provenientes de la parte superior de un talud se desploman, usualmente con planos de corte individuales, con una caída libre por parte de su recorrido. Suele ocurrir en taludes que presentan mayor verticalidad (Gijón Ferrér, 1988).



Figura 7. Caída de piedras en km. 259 RN12 Sur, La Castalia, San Marcos, Guatemala. (Fuente: Diario de los Altos, 2024.).

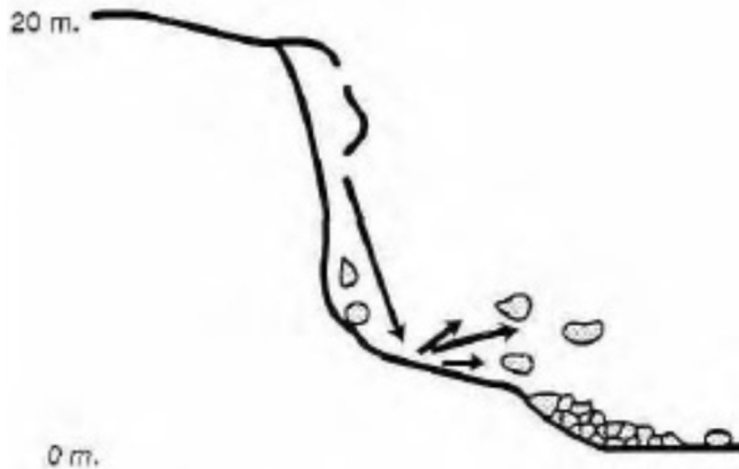


Figura 8. Esquema de desprendimientos o caídas. (Fuente: WP/WLI, 1990.).

e. Volcamientos

Similarmente a las caídas, estos implican bloques o masa de material que se desprende de taludes verticales o con pendientes elevadas. Este tipo de movimiento de masa ocurre en zonas volcánicas (Gijón Ferrér, 1988).

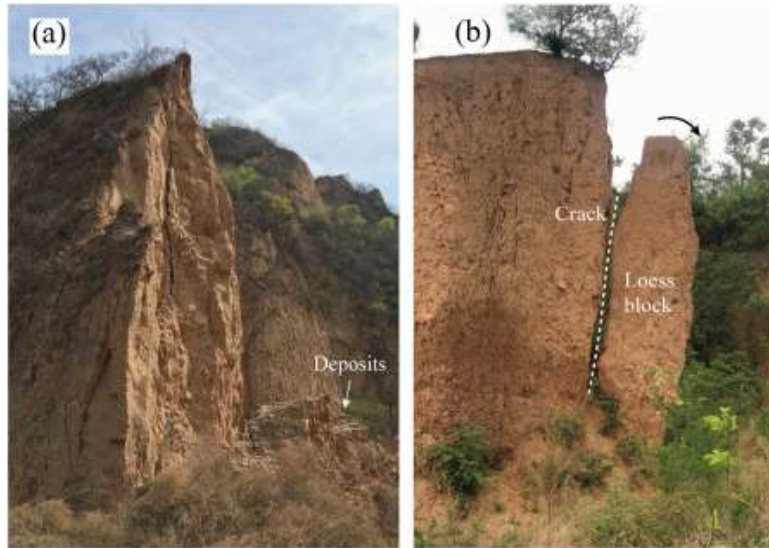


Figura 9. Volcamiento típico de suelo pulverulento.
(Fuente: Li, 2019.).

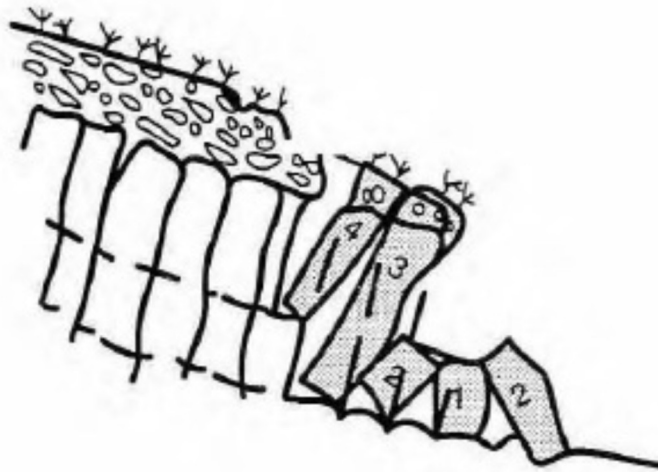


Figura 10. Esquema de volcamientos.
(Fuente: WP/WLI, 1990.).

3. Partes de un deslizamiento

Tras un evento de movimiento de masa, es posible identificar los siguientes componentes en el suelo desplazado y en el área alrededor al deslizamiento.

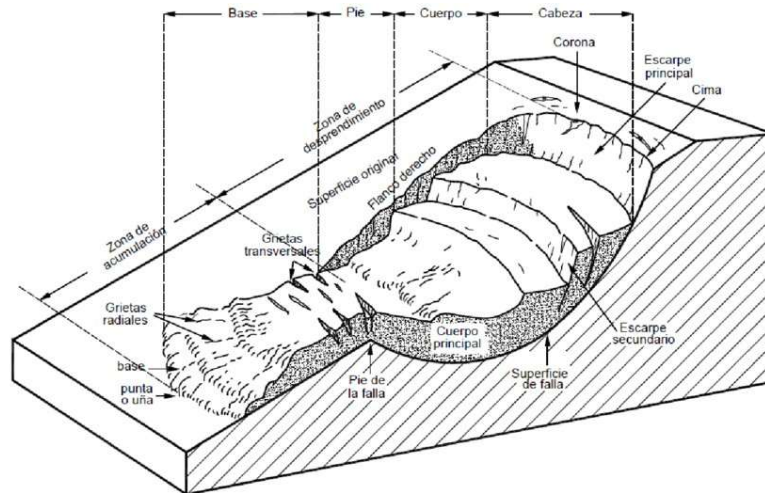


Figura 11. Componentes de un deslizamiento.
(Fuente: Cruden y Varnes, 1996).

a. Zona de cabecera

Es el área superior donde se origina el deslizamiento. Aquí se forma la *cicatriz de arranque*, una depresión que marca el inicio del movimiento del terreno (Cruden & Varnes, 1996).

b. Plano de deslizamiento

Es la superficie a lo largo de la cual se produce el movimiento. Puede ser una superficie plana, curvada o irregular, y su geometría influye directamente en la velocidad y estabilidad del deslizamiento (Highland & Bobrowsky, 2008).

c. Cuerpo del deslizamiento

Corresponde a la masa de material que se desplaza. Esta parte puede presentar deformaciones internas, como agrietamientos y desplazamientos diferenciales.

d. Cuerpo del deslizamiento

Zona de acumulación: Es el área donde se deposita el material transportado. Puede extenderse considerablemente y afectar infraestructura, cultivos o ecosistemas naturales.

e. Cuerpo del deslizamiento

Frente del deslizamiento: Parte inferior del deslizamiento, que suele mostrar una acumulación más densa de materiales y puede representar un peligro inmediato para estructuras aguas abajo (Guzzetti et al., 1999).

4. Estados de un deslizamiento

Los deslizamientos de tierra no ocurren de forma repentina, sino que atraviesan una serie de etapas que reflejan la evolución del fenómeno desde su inestabilidad inicial hasta la estabilización o reactivación del movimiento. Comprender estos estados permite evaluar la amenaza temporal y planificar medidas de mitigación adecuadas (Varnes, 1984).

a. Activo

Se encuentra en movimiento al momento de observación.

b. Suspendido

No ha presentado movimiento en los últimos 12 meses.

c. Reactivado

Deslizamiento activo que previamente fue inactivo.

d. Inactivo

Se movió en los últimos 12 meses, pero no se encuentra activo.

e. Latente

Deslizamiento inactivo que se puede volver a dar por las mismas causas que provocaron el movimiento inicial.

f. Antiguo

Un deslizamiento antiguo que no ha sido afectado por sus causas originales.

g. Estabilizado

Un deslizamiento inactivo al cual le han aplicado medidas correctivas.

h. Relicto

Deslizamiento inactivo que se dio por condiciones climáticas o geomorfológicamente diferentes a las que prevalecen actualmente.

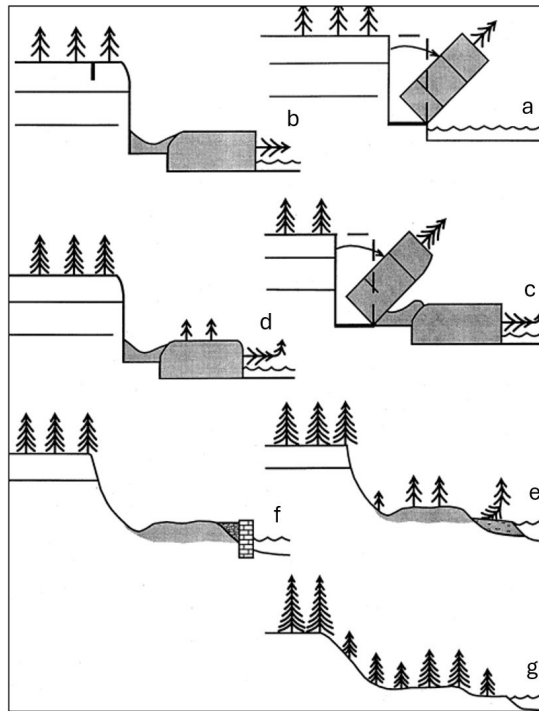


Figura 12. Estados de los deslizamientos.
(Fuente: WP/WLI,1993).

5. Causas principales de deslizamiento

a. Factores mecánicos del suelo

La estabilidad de un talud depende en gran medida de las propiedades mecánicas del suelo, las cuales determinan su resistencia al corte. Entre estas propiedades, la cohesión y el ángulo de fricción interna son fundamentales en el análisis geotécnico. La cohesión, representada como 'c', expresa la capacidad del suelo de mantenerse unido debido a fuerzas interparticulares, ya sea por cementación, atracción molecular o efectos electrostáticos. Esta propiedad es significativa en suelos finos como arcillas y limos, en donde los enlaces intergranulares tienen un papel importante. Por el contrario, en suelos granulares como arenas y gravas, la cohesión es prácticamente nula, y la resistencia al corte depende casi exclusivamente de la fricción interna (Coduto et al., 2011).

El ángulo de fricción interna, o ϕ , representa la resistencia que ofrece el suelo al desplazamiento debido al contacto entre partículas. Este valor depende del tamaño, forma, distribución granulométrica, rugosidad y grado de compactación del suelo. En general, suelos con partículas angulares y bien gradadas presentan un mayor ángulo de fricción en comparación con suelos mal gradados o de partículas redondeadas. La resistencia al corte total se puede calcular mediante el criterio de falla de Mohr-Coulomb, que se expresa como:

$$\tau = c + \sigma \cdot \tan(\phi)$$

*Ecuación 1. Resistencia al corte total de acuerdo con Mohr-Coulomb
(Fuente: Mohr-Coulomb).*

Donde:

- τ = Esfuerzo cortante
- c = Cohesión
- σ = Tensión de normal
- ϕ = Ángulo de rozamiento interno

También deben considerarse otros factores como la densidad relativa, que afecta la compactación y porosidad del suelo; una mayor densidad implica menor contenido de aire y mayor contacto entre granos, lo que contribuye a una mayor fricción interna. Asimismo, la estructura del suelo y la presencia de discontinuidades como planos de estratificación o fisuras pueden debilitar la masa del talud, facilitando el desarrollo de superficies de falla (Berman & Popov, 2020). El nivel freático es otro factor clave: cuando éste asciende dentro de un talud, se incrementa la presión de poros en zonas críticas. La elevación repentina del nivel freático, por ejemplo, tras tormentas o rupturas de canales, puede provocar fallas inmediatas. Además, en zonas donde existe infiltración lateral desde fuentes externas (como ríos o cuerpos de agua cercanos), la saturación prolongada actúa como un agente desestabilizador constante. En muchos casos, la falta de drenaje interno en los taludes naturales o intervenidos agrava la acumulación de agua, favoreciendo condiciones de licuación parcial o pérdida súbita de resistencia en suelos finos (Blight, 2013).

La permeabilidad del suelo, definida como la capacidad de este para permitir el paso del agua, condiciona la velocidad con la que el agua se redistribuye dentro del talud. Suelos arcillosos con baja permeabilidad tienden a retener el agua, manteniendo la presión de poros elevada durante largos periodos. Por el contrario, suelos arenosos o de grava, aunque permiten una rápida infiltración, pueden presentar inestabilidad si se encuentran sobre capas impermeables que actúan como planos de falla lubricados. Por tanto, el perfil estratigráfico y la presencia de interfaces permeables-impermeables deben ser cuidadosamente evaluados al estudiar la estabilidad de taludes (Smith & Fleming, 2014).

b. Infiltración y saturación del suelo

La infiltración es el movimiento del agua de la superficie del terreno hacia el suelo o la roca por medio de los poros o discontinuidades de este. Puede provocar una superficie de talud inundada, que ocurre porque la intensidad de la lluvia es mayor que la cantidad de agua que se puede infiltrar en la tierra. Una tasa de infiltración alta puede generar problemas de presión de agua interna, entre más, que pueden provocar un deslizamiento. A causa de esto, la habilidad del talud de auto drenarse, depende de las formaciones geológicas adyacentes al punto de infiltración. La presencia de mantos freáticos o de puntos de infiltración masiva de agua sobre el talud como ríos o cuerpos de agua contribuyen en la estabilidad de un talud (Suarez Días, 1998).

c. Taxonomía y series de suelos

La taxonomía de suelos es un sistema de clasificación científica que organiza los suelos en categorías jerárquicas con base en sus propiedades físicas, químicas y morfológicas. Esta clasificación permite entender el comportamiento del suelo bajo distintas condiciones, lo cual es crucial en estudios de susceptibilidad a deslizamientos. En países como Guatemala, la diversidad geológica y climática produce una gran variedad de tipos de suelo, muchos de los cuales presentan características que los hacen más vulnerables a inestabilidades, como una baja cohesión o una estructura débil (Brady & Weil, 2008).

Las series de suelos constituyen una unidad más específica dentro de la taxonomía, definida por horizontes del perfil del suelo con características similares en textura, estructura, drenaje y composición mineral. Estas series son fundamentales en estudios de riesgo porque determinan la capacidad del terreno para absorber agua, su permeabilidad y resistencia mecánica. Por ejemplo, suelos arcillosos con mal drenaje y alta plasticidad tienden a saturarse fácilmente, aumentando la probabilidad de deslizamientos en eventos de lluvia intensa (K. Smith & Mullins, 2013).

d. Factores detonantes

Los factores detonantes de deslizamientos son aquellos elementos o procesos inmediatos que desencadenan la falla del terreno. A diferencia de los factores condicionantes, que predisponen una zona a deslizarse, los detonantes son eventos puntuales que superan la resistencia interna del suelo. Entre los más frecuentes se encuentran las precipitaciones intensas, los sismos, el desmonte abrupto de vegetación y las actividades antrópicas como cortes inadecuados en laderas o construcciones mal planificadas (Guzzetti et al., 2005).

e. Sismicidad

De acuerdo con Moya y Corominas (1997, 2002), como fue citado en Fernández et al. (2016), se identificaron tres patrones meteorológicos clave que aumentan significativamente la probabilidad de ocurrencia de roturas de laderas y deslizamientos. Estos patrones son: (1) temporadas de lluvias intensas de corta duración; (2) períodos prolongados de lluvias de intensidad moderada a baja que se extienden por varios días o incluso semanas; y (3) estaciones caracterizadas por una humedad ambiental anormalmente elevada (Moya & Corominas, 1997). Estas condiciones hidrometeorológicas tienen la capacidad de modificar el equilibrio de las pendientes al incrementar el contenido de humedad del suelo, reducir la resistencia al corte y aumentar el peso propio del material, generando un entorno propicio para la inestabilidad del terreno.

- **Precipitación**

La precipitación es, en muchos casos, el factor detonante más común en regiones tropicales y montañosas como Guatemala. La infiltración del agua de lluvia incrementa la presión de poros y reduce la cohesión entre partículas del suelo, disminuyendo su estabilidad. Especialmente críticos son los eventos de lluvia acumulada durante varios días, donde la saturación progresiva

de las capas del suelo puede conducir a deslizamientos masivos incluso sin precipitaciones extraordinarias en el momento del colapso (Wieczorek, 1996).

Con el objetivo de comprender esta relación entre la precipitación y la ocurrencia de deslizamientos, diversos estudios han recurrido al análisis estadístico para identificar correlaciones representativas entre eventos de lluvia y procesos de remoción en masa. En particular, el trabajo de Remondo et al. (2008), citado por Rivas et al. (Rivas & Remondo, 2016), constituye un ejemplo valioso de esta aproximación. En dicho estudio, se identificaron seis episodios de lluvia intensa en el área de análisis que estuvieron directamente vinculados con la ocurrencia de múltiples deslizamientos superficiales y flujos de detritos. No obstante, los autores reconocen que la información no es concluyente, dado que no se contaba con datos precisos sobre las condiciones previas de estabilidad de los taludes afectados, lo que limita la capacidad de atribuir causalidad directa.

A pesar de dichas limitaciones, el análisis permitió establecer que, en los casos estudiados, el factor determinante para la activación de los deslizamientos fue la intensidad de la lluvia, más que su duración (Moya & Corominas, 1997). Este hallazgo resulta particularmente relevante, ya que sugiere que la susceptibilidad a deslizamientos puede responder de forma aguda a eventos de precipitación extrema, aunque breves, especialmente cuando el terreno ya presenta condiciones predisponentes como saturación previa o materiales sueltos. Por lo tanto, para estudios de pronóstico o evaluación de riesgo como el presente, es fundamental considerar parámetros que capturen la intensidad de la lluvia en ventanas temporales específicas, tales como la precipitación acumulada en los 7 días previos a un evento, en lugar de utilizar únicamente promedios mensuales o acumulados estacionales.

1. Metodologías Mora Vahrson de determinación de áreas susceptibles a deslizamientos

La metodología de Mora-Vahrson es un enfoque utilizado para determinar los riesgos de deslizamientos en zonas tropicales con actividad sísmica (Preciado, 2006). Dicha metodología contempla cinco factores, divididos en dos grupos, elementos pasivos y activos. Los factores pasivos representan riesgos inherentes o permanentes del área, como pendiente, litología, humedad, etc. Los factores activos consideran la intensidad sísmica e intensidad de lluvias, y representan los factores detonantes. A cada factor se le asigna un índice de influencia, Valor S, de acuerdo con un determinado sitio, y al combinar esto con un peso específico ponderado, se calcula un nivel relativo de amenaza, H. (Mora & Vahrson, 1994).

En donde se tienen dos variables, tomando en cuenta los factores pasivos y los factores activos:

$$H = (S_r \cdot S_l \cdot S_h) \cdot (T_s \cdot T_p)$$

*Ecuación 2. Nivel relativo de amenaza según Mora & Vahrson.
(Fuente: Mora & Vahrson, 1994).*

Donde:

- S_r = Factor de susceptibilidad por pendiente
- S_l = Factor de susceptibilidad por litología
- S_h = Factor de susceptibilidad por humedad
- T_s = Factor disparante por sismicidad
- T_h = Factor disparante por humedad

La clasificación de distintas litologías y sus susceptibilidades, de acuerdo con casos representativos de Costa Rica y Centroamérica es lo siguiente:

Litología	Calificación	Factor SI
Aluvi3n compacto permeable; caliza permeable, intrusiones ligeramente fisuradas, basalto, ignimbrita, gneis; bajo grado de meteorizaci3n, bajo nivel freático, fracturas limpias con rugosidad, alta resistencia al corte.	Bajo	1
Mayor grado de meteorizaci3n de dichas litologías y de rocas sedimentarias masivas duras; menor resistencia al corte y fracturas cortantes.	Moderado	2
Rocas volcánicas sedimentarias, intrusivas, metam3rficas considerablemente erosionadas, suelos regolíticos arenosos compactados, fracturamiento considerable y niveles freáticos fluctuantes.	Medio	3
Rocas considerablemente erosionadas, alteradas hidrotermalmente de cualquier tipo, fuertemente fracturadas y fisuradas, rellenas de arcilla; suelos piroclásticos y fluvio-lacustres poco compactados, mantos freáticos poco profundos.	Alto	4
Rocas extremadamente alteradas, suelos aluviales, coluviales y residuales de baja resistencia al corte, mantos freáticos poco profundos.	Muy alto	5

Tabla 1. Clasificaci3n por litología y valor SL.
(Fuente: Mora, Vahrson, 1994).

Valor de pendiente Rr (m/km²)	Calificaci3n	Factor Sr
0-75	Muy Bajo	0
76-175	Bajo	1
176-300	Moderado	2
301-500	Medio	3
501-800	Alto	4
>800	Muy Alto	5

Tabla 2. Valores relativos de relieve (Rr), su clasificaci3n y valores resultantes del factor de pendiente (Sr).
(Fuente: Mora, Vahrson, 1994).

Precipitación mensual promedio (mm/mes)	Valor
<125	0
125–250	1
>250	2

Tabla 3. Clasificación de valores promedio mensuales de precipitación.
(Fuente: Mora, Vahrson, 1994).

Intensidad (Escala Modificada de Mercalli)	Calificación	Factor Ts
III	Ligero	1
IV	Muy Bajo	2
V	Bajo	3
VI	Moderado	4
VII	Medio	5
VIII	Considerable	6
IX	Importante	7
X	Fuerte	8
XI	Muy Fuerte	9
XII	Extremadamente Fuerte	10

Tabla 4. Determinación del factor de intensidad sísmica como detonante para deslizamientos, empleando valores de intensidad de 100 años y la escala modificada de Mercalli.
(Fuente: Mora, Vahrson, 1994).

Precipitación máxima, n > 10 años; Tr = 100 años	Lluvia n < 10 años; promedio	Calificación	Factor Ts
<100 mm	<50 mm	Muy Bajo	1
101–200 mm	51–90 mm	Bajo	2
201–300 mm	91–130 mm	Medio	3
301–400 mm	131–175 mm	Alto	4
>400 mm	>175 mm	Muy Alto	5

Tabla 5. Índice de intensidad de precipitación resultante de la clasificación máxima de precipitaciones diarias para un periodo de retorno de 100 años. En caso de datos pluviométricos bajos se considera la columna 2.
(Fuente: Mora, Vahrson 1994).

2. Medidas de mitigación de deslizamientos

La mitigación de deslizamientos comprende un conjunto de estrategias diseñadas para reducir la susceptibilidad, la frecuencia o el impacto de los movimientos de ladera. Estas medidas pueden clasificarse en estructurales, que implican intervenciones físicas directas sobre el terreno, y no estructurales, que se enfocan en la planificación, monitoreo y gestión del riesgo. Su implementación depende del tipo de deslizamiento, las condiciones geotécnicas del sitio y la exposición de elementos vulnerables.

Entre las medidas estructurales más comunes se encuentran las obras de estabilización mecánica del talud. Estas incluyen la construcción de muros de contención, pantallas ancladas, refuerzo con geotextiles, o mallas metálicas que previenen caídas de rocas. Asimismo, la modificación geométrica del terreno mediante cortes controlados, rellenos compactados, o la reducción de la pendiente puede mejorar significativamente la estabilidad general (Wen et al., 2020). En terrenos con suelos saturables, la implementación de sistemas de drenaje superficial y subterráneo (como drenes horizontales, zanjas de coronación o cunetas de desvío) es fundamental para disminuir la presión de poro y evitar la pérdida de resistencia al corte (Robinson, 2022).

Desde el enfoque no estructural, la planificación territorial juega un papel clave. La delimitación de zonas de riesgo y la restricción del uso del suelo en laderas inestables permiten reducir la exposición humana a los peligros. Además, los sistemas de monitoreo y alerta temprana basados en sensores de humedad, inclinómetros o estaciones meteorológicas pueden detectar condiciones críticas antes de que ocurran deslizamientos, facilitando la evacuación preventiva (Huabin et al., 2005). La educación comunitaria y los programas de gestión del riesgo a nivel local complementan estas acciones, promoviendo una cultura de prevención.

Algunas soluciones son las siguientes:

a. Drenaje superficial y subterráneo de taludes

Una de las medidas más eficaces para reducir la inestabilidad del terreno es la implementación de sistemas de drenaje. El exceso de agua en el suelo puede incrementar la presión intersticial y reducir la resistencia al corte, condiciones propicias para deslizamientos. Por ello, se utilizan cunetas, zanjas de coronamiento, drenes y drenes subhorizontales que canalizan el agua fuera de la masa inestable (Reis et al., 2008). El mantenimiento de esta medida consiste en la limpieza periódica de conductos, revisión posterior a lluvias intensas, y reemplazo del material filtrante cuando se detecta obstrucción.



Figura 13. Drenes subhorizontales o californianos.
(Fuente: Civogal, s.f.).



Figura 14. Cunetas como sistema de drenaje para contención de taludes.
(Fuente: Cueva del ingeniero civil, s.f.).

b. Soil nailing (anclajes pasivos en el suelo)

Esta técnica consiste en insertar barras de acero (clavos) en perforaciones a lo largo de un talud. Las barras se fijan mediante lechada de cemento y actúan como refuerzo pasivo, estabilizando el terreno al aumentar su resistencia al corte. Dichas inclusiones van separadas en grillas a lo largo del talud (la separación varía de acuerdo con el tipo de suelo, altura de muro, tamaño de la sobrecarga, etc.), reforzadas con una o varias capas de acero y muro de concreto lanzado o geotextil. Es útil en laderas empinadas, taludes de carreteras y zonas urbanizadas con riesgo de colapso. En caso sea necesario, es importante colocar el drenaje requerido, ya sean drenes subhorizontales, barbacanas, drenaje francés, entre otros (Ho & Barratt, 2005).

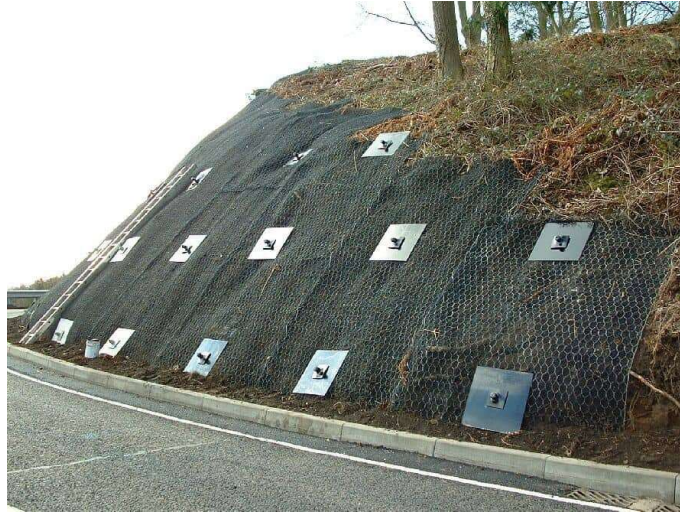


Figura 15. Talud protegido con soil nailing recubierto con geotextil.
(Fuente: Shah, M, 2024).



Figura 16. Muro de soil nailing recubierto con shotcrete (concreto lanzado).
(Fuente: UMA Geotechnical Construction).

c. Geo mantas y coberturas geo sintéticas

Las geo mantas son materiales geo sintéticos diseñados para controlar la erosión superficial de taludes y facilitar el establecimiento de vegetación. Estas coberturas se instalan directamente sobre la superficie del terreno expuesto para reducir la acción directa de la lluvia y el escurrimiento superficial sobre el suelo. Existen geo mantas sintéticas (como las fabricadas con polipropileno o poliéster) y geo mantas biodegradables (hechas de coco, yute o paja), que se degradan naturalmente con el tiempo mientras se desarrolla la cobertura vegetal permanente (Koerner, 2012).

Su anclaje al terreno se realiza comúnmente mediante estacas metálicas o de madera, grapas de acero en forma de “U” y, en algunos casos, se refuerzan con mallas metálicas para

mejorar su adherencia al terreno en zonas con pendientes pronunciadas. En taludes más inestables, también se combinan con técnicas de revegetación hidrosiembra o incluso con mallas tridimensionales que refuerzan el arraigo de las raíces (Koerner, 2012).



*Figura 17. Muro recubierto con geotextil.
(Fuente: ItalPlast, s.f.)*

d. Shotcrete (concreto lanzado)

El shotcrete, o concreto proyectado, es una mezcla de cemento, arena, grava fina y aditivos que se aplica neumáticamente sobre superficies irregulares a alta presión. Se utiliza principalmente para recubrir taludes inestables, protegiéndolos contra la erosión superficial, la intemperización y pequeños movimientos de suelo. Se puede aplicar en dos modalidades: vía seca (mezcla seca transportada por aire y mezclada con agua en la boquilla) o vía húmeda (mezcla húmeda bombeada directamente). Ambas técnicas se proyectan directamente sobre el talud utilizando mangueras y compresores de aire de alta presión. Para mejorar su adherencia y resistencia, el shotcrete se combina frecuentemente con mallas metálicas de refuerzo, barras de acero o fibras sintéticas (Das, 2010).



*Figura 18. Muro recubierto con concreto lanzado.
(Fuente: Geotechnical Engineering, Australia, s.f.)*

e. Muros de contención de concreto armado

Los muros de contención de concreto armado son estructuras rígidas diseñadas para resistir los empujes laterales del terreno, estabilizando zonas donde los taludes naturales no son viables o donde la infraestructura adyacente debe protegerse. Este tipo de muro es especialmente útil en carreteras de montaña, desarrollos urbanos en laderas y zonas con cortes verticales de terreno (Rodríguez, 2006).

Existen distintos tipos de muros según su configuración estructural: muros en voladizo, muros con contrafuertes, muros de gravedad, y muros anclados, entre otros. La selección del tipo depende de la altura, condiciones geotécnicas del terreno, y el tipo de carga que se espera soportar. Estos muros suelen incorporar drenajes internos como tubos PVC o drenes franceses, ubicados en la parte posterior del muro, para evitar la acumulación de agua y las presiones hidrostáticas que podrían comprometer su estabilidad.



Figura 19. Muro de contención de concreto armado.
(Fuente: Geotecnia y Mecánica de Suelos, 2023).

1. Legislación relacionada a deslizamientos o desastres naturales

a. Legislación nacional (Guatemala)

Coordinadora Nacional para la Reducción de Desastres (CONRED) – Decreto Legislativo 109-96 Este decreto establece la creación y funcionamiento del sistema CONRED como ente rector de la gestión del riesgo en Guatemala. La ley regula las acciones de prevención, mitigación, preparación, respuesta y recuperación ante fenómenos naturales o provocados. En su marco institucional, se establece la necesidad de elaborar mapas de amenazas y vulnerabilidad, siendo los deslizamientos uno de los fenómenos incluidos. Además, se contempla la obligación de las entidades públicas de incorporar criterios de gestión de riesgo en sus políticas y proyectos de desarrollo (de la República de Guatemala, 1996).

Acuerdo Número 06-2022 de CONRED (Guatemala, 18 de mayo de 2011) Aunque aprobado en 2022, este acuerdo actualiza procedimientos y lineamientos institucionales emitidos desde 2011. Define las metodologías de evaluación de daños y análisis de necesidades (EDAN), así como el uso de criterios técnicos para clasificar zonas de riesgo, entre ellas aquellas propensas

a deslizamientos. También establece protocolos de monitoreo, alerta y evacuación en zonas de amenaza alta o muy alta (CONRED, 2022).

Secretaría de Planificación y Programación de la Presidencia (SEGEPLAN) – Guía para el análisis y gestión del riesgo en proyectos de inversión pública Esta guía establece que todo proyecto de inversión financiado con fondos públicos debe evaluar su exposición a amenazas naturales, incluyendo los deslizamientos. Propone una metodología de análisis de amenazas, vulnerabilidad y riesgos, con la obligación de integrar medidas de mitigación en el diseño y ejecución de proyectos. Esto refuerza la planificación preventiva en obras civiles y de infraestructura (SEGEPLAN, 2018).

b. Legislación internacional y marcos de referencia

Marco de Sendai para la Reducción del Riesgo de Desastres 2015–2030 (ONU) Adoptado por la Asamblea General de las Naciones Unidas, este marco internacional establece como prioridad la reducción del riesgo de desastres a través del fortalecimiento de capacidades, identificación de amenazas y la planificación territorial con enfoque preventivo. En el caso de los deslizamientos, el Marco de Sendai impulsa la incorporación de sistemas de alerta temprana, inventarios de amenazas y la promoción de la resiliencia comunitaria frente a movimientos de ladera (United Nations Office for Disaster Risk Reduction (UNDRR), 2015).

Directrices de la Estrategia Internacional para la Reducción de Desastres (EIRD/UNDRR) Estas directrices, promovidas por la Oficina de las Naciones Unidas para la Reducción del Riesgo de Desastres (UNDRR), recomiendan a los países fortalecer su marco legal e institucional para abordar fenómenos como los deslizamientos mediante políticas públicas, regulaciones de uso del suelo y planificación del desarrollo sostenible (UNDRR, 2019).

Normas técnicas del Manual de Esferas del Proyecto Esfera: Aunque centrado en respuestas humanitarias, el Proyecto Esfera establece normas mínimas para la gestión de emergencias, incluyendo las relacionadas con deslizamientos. Requiere la identificación previa de zonas inseguras, la reubicación segura de comunidades y la planificación de asentamientos en función del riesgo geológico (Project, 2018).

B. Características geográficas del área de estudio – San Marcos, Guatemala

1. Descripción general

El departamento de San Marcos presenta una amplia variación altitudinal que influye en su diversidad climática y ecológica. Las elevaciones oscilan desde niveles cercanos al mar en las zonas costeras hasta las cimas de los volcanes más altos de Centroamérica. Por ejemplo, el volcán Tajumulco alcanza una altitud de 4,220 metros sobre el nivel del mar (msnm). En contraste, las áreas costeras del sur del departamento se encuentran a altitudes cercanas a 0 msnm. La altitud media del departamento es aproximadamente 1,345 msnm (Topographic Map, s.f.)

a. Pendiente

La topografía de San Marcos es variada, con pendientes que reflejan su compleja orografía. En las regiones montañosas, especialmente en las cercanías de los volcanes Tajumulco y Tacaná, las pendientes son pronunciadas, superando en algunos casos el 20%. Estas áreas presentan laderas empinadas que contribuyen a la susceptibilidad a deslizamientos de tierra. Por otro lado, las zonas costeras y planicies aluviales del sur del departamento tienen pendientes suaves, generalmente inferiores al 3%, lo que favorece actividades agrícolas y asentamientos humanos (SEGEPLAN, 2020).

b. Relieve

El relieve de San Marcos está dominado por la presencia de la Sierra Madre, que atraviesa el departamento de noroeste a sureste. Esta cadena montañosa alberga formaciones volcánicas significativas, como los mencionados volcanes Tajumulco y Tacaná. Además, el departamento cuenta con valles, mesetas y planicies costeras. Las tierras altas del norte y centro presentan un relieve accidentado con elevaciones que superan los 3,000 msnm, mientras que, hacia el sur, el terreno desciende gradualmente hacia la llanura costera del Pacífico, caracterizada por su relieve plano y suelos fértiles adecuados para la agricultura (MAGA, 2011).



*Figura 20. Mapa de relieve del departamento de San Marcos, Guatemala.
(Fuente: Yarr65 (nombre de usuario), s.f.).*

2. Fisiografía

El departamento de San Marcos dividido en cuatro regiones fisiográficas, divididas de acuerdo con sus características físicas, geomorfológicas, y tipos de suelos.

- Tierras altas cristalinas: Montañas con altas pendientes.
- Tierras altas o cadenas volcánicas: Se encuentran los 2 volcanes más altos, siendo el Tajumulco con una altura de 4,220 msnm, y el Tacaná con 4,093 msnm.
- Pendiente volcánica reciente: Son de reciente formación, montañas y volcanes que tienen una pendiente de entre 5 y 12%, con una elevación entre los 100 y 500 msnm.
- Llanura costera del pacífico: son tierras generalmente planas. Las elevaciones suelen ser menores a los 100 metros, y las características de drenaje de sus suelos son deficientes (SEGEPLAN, 2024).

En el siguiente cuadro se describen las regiones mencionadas:

Región fisiográfica	Gran paisaje	Sub-paisaje	Sub-paisaje
Tierras altas cristalinas	• Altiplano occidental	• Valle intercolinares	• Montañas o colinas
Tierras altas o cadenas	• Conos volcánicos • Montañas o colinas • Valles intercolinares	• Conos, faldas volcánicas	• Montañas o colinas ligeramente escarpadas • Montañas o colinas moderadamente escarpadas • Montañas o colinas fuertemente escarpadas
Pendiente volcánica reciente	• Pie del monte • Abanicos • Aluviales	• Parte apical • Parte distal	
Llanura costera del Pacífico	• Planicie costera • Áreas inundables	• Terraza alta • Terraza	

Tabla 6. Regiones fisiográficas departamento de San Marcos.
(Fuente: IGN, 2000).

3. Características meteorológicas

San Marcos se encuentra dentro de tres de las regiones climáticas de Guatemala, Región Pacífico, Región Bocacosta, Región Occidente (INSIVUMEH, 2020).

- **Región Bocacosta:** Región que atraviesa desde San Marcos hasta Santa Rosa y se ubica a lo largo de la región montañosa, en el descenso del altiplano hacia la planicie costera hacia el Océano Pacífico. Esta región cuenta con los índices de precipitación más altos del país. Los niveles de temperatura aumentan al aproximarse al litoral. El clima generalizado es de

género semicálido, sin estación fría definida, con carácter muy húmedo sin estación seca definida (Servicio Meteorológico de Guatemala, s.f.).

- **Región Pacífico:** Es una región cuya extensión abarca desde los 0 a los 600 msnm. Tiene precipitaciones con menos intensidad que la región Bocacosta, que empiezan a disminuir al acercarse al litoral marítimo. Tiene registros altos de temperatura. Dentro de esta región hay climas cálidos sin una estación fría definida, y su carácter es húmedo a semiseco (Servicio Meteorológico de Guatemala, s.f.).
- **Región Occidente:** Consiste en la región más montañosa del territorio nacional, con montañas cuyas altitudes se encuentran entre los 1,600 y 4,200 msnm. Los accidentes geográficos de la región generan una gran diversidad de microclimas. Estos climas varían entre templados y semifríos, siendo húmedos y semisecos.

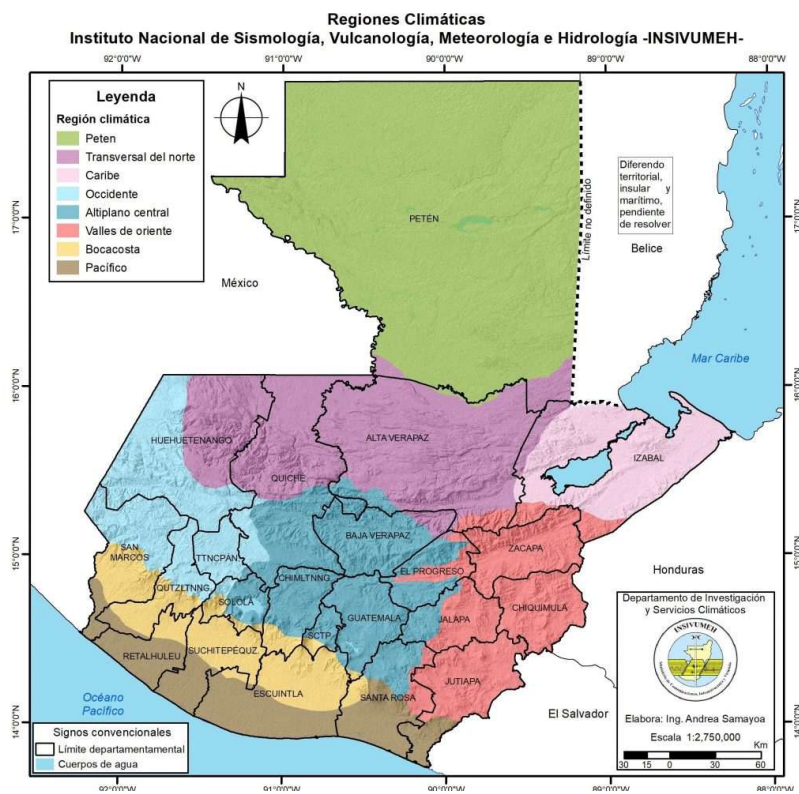


Figura 21. Mapa de regiones climáticas de Guatemala.
(Fuente: Samayoa, A., 2014).

a. Precipitación

El departamento de San Marcos se caracteriza por una alta pluviosidad, influenciada por su ubicación geográfica y su variada topografía. La temporada de lluvias se extiende aproximadamente desde principios de mayo hasta finales de octubre, con un pico en los meses de junio a septiembre. Durante este periodo, las precipitaciones mensuales pueden superar los 200 mm, alcanzando su máximo en septiembre con un promedio de 266 mm. En contraste, la

estación seca abarca de noviembre a abril, siendo enero el mes más seco con un promedio de 4 mm de precipitación (Weather Spark, s.f.)

La distribución espacial de la lluvia varía dentro del departamento. Las zonas costeras y de la Boca Costa presentan mayores acumulados anuales, con promedios que oscilan entre 3,100 y 4,000 mm, mientras que las áreas del altiplano occidental registran entre 900 y 1,200 mm anuales. Esta variabilidad en la precipitación está asociada a factores como la altitud, la orientación de las montañas y la influencia de sistemas meteorológicos regionales (SEGEPLAN, 2021).

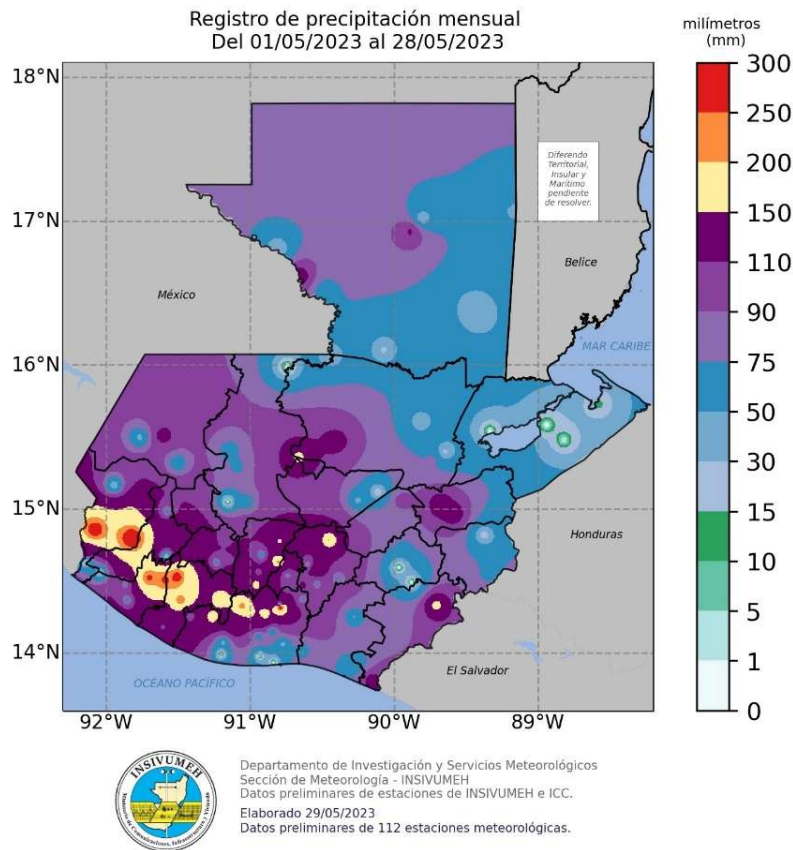


Figura 22. Mapa de registro de precipitación anual de Guatemala.
(Fuente: INSIVUMEH, 2023).

b. Viento

La velocidad del viento en San Marcos muestra variaciones estacionales. Durante los meses de diciembre a marzo, se observan velocidades promedio diarias que oscilan entre 5.5 y 6.2 km/h, siendo marzo el mes más ventoso del año. En contraste, julio es el mes más calmado, con velocidades promedio alrededor de 4.6 km/h (Weather Spark, s.f.).

La dirección predominante del viento varía a lo largo del año, aunque durante la temporada seca es común la presencia de vientos del norte, asociados a frentes fríos que afectan la región.

Estos vientos pueden intensificarse en las zonas montañosas, especialmente en los municipios ubicados en el altiplano, donde la topografía canaliza y acelera las corrientes de aire (INSIVUMEH, 2020).

4. Características geológicas

a. Características geotectónicas

San Marcos se sitúa en una región de significativa actividad tectónica, influenciada por la interacción de las placas tectónicas del Caribe y de Cocos. Esta convergencia ha dado lugar a la formación de la Sierra Madre de Chiapas, una cadena montañosa que atraviesa el departamento y alberga los picos más elevados de Centroamérica, como los volcanes Tajumulco (4,220 msnm) y Tacaná (4,092 msnm) (Denyer, 2009).

La dinámica de subducción de la placa de Cocos bajo la placa del Caribe es responsable de la actividad sísmica y volcánica en la región. Esta interacción ha generado estructuras geológicas complejas, incluyendo fallas y sistemas de fracturas que condicionan la sismicidad del área. Un ejemplo notable es el sismo de magnitud 7.4 ocurrido el 7 de noviembre de 2012, que afectó significativamente a San Marcos y departamentos vecinos (EERI & AGIES, 2012).

Además, en San Marcos se encuentra la Caldera de San Marcos, una estructura geológica de aproximadamente 12 km por 9 km, asociada a actividad volcánica pasada. Esta caldera alberga manifestaciones geotérmicas, como fuentes termales y áreas de alteración hidrotermal, indicando la presencia de sistemas geotérmicos activos con potencial para la generación de energía (Ministerio de Energía y Minas, 2015).

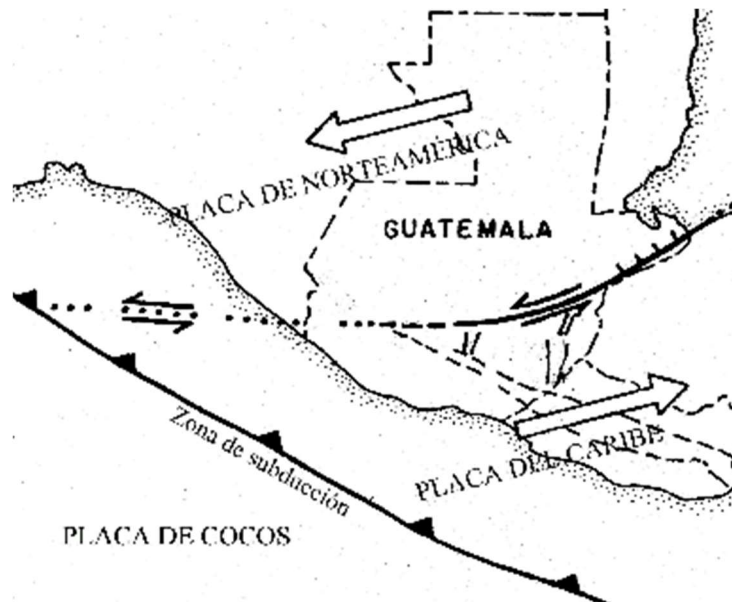


Figura 23. Mapa geotectónico del país de Guatemala.
(Fuente: Asturias, 2003).

b. Características geomorfológicas

El relieve de San Marcos es resultado de procesos geológicos y geomorfológicos que han modelado su superficie a lo largo del tiempo. La Sierra Madre de Chiapas domina el paisaje del norte del departamento, caracterizada por montañas y volcanes de gran altitud, como el Tajumulco y el Tacaná. Estas elevaciones presentan laderas escarpadas y valles profundos formados por la erosión fluvial y la actividad tectónica (MAGA, 2011).

Al sur de la sierra, el terreno desciende hacia la planicie costera del Pacífico, una extensa llanura aluvial formada por la deposición de sedimentos transportados por ríos como el Suchiate y el Naranjo. Esta región se caracteriza por suelos fértiles y pendientes suaves, lo que la hace propicia para la agricultura y el asentamiento humano (MAGA, 2011).

Entre las unidades geomorfológicas destacadas se encuentran los abanicos aluviales, como el del río Cabús, formado por la acumulación de materiales transportados desde las zonas altas hacia la planicie. Estos abanicos presentan pendientes moderadas y suelos de origen volcánico, adecuados para diversas actividades agrícolas (MAGA, 2011).

C. Sistemas de Información Geográfica (SIG)

1. Definición

Los Sistemas de Información Geográfica (SIG), también conocidos como Geographic Information Systems (GIS), son herramientas informáticas diseñadas para capturar, almacenar, procesar, analizar y visualizar datos geográficos o espaciales con el fin de apoyar la toma de decisiones en múltiples disciplinas. Su función va más allá de la simple representación cartográfica, permitiendo establecer relaciones entre variables espaciales y atributos temáticos asociados a entidades geográficas. En términos generales, un SIG combina bases de datos alfanuméricas con mapas digitales para producir análisis espaciales complejos que serían imposibles de realizar manualmente (Longley et al., 2015).

Uno de los elementos distintivos de los SIG es que manejan información con doble componente: una parte espacial, que hace referencia a la localización o geometría de un objeto (punto, línea o polígono), y una parte temática, que contiene los atributos o características asociadas a ese objeto (por ejemplo, tipo de suelo, cobertura vegetal, uso del terreno). Esta estructura permite modelar fenómenos geográficos en distintas escalas, con lo cual los SIG se vuelven esenciales para el estudio de procesos donde la posición relativa y la distribución espacial son factores clave, como en la planificación urbana, la gestión de riesgos naturales o los estudios ambientales (Santos Preciado, 2010b).

A diferencia de los sistemas de cartografía tradicionales, los SIG permiten trabajar de forma dinámica con grandes volúmenes de datos georreferenciados, integrando fuentes heterogéneas como imágenes satelitales, modelos digitales de elevación, datos censales, sensores remotos y registros de campo. Además, incluyen funciones analíticas avanzadas como el análisis de proximidad, superposición de capas (overlay), generación de buffers, interpolación espacial y

análisis multicriterio, lo cual amplía su aplicación en proyectos de ingeniería, ordenamiento territorial y monitoreo ambiental (Burrough & McDonnell, 1998).

En el contexto actual, los SIG no solo se utilizan como herramientas de escritorio, sino que se han expandido hacia plataformas web (WebGIS) y entornos móviles, facilitando el acceso y la gestión colaborativa de datos espaciales en tiempo real. Esto ha permitido su incorporación en sistemas de alerta temprana, gestión de infraestructuras urbanas inteligentes, análisis de big data espacial y toma de decisiones en políticas públicas. En suma, los SIG se han convertido en una pieza fundamental en el análisis geográfico aplicado a múltiples escalas y disciplinas (ESRI, 2022).

2. Usos de los sistemas de información geográfica

Los Sistemas de Información Geográfica (SIG) se han consolidado como una herramienta fundamental para abordar problemas de carácter espacial y geográfico en diversas disciplinas, gracias a su capacidad para integrar datos, analizar relaciones espaciales y representar fenómenos complejos. De acuerdo con Rhind (1990) (Santos Preciado, 2010), los problemas susceptibles de ser analizados mediante SIG pueden clasificarse en seis grandes categorías, de las cuales se destacan cinco por su aplicación directa en trabajos de análisis territorial, planificación, y evaluación de riesgo, como el desarrollado en esta investigación.

a. Localización directa

Esta es una de las funciones más básicas pero esenciales de un SIG. Consiste en obtener información temática sobre un punto o entidad geográfica específica. Al hacer clic sobre un punto del mapa, el sistema puede desplegar atributos asociados, como tipo de suelo, uso del terreno, altitud, densidad poblacional, infraestructura cercana, entre otros. Esta funcionalidad permite responder a la pregunta: ¿Qué hay en este lugar? y es clave para estudios de inventarios, catastros, monitoreo ambiental y gestión de servicios públicos (Longley et al., 2015).

b. Localización condicionada

También conocida como consulta espacial por atributos, esta función permite identificar zonas que cumplen con uno o varios criterios definidos. Por ejemplo, si se requiere encontrar áreas con pendiente menor al 15%, altitud inferior a 2000 msnm y suelos del tipo andosol, un SIG puede superponer capas temáticas y aplicar filtros lógicos para delimitar las zonas que cumplan simultáneamente con esas condiciones. Esta herramienta es ampliamente usada en ordenamiento territorial, selección de sitios para infraestructuras, zonas de conservación o instalación de servicios (K.T. Chang, 2018).

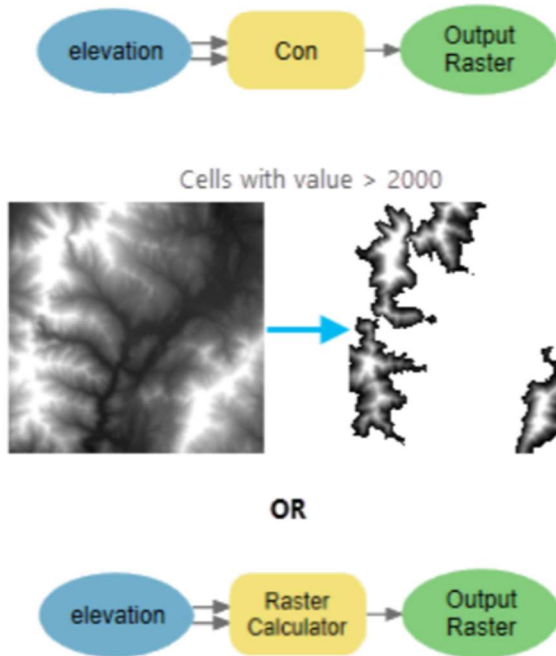


Figura 24. Localización condicionada usando parámetros de elevación con ArcGIS. (Fuente: ESRI - ArcGIS Pro).

c. Análisis de tendencias

Los SIG permiten comparar el mismo territorio en diferentes momentos del tiempo, facilitando estudios de evolución espacial. Al superponer imágenes satelitales o datos geospaciales históricos, es posible analizar cómo han cambiado variables como la cobertura vegetal, el uso del suelo, la expansión urbana o la incidencia de eventos naturales. Este tipo de análisis temporal es esencial para comprender dinámicas territoriales, modelar escenarios futuros o evaluar el impacto de políticas públicas y fenómenos climáticos (Gorelick et al., 2017).

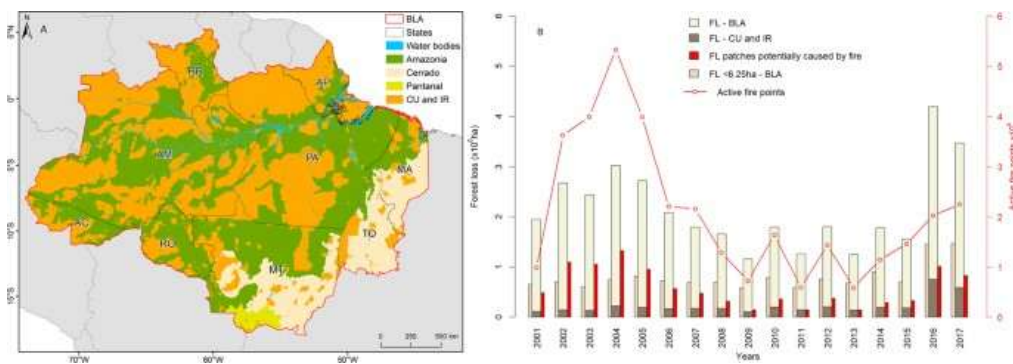


Figura 25. Uso de GIS para análisis en tendencias de deforestación en Brasil. (Fuente: Montbeller, C., et al).

d. Evaluación de rutas

Mediante herramientas de análisis espacial, los SIG permiten realizar análisis de redes para encontrar rutas óptimas entre dos o más puntos. Estos análisis toman en cuenta variables como distancia, pendiente, tipo de superficie, costos económicos o impacto ambiental. Son fundamentales en la planificación de caminos, líneas de transmisión, rutas de evacuación, distribución logística, entre otros. El análisis puede incluir obstáculos naturales, zonas restringidas o preferenciales, ofreciendo alternativas optimizadas en función del objetivo del estudio (ESRI, 2021).

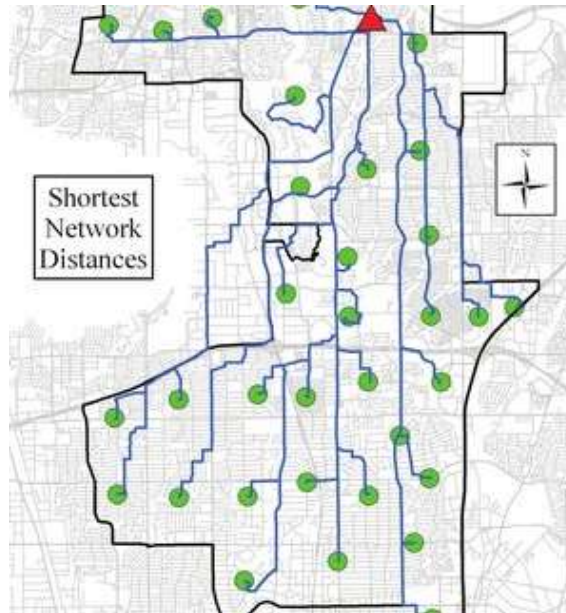


Figura 26. Análisis de redes y enrutamiento vehicular en GIS.
(Fuente: Curtin, K. 2007).

e. Modelos espaciales

Los SIG permiten crear modelos espaciales que representan fenómenos complejos del territorio, como la dispersión de contaminantes, el riesgo de inundación o deslizamientos, o la accesibilidad a servicios básicos. Estos modelos utilizan datos reales y algoritmos de simulación para generar representaciones digitales de procesos naturales o sociales, los cuales pueden ser manipulados para evaluar distintos escenarios. El modelado espacial es clave en la gestión del riesgo, el urbanismo predictivo y la planificación estratégica (Negrete, 2023).

3. Modelos de datos espaciales

Los Sistemas de Información Geográfica (SIG) utilizan principalmente dos modelos para representar datos espaciales: el modelo ráster y el modelo vectorial. Ambos tienen aplicaciones específicas según el tipo de análisis geográfico que se pretenda realizar, y su elección depende de la naturaleza del fenómeno que se desea representar, así como del nivel de detalle requerido.

a. Mapa ráster

El modelo de datos ráster representa el espacio geográfico mediante una matriz de celdas (píxeles) dispuestas en filas y columnas. Cada celda contiene un valor numérico que representa una característica específica de la superficie terrestre en esa ubicación, como precipitación, altitud, temperatura, humedad relativa, radiación solar, entre otros parámetros ambientales. Este modelo es ampliamente utilizado en análisis de variables continuas, imágenes satelitales y modelos digitales de elevación (DEM) (Olaya, 2014).

Una de las características del modelo ráster es que no almacena explícitamente las coordenadas de cada celda. En su lugar, se define una referencia espacial para la esquina superior izquierda, la resolución del píxel y la proyección, lo que permite inferir las coordenadas de cualquier celda a partir de su posición relativa dentro de la matriz. Esta estructura simplificada lo hace eficiente para realizar operaciones matemáticas, como superposición, cálculos por píxel y análisis espaciales de tipo zonal o de distancia (Pucha, 2021).

Sin embargo, el modelo ráster tiene limitaciones en la representación precisa de objetos definidos, como límites administrativos, parcelas o redes de infraestructura, ya que, al depender de una rejilla fija, puede producir generalizaciones y pérdida de detalle en bordes y formas irregulares.

b. Mapa vectorial

El modelo vectorial, en contraste, representa los objetos geográficos como entidades discretas mediante tres primitivas geométricas básicas: puntos, líneas y polígonos. Cada entidad está asociada a una tabla de atributos que describe sus características, permitiendo representar no solo su forma y ubicación, sino también información temática como nombre, clasificación, altura, uso del suelo, entre otros (Aliaga, 2017; Argentina, 2023).

Este modelo parte de la premisa de que la superficie terrestre está compuesta por objetos definidos, como caminos, lagos, edificios o árboles, y que es posible modelar el espacio mediante sus límites exactos. Por ejemplo, un río puede representarse como una línea, una parcela como un polígono y una torre de comunicación como un punto. El modelo vectorial resulta ideal para estudios de planificación urbana, redes de transporte, análisis de proximidad o gestión catastral, ya que permite realizar consultas espaciales complejas, como operaciones de topología, uniones espaciales y análisis de cobertura (Pucha, 2021).

A diferencia del modelo ráster, el vectorial permite una mayor precisión en la representación de formas y límites, aunque su manipulación puede requerir mayor capacidad computacional cuando se trabaja con grandes volúmenes de datos o geometrías complejas.

Ambos modelos no son excluyentes, sino complementarios. Muchos proyectos SIG exitosos integran datos ráster y vectoriales en sus análisis, utilizando la fuerza de cada uno según las necesidades del estudio.

4. Software de SIG

El uso de software especializado es fundamental para el desarrollo, análisis y visualización de información geoespacial en proyectos de ingeniería, medio ambiente y planificación territorial. Los Sistemas de Información Geográfica permiten integrar múltiples capas de datos vectoriales y ráster para generar análisis espaciales complejos. Existen diversas herramientas, tanto de código abierto como comerciales, que se han consolidado como referentes en el campo de los SIG.

a. QGIS

QGIS (Quantum GIS) es un software de código abierto ampliamente utilizado por su flexibilidad, comunidad activa y capacidad de integración con herramientas como GRASS, SAGA y Python. Es una opción accesible para usuarios de todo nivel, permitiendo realizar análisis geoespaciales avanzados sin costos de licencia. Entre sus fortalezas destacan el modelado geográfico, el procesamiento por lotes y la visualización cartográfica personalizada (Olaya, 2014).

b. ArcGIS

ArcGIS, desarrollado por ESRI, es una de las plataformas comerciales más robustas del mercado. Ofrece herramientas avanzadas para análisis espacial, geoprocésamiento, gestión de bases de datos geográficas (Geodatabases) y aplicaciones web. Su ecosistema incluye ArcMap, ArcGIS Pro y ArcGIS Online, integrándose con servicios de nube y soluciones empresariales. Es especialmente útil en instituciones gubernamentales y proyectos de gran escala (Longley et al., 2015; Environmental Systems Research Institute (ESRI), 2022).

c. Google Earth Engine (GEE)

Google Earth Engine es una plataforma en la nube diseñada para el análisis masivo de datos satelitales. Su acceso a colecciones como Landsat, Sentinel y CHIRPS permite extraer variables ambientales históricas, ideal para estudios sobre cobertura vegetal, hidrología y riesgo. Gracias a su lenguaje basado en JavaScript o Python, GEE facilita la automatización de procesos y la generación de productos espaciales sin necesidad de infraestructura local (Gorelick et al., 2017).

5. Archivos compatibles con SIG

Los Sistemas de Información Geográfica (SIG) manejan diversos tipos de archivos que permiten representar datos espaciales (ubicación y forma de elementos geográficos) y atributos temáticos (información asociada). Estos archivos se dividen generalmente en tres grandes grupos: vectoriales, ráster y archivos auxiliares o de configuración (Burrough & McDonnell, 1998; Longley et al., 2015).

a. Archivos vectoriales

Representan objetos geográficos mediante puntos, líneas y polígonos. Son útiles para representar límites administrativos, carreteras, edificaciones, redes, etc.

- `.shp`, `.shx`, `.dbf`, `.prj` — Formato Shapefile de ESRI, uno de los más utilizados a nivel mundial para vectores simples, aunque requiere múltiples archivos complementarios (Environmental Systems Research Institute (ESRI), 2022).
- `.geojson` — Formato basado en JSON, ampliamente usado en aplicaciones web geoespaciales por su compatibilidad con APIs y navegadores modernos (Gorelick et al., 2017).
- `.kml` / `.kmz` — Diseñados inicialmente para Google Earth, permiten la visualización de datos geográficos en entornos interactivos y web.
- `.gpkg` — GeoPackage, formato moderno basado en SQLite, recomendado por el Open Geospatial Consortium (OGC) para la interoperabilidad de datos vectoriales y ráster en un solo archivo ((OGC), 2021).

b. Archivos ráster

Representan datos espaciales continuos como imágenes, modelos de elevación, temperatura, cobertura del suelo, etc., mediante una matriz de píxeles

- `.tif` / `.tiff` — GeoTIFF, formato que almacena imágenes georreferenciadas y metadatos espaciales incrustados.
- `.asc` — ASCII Grid, usado frecuentemente para modelos digitales de terreno y análisis hidrológicos (Maguire et al., 2005)
- `.img` — Formato de ERDAS Imagine, común en procesamiento de imágenes satelitales.

c. Archivos auxiliares o de proyecto

- `.qgz` / `.qgs` — Formatos de proyectos QGIS que guardan configuraciones de visualización y vínculos a archivos fuente (Team, 2023).
- `.lyr` — Archivos de capa de ArcGIS, utilizados para almacenar la simbología de capas sin alterar los datos originales.
- `.xml`, `.aux.xml` — Archivos de metadatos y configuraciones asociadas a capas ráster o vectoriales.

La correcta identificación y uso de estos formatos es esencial para asegurar la interoperabilidad entre software GIS, preservar la calidad de los datos y facilitar el análisis geoespacial eficiente (Burrough & McDonnell, 1998; Longley et al., 2015).

6. Sistemas de Coordenadas de Referencia (SRC)

Los Sistemas de Coordenadas de Referencia (SCR) son fundamentales en los Sistemas de Información Geográfica (SIG), ya que permiten ubicar con precisión elementos geoespaciales sobre la superficie terrestre. Un SCR define cómo se representan las coordenadas (como latitud y longitud o coordenadas planas X, Y) en relación con una superficie terrestre, a partir de un datum geodésico, una proyección cartográfica y una unidad de medida. Existen SCR globales,

como WGS 84, y otros locales adaptados a regiones específicas para mejorar la precisión geométrica (K.-T. Chang, 2019).

Cada sistema posee un código EPSG (European Petroleum Survey Group), que lo identifica de forma única. Por ejemplo, el EPSG:4326 corresponde al sistema WGS 84 que usa coordenadas geográficas (grados decimales), mientras que el EPSG:3857 es una proyección métrica utilizada ampliamente en servicios web como Google Maps. Para análisis más precisos a nivel local, se utilizan sistemas proyectados como el EPSG:32615 (UTM zona 15N con datum WGS 84), ideal para regiones como Guatemala, ya que permite trabajar con distancias y áreas en metros (EPSG Geodetic Parameter Registry, 2024; Hofmann-Wellenhof et al., 2011).

Código EPSG	Nombre del sistema	Descripción	Rango típico
4326	WGS 84	Sistema global basado en coordenadas geográficas (latitud/longitud en grados).	Grados: Latitud: -90° a 90° Longitud: -180° a 180°
3857	WGS 84 / Pseudo-Mercator	Proyección métrica usada en mapas web; distorsiona áreas cerca de los polos.	Metros: X: aprox. -20m a 20m Y: aprox. -85° a 85°
32615	WGS 84 / UTM zona 15N	Proyección UTM adecuada para Guatemala; coordenadas en metros, útil para análisis de precisión local.	Metros X: 166,021 a 833,978 Y: aprox. 0 a 10,000,000

Tabla 7. Ejemplos de Sistemas de Coordenadas de Referencia (SCR).
(Fuente: Elaboración propia).

7. Aplicaciones directas de los sistemas de información geográfica

a. Infraestructura

Los Sistemas de Información Geográfica (SIG) se han consolidado como una herramienta indispensable en la gestión de infraestructura urbana y rural, ya que permiten almacenar, visualizar y analizar datos espaciales y alfanuméricos de forma integrada. Esta capacidad es especialmente útil en sectores como las redes de servicios públicos —electricidad, agua potable, alcantarillado, gas, telecomunicaciones— donde se requiere precisión tanto en la ubicación física de los elementos como en su estado funcional. A través de los SIG es posible generar modelos de redes subterráneas o aéreas, planificar su expansión, detectar zonas críticas y programar mantenimientos preventivos. Además, la integración de los SIG con sensores remotos y sistemas SCADA ha permitido avanzar hacia esquemas de infraestructura inteligente (Lo & Yeung, 2007).

b. Gestión territorial

En el campo de la planificación territorial, los SIG permiten elaborar catálogos georreferenciados de la infraestructura existente, áreas urbanas, zonas agrícolas, forestales o de conservación, así como delimitar polígonos de uso del suelo, propiedad y regulación catastral. Gracias a su capacidad de sobreponer capas temáticas, los SIG facilitan la evaluación multicriterio de terrenos, ayudando a tomar decisiones sobre ordenamiento territorial, zonificación urbana, y evaluación de conflictos de uso del suelo. Esta tecnología permite visualizar dinámicamente el crecimiento urbano, la expansión de lotificaciones y la interacción entre infraestructura y medio ambiente (Maguire et al., 2005).

c. Ambiente

En el ámbito ambiental, los SIG se utilizan ampliamente para la monitorización, modelado y gestión de los recursos naturales. Se aplican en el seguimiento de procesos como la deforestación, la erosión de suelos, la contaminación hídrica, el cambio de cobertura vegetal y los incendios forestales. A través de la combinación de imágenes satelitales y sensores remotos con bases de datos históricas, se pueden realizar análisis espaciales para identificar tendencias ecológicas y proyecciones de deterioro ambiental. Además, se emplean para delimitar zonas de conservación, áreas protegidas y corredores ecológicos, así como para evaluar el impacto ambiental de proyectos de infraestructura mediante modelos de simulación (Burrough & McDonnell, 1998).

d. Equipamiento social:

Los SIG permiten mejorar la gestión del equipamiento social al identificar la ubicación, cobertura y accesibilidad de servicios esenciales como hospitales, escuelas, centros comunitarios, mercados, etc. Al integrar variables sociodemográficas como densidad de población, índices de pobreza o necesidades básicas insatisfechas, se puede optimizar la planificación de nuevos equipamientos, focalizando intervenciones en zonas de mayor vulnerabilidad. De igual forma, permiten simular escenarios de crecimiento poblacional o reasignación de servicios en contextos de reubicación o emergencia, apoyando políticas públicas más equitativas y sostenibles (Longley et al., 2015).

e. Tráfico:

En el campo del transporte y movilidad urbana, los SIG son una herramienta fundamental para la modelación de flujos vehiculares y peatonales. A través del análisis espacial de tramos viales, intersecciones y patrones de uso, se pueden identificar cuellos de botella, zonas de congestión y rutas óptimas. También son utilizados para planificar rutas de transporte público, analizar tiempos de recorrido, e implementar sistemas inteligentes de movilidad. Combinando los SIG con datos en tiempo real provenientes de sensores o GPS, es posible diseñar sistemas de control adaptativo del tráfico, simulaciones de evacuación y planes de contingencia ante desastres (Zheng et al., 2011).

8. Georreferenciación:

La georreferenciación es la práctica que permite relacionar la información de la posición entre documentos o archivos cartográficos con procedencias distintas. Este proceso permite determinar la posición de un elemento en un sistema de coordenadas espaciales diferente al que se encuentra (Dávila Martínez & Camacho Arranz, 2012). Este proceso es esencial para integrar y comparar información geográfica que puede haberse obtenido mediante distintos métodos o en diferentes momentos históricos, como mapas antiguos, imágenes de satélite, y levantamientos topográficos (Goodchild, 1993).

La precisión de este proceso depende en gran medida de la calidad de los puntos de control utilizados, que son referencias conocidas y fijas en ambos sistemas de coordenadas. La capacidad de transformar y unificar datos espaciales en un marco de referencia común es crucial para numerosas aplicaciones, como la planificación urbana, el monitoreo ambiental, y la gestión de recursos naturales (Longley et al., 2015).

A través de la georreferenciación, se puede establecer una correspondencia precisa entre las coordenadas de los elementos en diferentes sistemas, facilitando así su análisis conjunto. Esto se realiza encontrando la relación entre las posiciones de un elemento en ambos sistemas, de manera que se encuentre la posición homóloga del elemento de un sistema a otro (Londoño, 1999).

9. Modelo de los datos:

Un modelo de datos es el conjunto de reglas utilizadas para representar las diferentes entidades almacenadas en una base de datos. Las entidades y variables de los modelos de datos de SIG son los siguientes:

a. Realidad perceptible

Es el nivel propio de las personas encargadas de gestionar y planificar el espacio (Sarría, 2014).

b. Modelo conceptual

Es el nivel de científicos, como geólogos, geógrafos, ecólogos, hidrogeólogos, etc. encargados de desarrollar, verificar y aplicar teorías e hipótesis relacionadas a variables y procesos que se llevan en un espacio. Existen dos tipos fundamentales; entidades y variables (Sarría, 2014).

c. Modelo digital o estructura de datos

Desarrolladores e informáticos de SIG que optimizan las estructuras de datos para almacenar la información y ampliar el repertorio de herramientas del SIG del modelo y base de datos, de acuerdo con las necesidades de los científicos y técnicos en SIG (Sarría, 2014).

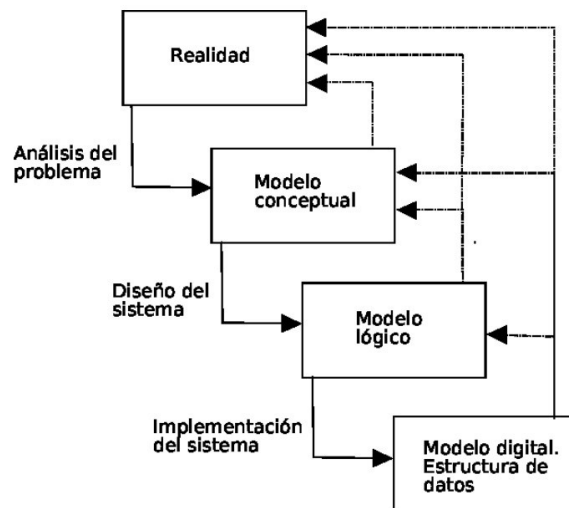


Figura 27. Esquema de la modelización de datos en SIG.
(Fuente: Sarría, F.A, s.f.)

D. Evaluación y monitoreo de deslizamientos:

1. Evaluación de riesgos:

a. Consecuencias de deslizamientos:

Los deslizamientos representan una amenaza significativa para la integridad física de las personas y para el funcionamiento de infraestructuras críticas. Según los Centros para el Control y la Prevención de Enfermedades (CDC), los impactos más inmediatos se relacionan con lesiones traumáticas ocasionadas por la caída súbita de rocas, árboles, o escombros. Las víctimas suelen encontrarse en viviendas ubicadas en laderas inestables, en caminos rurales o al pie de taludes. En particular, los flujos de escombros (debris flows), caracterizados por su alta velocidad y densidad, pueden arrastrar rocas de tamaño considerable, árboles y estructuras livianas, causando heridas graves o incluso la muerte a personas atrapadas en su trayectoria (Center for Disease Control (CDC), 2022).

Además del riesgo directo sobre la vida humana, los deslizamientos tienen la capacidad de interrumpir redes de infraestructura crítica como carreteras, puentes, líneas eléctricas y sistemas de alcantarillado. Estas interrupciones pueden limitar el acceso a servicios médicos, cortar el suministro de agua potable o electricidad, y dificultar las labores de rescate y evacuación. Las fallas en redes sanitarias también pueden dar lugar a contaminación de fuentes hídricas y brotes de enfermedades infecciosas, sobre todo en contextos rurales con baja capacidad de respuesta institucional. Por estas razones, el impacto de un deslizamiento no se limita al evento inicial, sino que puede extenderse por días o semanas a través de efectos colaterales (Guzzetti, Cardinali et al., 2005).

b. Impacto socioeconómico:

Desde la perspectiva de la evaluación de desastres, la Comisión Económica para América Latina y el Caribe (CEPAL) ha establecido una metodología ampliamente utilizada para

cuantificar los impactos en tres niveles: daños directos, daños indirectos y efectos macroeconómicos. Los daños directos incluyen la pérdida o destrucción de bienes físicos — viviendas, infraestructura vial, redes de servicios— y la afectación de tierras productivas, que reducen el capital tangible disponible para las personas, empresas o instituciones (para América Latina y el Caribe (CEPAL), 1991).

Los daños indirectos se reflejan en la interrupción de actividades económicas, como la agricultura, el comercio o el transporte. Por ejemplo, un deslizamiento que bloquea una carretera principal puede impedir la comercialización de productos agrícolas, reducir los ingresos de las familias y elevar los costos de transporte. Asimismo, los recursos que deben ser destinados a reparaciones, atención de emergencia y reubicación implican un gasto no programado para los gobiernos y las comunidades afectadas.

Finalmente, los efectos secundarios o impactos a nivel macroeconómico pueden incluir un incremento en el gasto público, desbalances fiscales, contracción de la inversión y reducción del crecimiento económico nacional. En países de ingresos medios y bajos, estos efectos pueden amplificar desigualdades estructurales, generar migraciones forzadas y comprometer la sostenibilidad de proyectos de desarrollo local. En este contexto, los deslizamientos deben ser entendidos no solo como fenómenos naturales, sino como detonantes de crisis multidimensionales, cuya magnitud depende de la vulnerabilidad previa y la capacidad de respuesta institucional (Fernández & Mora, 2022).

2. Modelos de predicción de deslizamientos:

Los modelos de predicción se pueden emplear para cualquier tipo de amenaza o evento natural, que se pueda relacionar a una condición o parámetro. Tomando en cuenta que los deslizamientos son eventos provocados por una serie de factores, esto indica que se puede determinar la posibilidad de ocurrencia de un deslizamiento con base en un criterio preestablecido (Guzzetti, Reichenbach et al., 2005).

El criterio para determinar la prevalencia a deslizamiento de un punto en el espacio relaciona aspectos como el tipo de suelo y sus características mecánicas, pendiente del suelo (topografía), y otros factores accidentales, como la sismicidad y la precipitación, que no se pueden predecir ni en cuanto a su posición en el tiempo, ni su intensidad o fuerza (Guzzetti, Reichenbach et al., 2005).

Los modelos de propensión a deslizamientos, utilizando las bases de datos de los SIG, son capaces de determinar la población afectada por los deslizamientos. Tomando en cuenta los diferentes actores que contribuyen a los deslizamientos, se puede generar una ecuación para determinar el índice de prevalencia a un deslizamiento utilizando los valores máximos como el índice 1, obteniendo una suma de los diferentes parámetros con un factor de importancia determinado mediante un análisis de deslizamientos históricos. Utilizando mapas de inventario como de registro sísmico y pluvial, se pueden calcular los factores disparates que obtengan el índice final (Gui et al., 2023).

3. Inventarios de deslizamientos:

a. Criterios:

Los inventarios de deslizamientos son herramientas fundamentales para el estudio, la prevención y la gestión del riesgo geológico. Su elaboración parte del reconocimiento y análisis de múltiples variables físicas y ambientales que pueden incidir en la ocurrencia de estos fenómenos. Entre los principales criterios geomorfológicos, se consideran aspectos como la pendiente del terreno, la forma de las laderas, la curvatura del relieve, la orientación de la pendiente (aspect) y la altitud relativa. Estudios han demostrado que terrenos con pendientes superiores a 25° presentan una mayor susceptibilidad a deslizamientos, especialmente cuando combinan materiales inestables y condiciones de saturación (Guzzetti et al., 2012).

Los criterios geotécnicos incluyen parámetros relacionados con el tipo de suelo o roca, su litología, grado de meteorización, textura, cohesión y permeabilidad. Suelos finos con alta plasticidad y baja capacidad drenante, como las arcillas expansivas, son más propensos a la inestabilidad bajo condiciones de saturación (Highland Bobrowsky, 2008). Por su parte, los criterios climáticos más relevantes suelen estar relacionados con la precipitación acumulada en días previos, la intensidad de lluvias en corto plazo, y el contenido de humedad del suelo. Estos factores afectan la presión de poros y la resistencia al corte, desencadenando deslizamientos en zonas vulnerables.

Otro componente esencial es la información histórica y de campo. La observación de cicatrices antiguas, la documentación de eventos pasados y la interacción con comunidades locales son estrategias clave para delimitar áreas con antecedentes de inestabilidad. Estos inventarios históricos, cuando están bien georreferenciados, permiten hacer análisis espaciales de recurrencia, densidad de eventos y evolución temporal de los deslizamientos, lo que fortalece los modelos de pronóstico y zonificación de riesgo (Van Westen et al., 2006).

b. Métodos:

Los métodos para elaborar inventarios de deslizamientos han evolucionado con el desarrollo de tecnologías geoespaciales. Tradicionalmente, se realizaban mediante observación de campo, reconocimiento visual y análisis de fotografías aéreas. Actualmente, se utilizan métodos modernos como el análisis de imágenes satelitales multiespectrales, modelos digitales de elevación (DEM), y técnicas de teledetección y SIG (Sistemas de Información Geográfica), los cuales permiten identificar patrones morfológicos y cambios en la superficie terrestre con alta precisión (Galli et al., 2008).

Existen tres tipos principales de inventarios:

- **Eventuales:** Registran deslizamientos ocurridos en un solo evento desencadenante.
- **Históricos:** Agrupan eventos a lo largo del tiempo mediante fuentes documentales, registros de prensa o informes técnicos.

- **Geomorfológicos:** Interpretan las formas del terreno para identificar cicatrices de deslizamientos antiguos, incluso si no están activos actualmente.

Todos estos enfoques pueden ser integrados para crear inventarios multitemporales, más completos y útiles para estudios de susceptibilidad o riesgo (Guzzetti et al., 2012).

A pesar del avance tecnológico, la interpretación visual sigue siendo un componente importante en el proceso de inventario. El criterio del evaluador, su conocimiento del área y su experiencia influyen directamente en la precisión y calidad del producto final. No existen protocolos normativos universales para la elaboración de inventarios; sin embargo, algunos organismos internacionales como el USGS o la FAO han propuesto guías metodológicas para unificar criterios y facilitar la comparación entre estudios (Highland Bobrowsky, 2008). Los mejores inventarios son aquellos que combinan fuentes diversas, validan la información mediante trabajo de campo, y son mantenidos y actualizados en plataformas accesibles para su uso en planificación territorial y prevención de desastres.

E. Fundamentos del aprendizaje automático:

1. Machine Learning:

El aprendizaje automático Machine Learning es una subdisciplina de la inteligencia artificial que se enfoca en el desarrollo de algoritmos capaces de aprender a partir de los datos, con el objetivo de realizar predicciones o tomar decisiones sin necesidad de ser programados explícitamente para cada tarea específica (Bishop, 2006). A diferencia de los sistemas tradicionales basados en reglas definidas manualmente, el aprendizaje automático extrae automáticamente patrones relevantes del conjunto de datos y adapta su comportamiento con base en la experiencia previa.

Este enfoque es especialmente útil en contextos donde los fenómenos a modelar son complejos, multivariados y no lineales, como ocurre en muchos procesos ambientales y geográficos (Vieira, Lacerda et al., 2022). En particular, los sistemas de información geográfica (SIG) y plataformas como Google Earth Engine generan grandes volúmenes de datos espaciales y temporales, cuya complejidad hace difícil su análisis mediante técnicas estadísticas convencionales. El Machine Learning permite aprovechar esta riqueza de datos para construir modelos predictivos que identifican relaciones ocultas entre múltiples factores condicionantes.

Los algoritmos de aprendizaje automático se pueden clasificar en tres categorías principales: aprendizaje supervisado, no supervisado y por refuerzo (Goodfellow et al., 2016). En el aprendizaje supervisado, el modelo se entrena a partir de un conjunto de datos etiquetado, donde cada observación está asociada a una salida conocida, permitiendo construir una función de predicción. En el aprendizaje no supervisado, por el contrario, se analizan datos no etiquetados para encontrar estructuras subyacentes, como agrupamientos o patrones de correlación. El aprendizaje por refuerzo, finalmente, implica que un agente tome decisiones secuenciales optimizando una función de recompensa.

2. Aprendizaje supervisado:

El aprendizaje supervisado es una rama del aprendizaje automático que se basa en el uso de datos etiquetados, es decir, conjuntos de entrada que ya cuenta con una salida conocida. El objetivo principal de este enfoque es entrenar modelos para que sean capaces de predecir resultados o clasificar nuevas entradas con base en patrones aprendidos a partir de ejemplos previos. A diferencia del aprendizaje no supervisado, aquí el modelo cuenta con una “verdad” o etiqueta que le permite ajustar sus parámetros durante el entrenamiento para minimizar el error entre la predicción y el resultado esperado (Hastie, et al., 2009).

Las técnicas más comunes en el aprendizaje supervisado incluyen la regresión (lineal y logística), los árboles de decisión, las máquinas de soporte vectorial (SVM) y las redes neuronales. Estos algoritmos se utilizan ampliamente en tareas como clasificación de texto, reconocimiento de imágenes, predicción de precios y diagnóstico médico (Murphy, 2012). El proceso de entrenamiento implica alimentar al modelo con pares de entrada-salida, permitiéndole aprender la relación entre las características y las etiquetas, lo que facilita la generalización a nuevos datos.

Una de las principales ventajas del aprendizaje supervisado es su capacidad para generar modelos precisos cuando se dispone de datos etiquetados de calidad. Sin embargo, también presenta desafíos importantes, como la necesidad de grandes volúmenes de datos etiquetados, lo cual puede ser costoso o inviable en ciertos contextos. A pesar de ello, este enfoque ha demostrado ser altamente efectivo en aplicaciones prácticas como la detección de fraudes, la segmentación de clientes y la predicción de eventos futuros (Aggarwal, 2014).

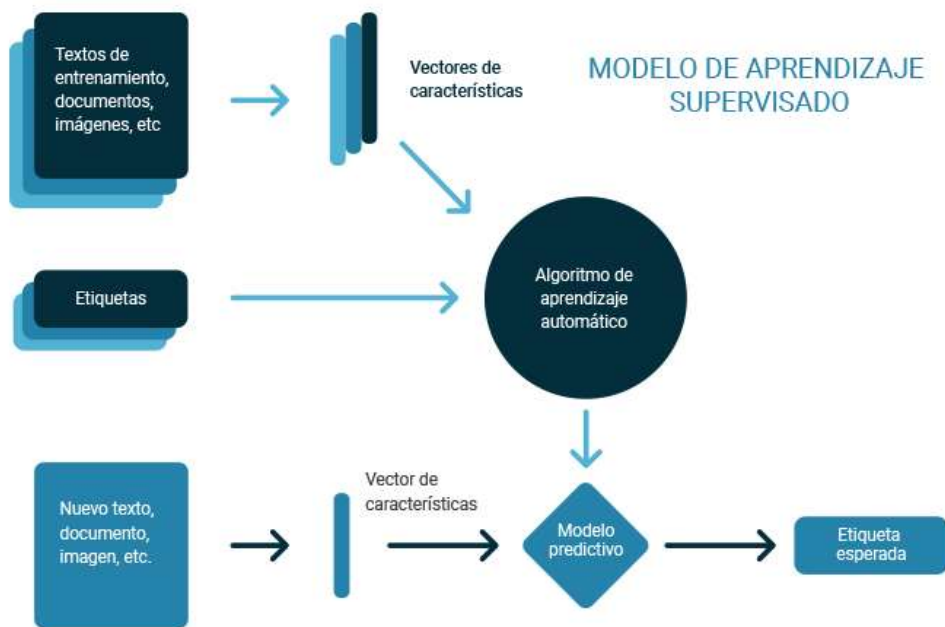


Figura 28. Diagrama de flujo de aprendizaje supervisado.
(Fuente: Gonzales, J. 2018).

3. Aprendizaje no supervisado:

El aprendizaje no supervisado es una rama del aprendizaje automático que trabaja con datos sin etiquetar, es decir, sin una salida conocida asociada a cada entrada. El objetivo principal de este enfoque es descubrir estructuras ocultas, patrones o relaciones dentro de los datos, sin intervención humana directa. A diferencia del aprendizaje supervisado, aquí el modelo no cuenta con una respuesta correcta con la cual comparar su desempeño durante el entrenamiento (Hastie et al., 2009).

Las técnicas más comunes en el aprendizaje no supervisado incluyen el análisis de conglomerados (clustering), la reducción de dimensionalidad y la detección de anomalías. Algoritmos como k-means, DBSCAN o el análisis de componentes principales (PCA) son ampliamente utilizados en contextos donde se necesita explorar grandes volúmenes de datos para identificar agrupaciones naturales o simplificar su representación (Murphy, 2012). Este tipo de aprendizaje es especialmente valioso en estudios exploratorios o cuando se desea generar hipótesis a partir de datos complejos y sin estructura aparente.

Una de las principales ventajas del aprendizaje no supervisado es su capacidad para adaptarse a datos nuevos y sin etiquetar, lo que lo hace útil en aplicaciones donde el etiquetado manual sería costoso o inviable. Sin embargo, también presenta desafíos importantes, como la dificultad para evaluar objetivamente el rendimiento del modelo y la interpretación de los resultados. A pesar de ello, este enfoque ha demostrado ser útil en áreas como segmentación de imágenes, agrupamiento de comportamientos, y estudios geoespaciales donde se busca identificar patrones naturales sin conocimiento previo del terreno (Aggarwal, 2014).

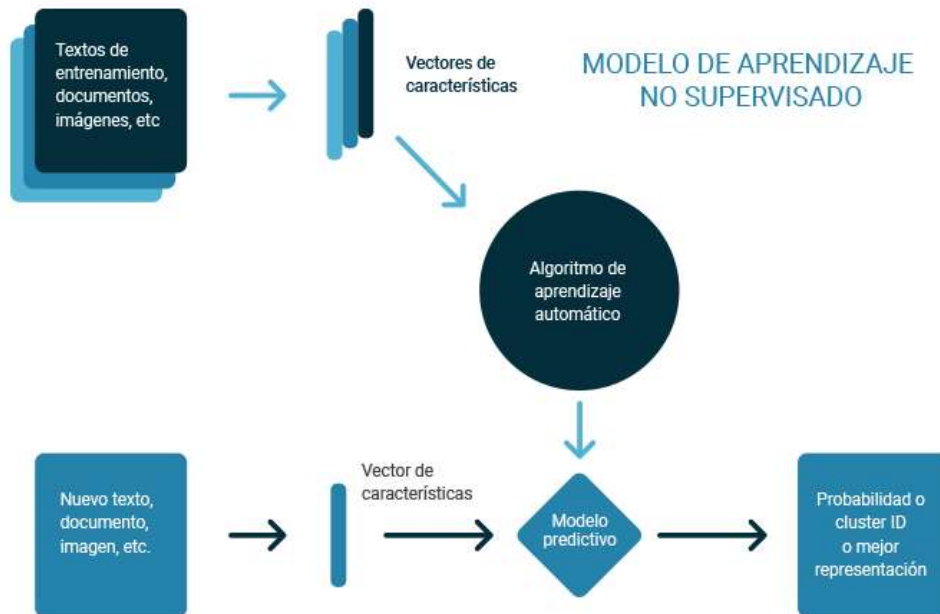


Figura 29. Diagrama de flujo de aprendizaje no supervisado. (Fuente: Gonzales, J. 2018).

4. Random Forest:

a. Concepto:

Random Forest es un algoritmo de clasificación y regresión basado en árboles de decisión que pertenece a la categoría de métodos ensemble. Su funcionamiento se basa en la combinación de múltiples árboles construidos a partir de subconjuntos aleatorios del conjunto de datos, reduciendo así la varianza del modelo y aumentando su estabilidad (Breiman, 2001).

b. Funcionamiento:

Utiliza la técnica de bagging (bootstrap aggregating), en la cual se entrenan múltiples árboles en paralelo, cada uno con una muestra aleatoria del dataset. En el caso de clasificación, cada árbol vota por una clase, y la clase final es la que reciba más votos.

c. Categoría de aprendizaje automático:

Pertenece al aprendizaje supervisado, ya que utiliza datos etiquetados (clases conocidas) para entrenar el modelo y luego hacer predicciones sobre nuevos datos.

d. Modelos y parámetros:

- **n_estimators (Número de árboles):** Indica cuántos árboles de decisión se generarán dentro del bosque. Un mayor número de árboles mejora la estabilidad del modelo y su precisión, ya que reduce la varianza, pero también incrementa el tiempo de cómputo. Generalmente, se selecciona un valor alto (como 100 o más) para obtener predicciones más robustas. Sin embargo, después de cierto punto, agregar más árboles ofrece beneficios marginales.
- **max_depth (Profundidad máxima de árbol):** Define cuántos niveles puede tener cada árbol de decisión. Si se permite una profundidad ilimitada, los árboles pueden ajustarse excesivamente al conjunto de entrenamiento, provocando sobreajuste. Establecer una profundidad máxima ayuda a controlar la complejidad del modelo y mejora su capacidad de generalización. Es un parámetro clave en el balance entre sesgo y varianza. Geomorfológicos: Interpretan las formas del terreno para identificar cicatrices de deslizamientos antiguos, incluso si no están activos actualmente.
- **min_samples_leaf (Muestras mínimas por hoja):** Especifica el número mínimo de muestras requeridas para definir una hoja (nodo terminal) en un árbol. Aumentar este valor obliga a que las hojas contengan más observaciones, lo cual simplifica los árboles y reduce el riesgo de sobreajuste. Es útil para regular la fragmentación del conjunto de datos en nodos poco representativos. También contribuye a modelos más interpretables.
- **random_state (Semilla aleatoria):** Controla la aleatoriedad del muestreo de datos y de las divisiones de nodos. Al fijar un valor, se garantiza la reproducibilidad de los resultados del modelo, permitiendo que los mismos datos y parámetros generen siempre los mismos árboles. Es esencial para fines de validación, documentación y comparación entre experimentos. No afecta directamente el rendimiento, pero sí su estabilidad.

Estos parámetros influyen directamente en la capacidad de generalización del modelo, y su ajuste adecuado es fundamental para obtener resultados confiables sin incurrir en sobreajuste (Liaw & Wiener, 2002).

e. Función objetivo:

La función objetivo en Random Forest es implícita, ya que no optimiza una función de pérdida en conjunto, sino que cada árbol se entrena de forma independiente con el objetivo de minimizar el error en su conjunto bootstrap. Posteriormente, los resultados se agregan mediante votación (clasificación) o promedio (regresión).

f. Conjunto de árboles de decisión:

Cada árbol en el bosque es un clasificador débil, pero su combinación logra un clasificador robusto. Los árboles son no podados (crecen hasta que se cumple un criterio de parada), lo que permite capturar patrones complejos, pero también podría llevar al sobreajuste, mitigado por el promedio entre árboles.

g. Complejidad del modelo:

Aunque computacionalmente más costoso que un solo árbol, Random Forest es altamente paralelizable y escalable. Su complejidad depende del número de árboles y de la profundidad de cada uno, lo cual se puede ajustar para mejorar rendimiento y tiempos de cómputo (Breiman, 2001).

h. Interpretabilidad y utilidad práctica:

Random Forest proporciona una métrica de importancia de variables, útil para identificar los predictores más influyentes. Este aspecto es especialmente valioso en aplicaciones de geociencias y planificación territorial (Liaw & Wiener, 2002).

5. Algoritmo - XGBoost:

a. Concepto:

XGBoost (Extreme Gradient Boosting) es un algoritmo de tipo boosting basado en árboles de decisión. Se caracteriza por su alto rendimiento, escalabilidad y precisión, y ha sido ampliamente utilizado en competiciones de ciencia de datos y en aplicaciones con datos complejos o de alta dimensionalidad (Chen & Guestrin, 2016).

b. Funcionamiento:

A diferencia de Random Forest, los árboles se construyen de forma secuencial. Cada nuevo árbol se entrena para corregir los errores del anterior, ajustándose a los residuos de predicción. Este enfoque permite una convergencia más rápida y un aprendizaje más fino.

c. Categoría de aprendizaje automático:

Como Random Forest, pertenece al aprendizaje supervisado. Utiliza una función de pérdida que guía el entrenamiento del modelo a lo largo de cada iteración.

d. Modelos y parámetros:

- **n_estimators (Número de árboles):** Este parámetro define cuántos árboles de decisión se construyen secuencialmente durante el proceso de entrenamiento. A mayor número de árboles, el modelo tiene mayor capacidad para aprender patrones complejos. Sin embargo, un número excesivo puede llevar al sobreajuste si no se acompaña de técnicas de regularización adecuadas. En la práctica, se busca un equilibrio entre rendimiento y eficiencia computacional.
- **[Taza de aprendizaje]:** Representa el peso que se le asigna a cada nuevo árbol durante el ensamblaje. Valores más bajos hacen que el modelo aprenda más lentamente, lo cual puede mejorar la generalización y reducir el riesgo de sobreajuste. No obstante, una tasa de aprendizaje muy baja puede requerir más árboles y aumentar el tiempo de entrenamiento. Se recomienda ajustar este parámetro junto con n_estimators para obtener un desempeño óptimo.
- **max_depth (Profundidad máxima):** Controla la complejidad individual de cada árbol de decisión. Un valor alto permite capturar interacciones complejas entre variables, pero también puede causar sobreajuste si no se combina con regularización. Por el contrario, árboles poco profundos generalizan mejor, aunque con menor capacidad de ajuste. Este parámetro actúa como barrera directa frente a la complejidad del modelo.
- **booster (Tipo de algoritmo base):** Define el tipo de modelo de boosting que se utilizará: puede ser gbtree (árboles), gblinear (regresión lineal) o dart (una variante con dropout en árboles). En problemas no lineales y con interacciones entre variables, el booster gbtree suele ser el más utilizado. La elección del booster influye en la forma en que se construyen los árboles y en la estrategia de optimización. Este parámetro afecta tanto la precisión como el tiempo de entrenamiento.
- **lambda (Regularización L2):** Este término penaliza la magnitud de los coeficientes en los nodos de decisión para reducir la complejidad del modelo. La regularización L2 ayuda a prevenir el sobreajuste, especialmente cuando se manejan muchos predictores correlacionados. Un valor alto de lambda hace que el modelo sea más conservador, favoreciendo árboles más simples. Es útil en escenarios con ruido o gran dimensionalidad de variables.
- **alpha (Regularización L1):** Similar a lambda, pero penaliza el número de variables utilizadas en las divisiones de los árboles. La regularización L1 tiende a hacer un modelo más disperso, forzando a que algunos pesos sean exactamente cero, lo que puede interpretarse como una forma de selección automática de variables. Es especialmente útil cuando se sospecha que no todas las variables son relevantes. Combinada con L2, contribuye a controlar el sobreajuste de manera más robusta

(Chen & Guestrin, 2016).

e. Función objetivo:

XGBoost optimiza una función objetivo compuesta por dos partes: una función de pérdida (como log-loss o error cuadrático) y un término de regularización que penaliza la complejidad del modelo. Esta regularización es clave para su capacidad de generalización.

f. Conjunto de árboles de decisión:

XGBoost construye un conjunto de árboles de decisión de forma secuencial, donde cada árbol actúa como un modelo regresor que intenta predecir los errores (residuos) del conjunto anterior. A diferencia de métodos que generan árboles complejos y profundos, XGBoost opta por árboles más simples y específicos, los cuales contribuyen gradualmente a mejorar el modelo global a través del método de gradient boosting. Esta arquitectura permite que el modelo se enfoque en corregir de manera progresiva las observaciones mal clasificadas, optimizando una función de pérdida diferenciable en cada iteración (Chen & Guestrin, 2016; James et al., 2013). La suma de estos árboles “débiles” permite alcanzar un clasificador robusto y preciso.

g. Complejidad del modelo:

XGBoost es un modelo más complejo en términos computacionales que Random Forest, debido a su naturaleza secuencial y al hecho de que requiere la optimización de una función objetivo en cada iteración. Sin embargo, cuenta con importantes optimizaciones que lo hacen eficiente y escalable, como el uso de técnicas de poda por profundidad máxima, procesamiento paralelo de operaciones internas, y un manejo eficiente de valores faltantes mediante asignación automática en cada nodo (Chen & Guestrin, 2016; Hong & Pradhan, 2019). Además, incorpora regularización L1 y L2 de manera nativa, lo cual mejora su capacidad de generalización en contextos de alta dimensionalidad o correlación entre variables (He & Garcia, 2009; Liaw & Wiener, 2002). Gracias a estas características, XGBoost ha sido considerado uno de los algoritmos más competitivos para tareas de predicción en ciencia de datos aplicada.

h. Interpretabilidad y utilidad práctica:

XGBoost permite extraer métricas de importancia de variables, y se puede interpretar mediante técnicas como SHAP. Su robustez y flexibilidad lo hacen ideal para aplicaciones donde se requieren predicciones precisas sobre fenómenos complejos, como deslizamientos en regiones geográficamente heterogéneas (Hong & Pradhan, 2019).

6. Comparación general de algoritmos basados en árboles:

Los algoritmos basados en árboles de decisión constituyen una familia de modelos ampliamente utilizados en tareas de clasificación y regresión, particularmente en contextos donde se manejan variables de diversa naturaleza y estructuras no lineales complejas. Dentro de esta familia destacan algoritmos como Decision Trees, Random Forest y XGBoost, cada uno con fortalezas particulares en función del tipo de datos y la precisión deseada. Una diferencia fundamental entre Random Forest y XGBoost radica en la estrategia de ensamblado que utilizan. Mientras que Random Forest emplea una técnica de "bagging"(bootstrap aggregating), generando múltiples árboles en paralelo a partir de subconjuntos aleatorios del conjunto de datos,

XGBoost se basa en un enfoque de "boosting", donde los árboles se construyen de forma secuencial y cada uno intenta corregir los errores del anterior (Breiman, 2001; Chen & Guestrin, 2016). Esto hace que Random Forest tienda a ser más robusto frente al sobreajuste en conjuntos de datos ruidosos, mientras que XGBoost puede capturar relaciones complejas con mayor precisión, especialmente cuando se ajustan cuidadosamente sus hiperparámetros (Liaw & Wiener, 2002; Pheng & Thapa, 2021).

Ambos modelos pueden manejar variables categóricas y numéricas, toleran valores faltantes, y son capaces de modelar interacciones no lineales. Sin embargo, XGBoost incluye optimizaciones como el manejo eficiente de memoria, técnicas de pruning y regularización L1/L2, lo cual le proporciona una ventaja computacional y de generalización cuando se trabaja con conjuntos de datos grandes o con muchas variables (Chen & Guestrin, 2016).

7. Evaluación de modelos de clasificación:

a. Matriz de confusión:

La matriz de confusión es una herramienta fundamental para evaluar el desempeño de un modelo de clasificación. Esta matriz muestra la cantidad de verdaderos positivos (TP), falsos positivos (FP), verdaderos negativos (TN) y falsos negativos (FN), permitiendo evaluar de forma detallada los aciertos y errores del modelo. En aplicaciones de gestión de riesgo, como la predicción de deslizamientos, el análisis de falsos negativos es especialmente relevante, ya que implican no detectar zonas de alto riesgo (Fell et al., 2008; Guzzetti et al., 2005).

Además de mostrar estos valores básicos, la matriz permite calcular métricas como la precisión, el recall, la especificidad y el F1-score, lo que proporciona una evaluación más completa del rendimiento del modelo. Por ejemplo, la precisión evalúa qué proporción de las predicciones positivas fueron correctas, mientras que el recall indica qué proporción de los eventos reales fueron correctamente identificados. El F1-score, por su parte, combina precisión y recall en una sola métrica armónica (Chicco & Jurman, 2020). Estas métricas derivadas se calculan directamente a partir de los valores de la matriz, por lo que su correcta interpretación depende del contexto y los objetivos del modelo.



Figura 30. Matriz de confusión de modelos de aprendizaje automático.
(Fuente: Barrios, J.I. 2019).

b. Medidas de evaluación de desempeño:

La precisión (precision) indica la proporción de predicciones positivas que fueron correctas ($TP / (TP + FP)$), mientras que el recall mide la capacidad del modelo para identificar todos los verdaderos positivos ($TP / (TP + FN)$). El F1-score, como media armónica entre precisión y recall, ofrece una métrica balanceada que es especialmente útil cuando existe un desbalance entre clases, como es común en fenómenos naturales (Hong & Pradhan, 2019; Saito & Rehmsmeier, 2015).

c. Accuracy

La precisión general (accuracy) es una métrica intuitiva que indica el porcentaje total de predicciones correctas. No obstante, su utilidad se ve limitada cuando las clases están desbalanceadas. En escenarios donde los eventos positivos (como los deslizamientos) son escasos en comparación con los negativos, un modelo puede obtener alta precisión simplemente prediciendo la clase mayoritaria, sin aportar valor real. Por ello, se recomienda siempre complementar esta métrica con otras más específicas (He & Garcia, 2009).

d. ROC AUC

La curva ROC (Receiver Operating Characteristic) y su correspondiente área bajo la curva (AUC) son métricas clave para evaluar el desempeño global de un clasificador. La curva representa gráficamente la relación entre la tasa de verdaderos positivos (TPR) y la tasa de falsos positivos (FPR). Un AUC cercano a 1 indica excelente capacidad discriminativa, mientras que un valor cercano a 0.5 sugiere un comportamiento aleatorio. Esta métrica es especialmente útil en contextos con clases desbalanceadas, ya que no depende de un único umbral de decisión (Pheng & Thapa, 2021; Saito & Rehmsmeier, 2015).

e. Validación cruzada y técnicas de entrenamiento/prueba

La validación cruzada (crossvalidation) es una técnica que permite evaluar la generalización del modelo al dividir el conjunto de datos en múltiples subconjuntos (folds), entrenando y probando el modelo en distintas combinaciones. Esto reduce la varianza asociada a una única partición y ofrece una estimación más robusta del rendimiento del modelo. Las técnicas más comunes son la k-fold y la validación estratificada, esta última recomendable cuando hay desbalance en las clases (James et al., 2013).

f. Consideraciones en data sets geoespaciales:

En el caso de datos geoespaciales, existen desafíos particulares asociados a la autocorrelación espacial y la dependencia entre observaciones cercanas. Estos factores pueden llevar a sobreestimaciones del rendimiento del modelo si no se controlan adecuadamente. Por ello, se recomienda realizar separación geográfica entre datos de entrenamiento y prueba, o aplicar métodos de validación espacial como spatial k-fold cross-validation (Roberts, 2017). Además, la resolución espacial de las variables, la calidad de los datos de entrada y la representación de clases deben considerarse cuidadosamente al diseñar el modelo.

VI. METODOLOGÍA

La presente metodología describe el proceso general seguido para generar un mapa de predicción de deslizamientos en el departamento de San Marcos, Guatemala, mediante el uso de sistemas de información geográficos y machine learning. El enfoque parte de la elaboración de un inventario histórico de eventos de deslizamiento, integrando información geoespacial multifuente y variables condicionantes como elevación, pendiente, precipitación acumulada, cobertura vegetal (NDVI), y datos edafológicos (suelo), geológicos y taxonómicos. A partir de este inventario, se construyó una base de datos estructurada compuesta por puntos de deslizamiento y puntos de no deslizamiento generados aleatoriamente bajo criterios espaciales. Los datos fueron procesados y enriquecidos mediante la plataforma Google Earth Engine (GEE) y herramientas de SIG (QGIS). Posteriormente, se entrenó un modelo de clasificación supervisada tipo Random Forest, ajustando sus parámetros para obtener un desempeño óptimo, evaluado mediante métricas como exactitud, precisión, sensibilidad y el área bajo la curva ROC. Finalmente, el modelo fue aplicado a puntos distribuidos de forma regular sobre el área de estudio, permitiendo generar una superficie de predicción especializada de la susceptibilidad a deslizamientos.

A. Delimitación del área de estudio

Se seleccionó el departamento de San Marcos, Guatemala, como área de estudio debido a su alta susceptibilidad a procesos de remoción en masa, asociada tanto a sus características geográficas como a su actividad sísmica y volcánica. Esta región se encuentra en una zona montañosa con fuertes pendientes. Además, los registros históricos evidencian una alta frecuencia de eventos, lo cual la convierte en una zona prioritaria para estudios de riesgo y prevención. La delimitación del área de estudio se realizó utilizando las capas administrativas oficiales disponibles en el geoportal del Instituto Geográfico Nacional (IGN), asegurando precisión y consistencia geográfica. Todas las capas temáticas utilizadas fueron recortadas a dicho límite departamental, garantizando un análisis centrado exclusivamente en el territorio de San Marcos.

B. Recolección y preparación del inventario de deslizamientos

Como primera instancia, se debe recolectar la información necesaria para generar un modelo de entrenamiento. Dichos datos se obtienen a partir de inventarios de deslizamientos recolectados a lo largo de los años (2008-2023) por parte de SE-CONRED (Secretaría Ejecutiva de la Coordinadora Nacional de Reducción de Desastres). Estos inventarios incluyen información como la ubicación del incidente, el tipo de incidente (sismo, deslizamiento, temperaturas bajas, etc.). Con base en el enfoque de este estudio, se utilizan solamente las columnas de latitud (LAT), longitud (LONG), tipo de incidente (Incidente Tipo Incidente), fecha del incidente (Fecha Incidente), y el departamento, debido a que se está obteniendo para el departamento de San Marcos. Tras filtrar la información necesaria, se procede a recopilar la información en un solo

Excel, organizado por fecha por facilidad. Para integrarlo a QGIS, se agrega una nueva capa vectorial a partir de un archivo delimitado por comas (csv), en donde se asigna el valor de X a la columna de longitud (LONG) y el valor de Y a la columna de latitud (LAT). Se selecciona que la capa se asigne al CRS de 4236 WGS 84.

Para la elaboración de este estudio, se utilizaron los siguientes programas o software:

Archivo	Descripción
QGIS 3.36.2	Se utilizó QGIS como software de SIG para manejar toda la información geoespacial a través de capas, polígonos y puntos.
Python 3.13 (64-bit)	Generación de modelos de entrenamiento y predicción con algoritmos de Random Forest y XGBoost.
Thonny	Ambiente de Python para ejecución de programas de entrenamiento, obtención de resultados (reporte de clasificación, matriz de confusión) y mapa de susceptibilidad.
Excel	Procesamiento y depuración de archivos de inventarios de incidentes obtenidos de SE-CONRED, y manejo de archivos csv.

Tabla 8. Software o programas empleados para este trabajo.
(Fuente: Elaboración propia).

De igual forma, se utilizaron los siguientes archivos:

Archivo	Descripción	Fuente
Incidentes-2008.xlsx	Inventario de incidentes varios del año 2008	SE-CONRED
Incidentes-2009.xlsx	Inventario de incidentes varios del año 2009	SE-CONRED
Incidentes-2010.xlsx	Inventario de incidentes varios del año 2010	SE-CONRED
Incidentes-2011.xlsx	Inventario de incidentes varios del año 2011	SE-CONRED
Incidentes-2012.xlsx	Inventario de incidentes varios del año 2012	SE-CONRED
Incidentes-2013.xlsx	Inventario de incidentes varios del año 2013	SE-CONRED
Incidentes-2014.xlsx	Inventario de incidentes varios del año 2014	SE-CONRED
Incidentes-2015.xlsx	Inventario de incidentes varios del año 2015	SE-CONRED
Incidentes-2016.xlsx	Inventario de incidentes varios del año 2016	SE-CONRED
Incidentes-2017.xlsx	Inventario de incidentes varios del año 2017	SE-CONRED
Incidentes-2018.xlsx	Inventario de incidentes varios del año 2018	SE-CONRED
Incidentes-2019.xlsx	Inventario de incidentes varios del año 2019	SE-CONRED
Incidentes-2020.xlsx	Inventario de incidentes varios del año 2020	SE-CONRED
Incidentes-2021.xlsx	Inventario de incidentes varios del año 2021	SE-CONRED
Incidentes-2022.xlsx	Inventario de incidentes varios del año 2022	SE-CONRED
Incidentes-2023.xlsx	Inventario de incidentes varios del año 2023	SE-CONRED
Inventario_deslizamientos	Inventario de deslizamientos filtrado, con columnas de latitud, longitud, fecha de incidente abarcando el límite departamental de San Marcos.	Elaboración propia.

Archivo	Descripción	Fuente
SM_Deslizamientos.epsg .csv	Puntos de inventario de deslizamientos proyectado en SRC 32615 - WGS 84 / UTM zone 15N	Elaboración propia.

Tabla 9. Archivos utilizados para la construcción del inventario de deslizamientos.
(Fuente: Elaboración propia).

C. Recolección y preparación de información de datos de entrenamiento

1. Recolectar información geográfica de Guatemala

Para generar la base de datos de entrenamiento, se requieren mapas con la información geológica, taxonómica, etc. del país de Guatemala. Dicha información se puede solicitar con el Ministerio de Agricultura y Ganadería (MAGA) o se puede acceder en Geo portales como el del SEGEPLAN o ArcGIS Online.

Información	Descripción	Fuente
Taxonomía	Mapa de taxonomía de suelos de Guatemala.	MAGA
Series de suelos	Mapa de serie de suelos del territorio de Guatemala, que incluye orden de suelos.	MAGA
Geología	Mapa de geología del territorio de Guatemala.	MAGA

Tabla 10. Fuentes de información utilizadas en el análisis.
(Fuente: Elaboración propia).

La información contenida dentro de dichos archivos incluye lo siguiente. Se emplearon solamente las filas que pudieran influir en el análisis para evitar sobrecargar el modelo.

Columna de atributo	Descripción	Valores
MAT_ORIG	Material de origen.	Arena de mar, ceniza volcánica, caliza, material aluvial, etc.
RELIEVE	Pendiente.	Plano, casi plano, inclinado, etc.
CLAS_DRE	Drenaje del suelo.	Buen drenaje, excesivo, moderado, etc.
TEX_SUP	Textura superficial.	Arcilla, arena, franca, franco arcillo-arenosa, etc.
C_TEXSUB	Clasificación general de la textura superficial.	Gruesa, media, moderadamente fina, etc.
TEX_SUB	Textura del subsuelo.	Arcilla, arena, caliza fracturada, ceniza no intemperizada, etc.
C_TEXSUB	Clasificación general del subsuelo.	Gruesa, media, moderadamente fina, etc.

Columna de atributo	Descripción	Valores
PRO_EF	Profundidad efectiva del perfil.	10, 40, 50, etc.
C_PROF	Categoría de profundidad.	Muy delgado, delgado, mediano, profundo, etc.
RIES_ERO	Riesgo de erosión.	Bajo, regular, alto, etc.

Tabla 11. Atributos en mapas de Series de Suelos de Guatemala.
(Fuente: MAGA, 2020).

Columna de atributo	Descripción	Valores
UNIDAD	Clasificación general del suelo basada en su génesis y características.	Dv-Dd-Pd, Dd-Ds-Ld, Dd-Eo, LsVs, etc.
ORDEN	Categoría principal en taxonomía que refleja procesos de formación.	Andisoles, Inceptisoles, Entisoles, Mollisoles, etc.
SUBORDEN	Subcategoría del orden que permite afinar las propiedades edáficas y su uso.	Buen drenaje, excesivo, moderado, etc.
CAR_SUBO	Suborden o subdivisión del orden que define condiciones específicas de humedad y temperatura.	Arquico, arenosos, comunes, Udico, etc.
POT_FER	Potencial de fertilidad del suelo, relacionado con la productividad vegetal.	Bajo, medio, alto, etc.
CAR_FIS	Calidad física del suelo.	Mala, regular, buena, etc.
HUM_SUE	Nivel de humedad disponible en el suelo para las plantas.	Adecuada, deficiente, excesiva, variable.
SAT_BAS	Grado de saturación con bases (calcio, magnesio, potasio, sodio) del suelo.	Baja, regular a alta, alta, etc.
C_MO	Contenido de materia orgánica en el suelo.	Bajo, regular a bajo, regular a alto, etc.
LIMIT	Limitaciones del suelo como profundidad, pedregosidad o retención de agua.	Baja fertilidad, acidez, deficiencia de humedad, exceso de humedad, etc.

Tabla 12. Atributos taxonómicos y de calidad del suelo de Guatemala
(Fuente: MAGA, 2020).

D. Preparación de información en capas de SIG (QGIS)

1. Abrir archivos con información geográfica de Guatemala en software de SIG, en este caso se utilizó QGIS.
 - a. Utilizando el polígono de la delimitación departamental de San Marcos, se recortan los mapas de manera que se pueda contar solamente con la información correspondiente geográficamente a San Marcos. Esto facilitará la revisión de los atributos, en caso haya valores nulos.
 - b. Vector > Herramientas de geoproceto > cortar.
 - c. Utilizando el polígono de la delimitación departamental de San Marcos, se recortan los mapas de manera que se pueda contar solamente con la información correspondiente geográficamente a San Marcos. Esto facilitará la revisión de los atributos, en caso haya valores nulos.
2. Convertir las variables categóricas a numéricas.
 - a. Generar una leyenda para cada columna.
 - b. Generar una nueva columna para cada columna original, renombrada "Nombre_original_N".
 - c. Habilitar la edición de la capa, utilizar la calculadora de campos.
 - d. Redefinir la columna que se está trabajando, utilizando un CASE de acuerdo para cada columna. Abajo se presenta el CASE para el atributo de Orden en la capa de taxonomía.

```
CASE
  WHEN "ORDEN" = 'Andisoles' THEN 1
  WHEN "ORDEN" = 'Inceptisoles' THEN 2
  WHEN "ORDEN" = 'Entisoles' THEN 3
  WHEN "ORDEN" = 'Mollisoles' THEN 4
  WHEN "ORDEN" = 'Ultisoles' THEN 5
  WHEN "ORDEN" = 'Mollisoles' THEN 6
  WHEN "ORDEN" = 'Alfisoles' THEN 7
  WHEN "ORDEN" = 'Vertisoles' THEN 8
  ELSE 0
END
```

- e. Repetir paso para todas las columnas en las capas empleadas.

E. Generación de puntos de no deslizamiento

El modelo predictivo se basa en aprendizaje supervisado, por lo que fue necesario construir una base de datos con puntos etiquetados como eventos de deslizamiento (clase 1) y puntos sin deslizamiento (clase 0), que cumplan con condiciones de representatividad espacial.

Para evitar el sesgo por autocorrelación espacial y garantizar independencia entre clases, los puntos de no deslizamiento fueron seleccionados a una distancia mínima de 1.50 km de cualquier punto de deslizamiento, tomando en cuenta el área de distribución de puntos. Esta separación permite asegurar que las condiciones ambientales y geográficas difieran suficientemente, lo que mejora la robustez del modelo y reduce el riesgo de sobreajuste.

1. Recortar capas.

- a. Utilizando el polígono de la delimitación departamental de San Marcos, se recortan los mapas de manera que se pueda contar solamente con la información correspondiente geográficamente a San Marcos. Esto facilitará la revisión de los atributos, en caso haya valores nulos.
- b. Vector > Herramientas de geoprocso > cortar.
- c. Generar un Buffer del inventario de deslizamientos.
 - Reproyectar la capa de inventario de deslizamientos a SRC 32615 - WGS 84 / UTM zone 15N, de manera que se pueda utilizar una unidad de buffer en metros.
 - Vector > Herramientas de geoprocso > Buffer.

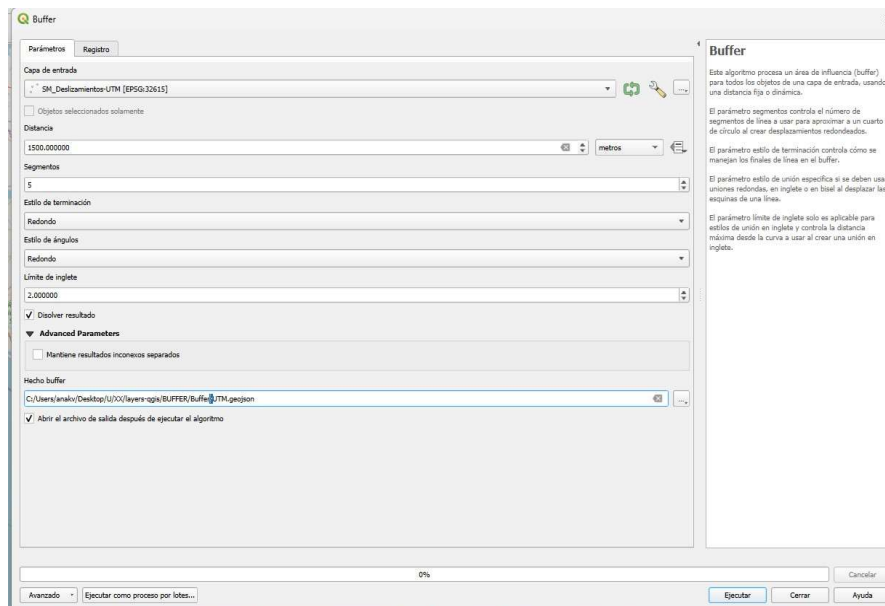


Figura 31. Generación de buffer de puntos de deslizamientos.
(Fuente: Elaboración propia).

- Indicar dirección a la cual se guardará el archivo resultante, se puede guardar archivo como: geojson, ESRI Shapefile, Geopackage, etc.
- Ejecutar procesos con valores y campos indicados en la ilustración X.
- Verificar que la capa de buffer sea visible en el lienzo de mapa en QGIS y haga sentido de acuerdo con los parámetros indicado.

2. Crear una capa de diferencia de límite de área de estudio con Buffer.

- a. Usando el límite departamental de San Marcos, se debe de restar el área del polígono limitante con la capa de buffer de 1.50 km.
- b. Verificar que las capas tengan el mismo CRS.
- c. Vector > Herramientas de geoprocresamiento > Corte.

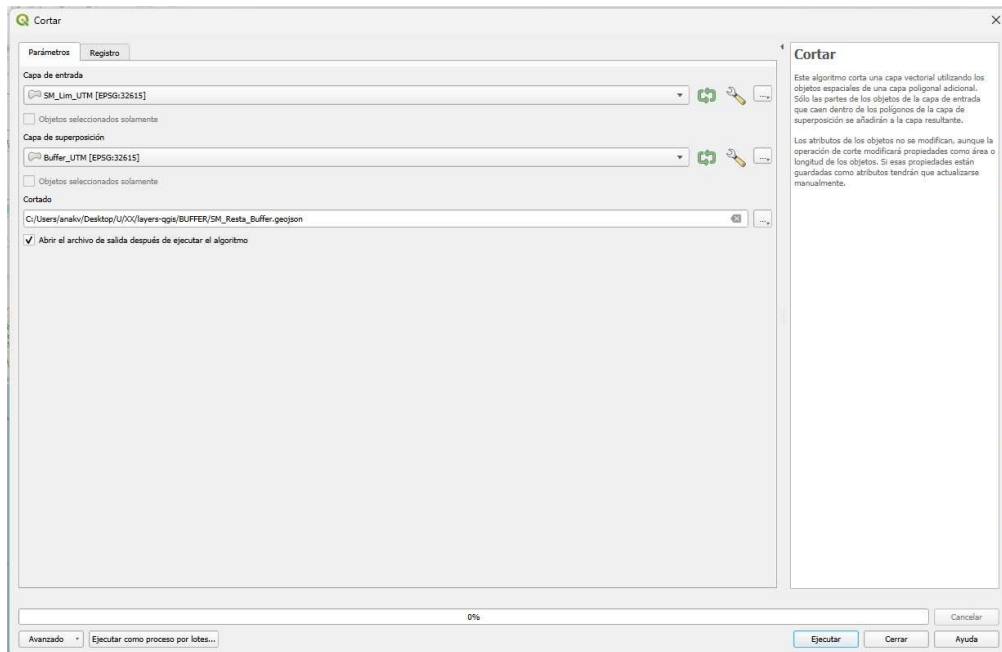


Figura 32. Generación de corte de límite departamental y buffer.
(Fuente: Elaboración propia).

- d. Indicar dirección a la cual se guardará el archivo resultante, al igual que el tipo de archivo.
- e. Ejecutar procesos con valores y campos indicados en la figura 25.
- f. Verificar que la capa del corte sea visible en el lienzo de mapa en QGIS y haga sentido de acuerdo con los parámetros indicados y las capas de entrada y de superposición.

3. Crear una capa de diferencia de límite de área de estudio con Buffer.
 - a. Vector > Herramientas de investigación > Puntos aleatorios dentro de polígonos
 - b. Generar 282 puntos debido a que esa es la cantidad de puntos dentro del inventario de deslizamientos. De acuerdo con el estudio realizado por Liu, X, Li, W. et al (2011), se demostró que usar conjuntos de entrenamiento balanceado mejora el desempeño del modelo a comparación de cuando hay desbalance entre las clases.

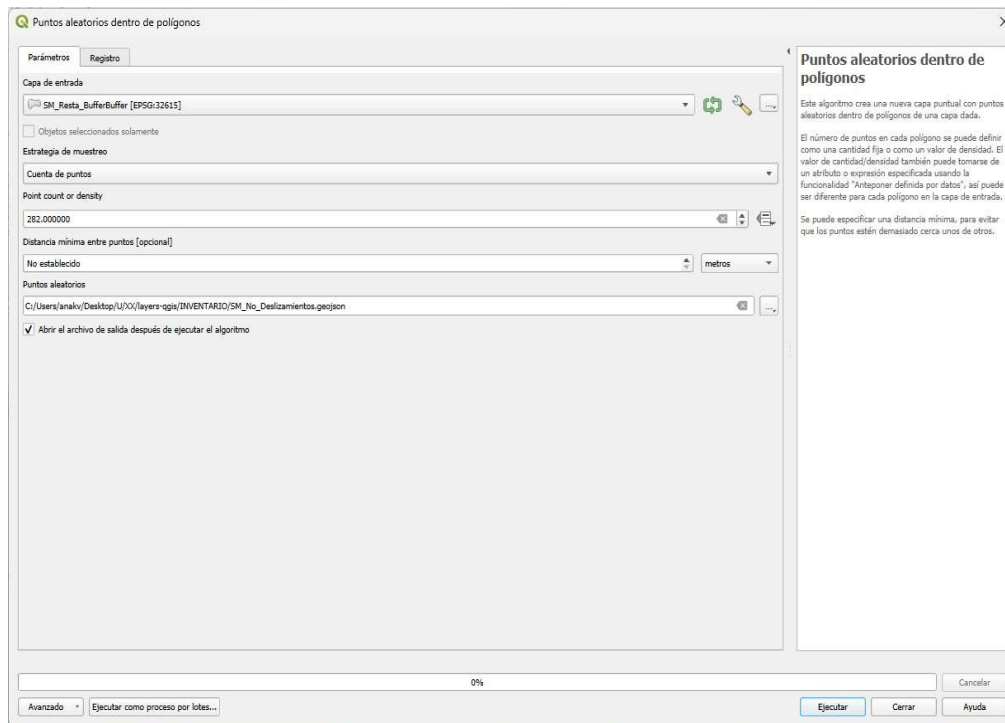


Figura 33. Generación de puntos aleatorios dentro de polígono con resta de Buffer.
(Fuente: Elaboración propia).

- c. Indicar dirección a la cual se guardará el archivo resultante, al igual que el tipo de archivo.
 - d. Ejecutar procesos con valores y campos indicados en la figura 26.
 - e. Verificar que la capa del corte sea visible en el lienzo de mapa en QGIS y haga sentido de acuerdo con los parámetros indicados y las capas de entrada y de superposición.
4. Agregar atributos a puntos de no deslizamiento.
 - a. Agregar 3 columnas en capa de puntos aleatorios (no deslizamientos).

b. Tabla de atributos > Habilitar edición > Crear nueva columna > Nombre: "Fecha", Categoría: Date (Fecha)/Nombre: "LAT", Categoría: Número decimal/Nombre: "LONG", Categoría: Número decimal.

- Fecha: `to_date('2008-01-01')+make_interval(days := rand(0, 5844))`.
- LAT (Latitud): `y$(geometry)`
- LONG (Longitud): `y$(geometry)`

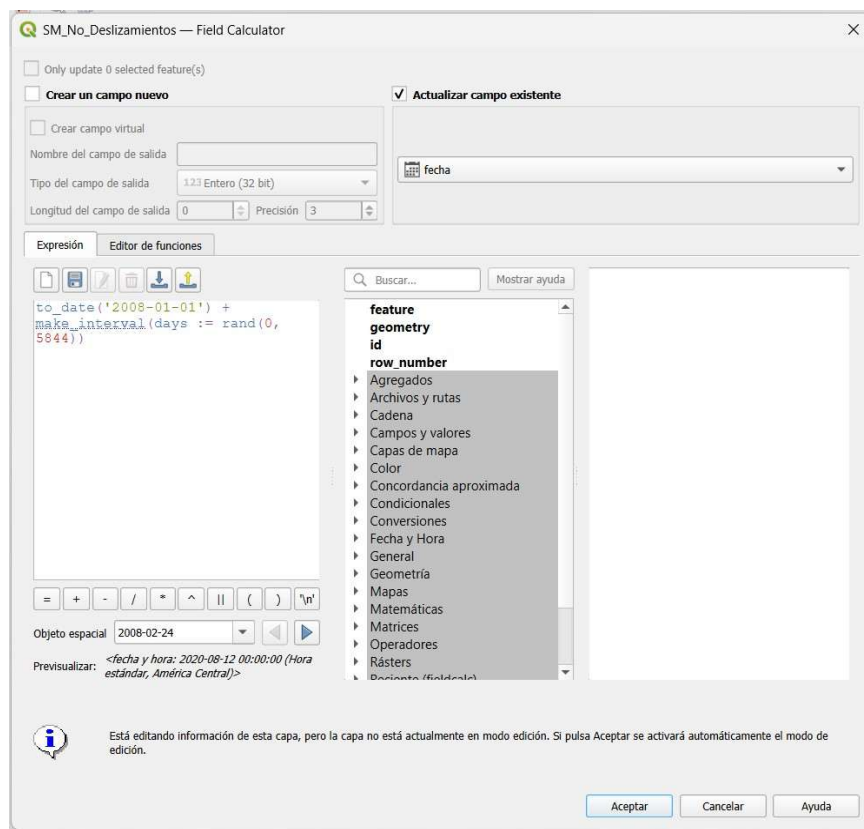


Figura 34. Generación de fechas aleatorias para puntos de no deslizamiento. (Fuente: Elaboración propia).

- c. Generar una columna de landslide.
- d. Tabla de atributos > Habilitar edición > Crear nueva columna > Nombre: "landslide", Categoría: Int-32 (Número entero)
- e. Usando la calculadora de campo (Ctrl+M) desde el panel de tabla de atributos, reescribir columna de fecha creada en paso anterior y darle valor de "0" para la expresión.

F. Generación de inventario total

1. Verificación de información.
 - a. Verificar que ambas capas tengan los mismos atributos.
 - b. Verificar que ambas capas tienen el mismo SRC.
 - c. El valor de landslide es de 0 para los no deslizamientos y 1 para los deslizamientos.
2. Unir capas.
 - a. `Vector > Herramientas de gestión de datos > Unir capas vectoriales`
 - b. Revisar que el conjunto de puntos aparece en el lienzo del mapa y es coherente con las capas de entrada y parámetros utilizados.
 - c. Exportar capa a un archivo csv o archivo shapefiles guardado como una carpeta comprimida (.zip).
 - d. Cargar archivo .csv o shp (.zip) como un asset a GEE.

G. Extracción de valores para entrenamiento del modelo

La extracción de variables geoespaciales para puntos de estudio se realizó utilizando la plataforma de procesamiento en la nube Google Earth Engine (GEE), debido a su acceso a colecciones satelitales históricas, capacidad de análisis de grandes volúmenes de datos y herramientas específicas para preprocesamiento como enmascaramiento de nubes y cálculo de índices espectrales. El objetivo principal fue asociar a cada punto del inventario variables ambientales en la fecha del evento, considerando tanto factores permanentes como condicionales.

1. Preparación de datos de entrada: Se cargó un inventario georreferenciado de deslizamientos con coordenadas (latitud y longitud) y la fecha del evento asociada a cada punto. Este inventario fue convertido en una colección de tipo `FeatureCollection` dentro de GEE, permitiendo su manipulación mediante funciones `map`. Para garantizar la correcta interpretación temporal, se transformaron los campos de fecha a objetos de tipo `ee.Date`. Esto permitió realizar operaciones basadas en ventanas temporales relativas a cada evento.
2. Se utilizó el modelo digital de elevación SRTM (Shuttle Radar Topography Mission) provisto por USGS (USGS/SRTMGL1_003). A partir de este modelo se extrajo:
 - a. Elevación: valor absoluto en metros.
 - b. Pendiente: Calculada mediante la función `ee.Terrain.slope()`.

3. Precipitación acumulada (CHIRPS)

Para la estimación de precipitación se empleó la colección CHIRPS Daily (UCSBCHG/CHIRPS/DAILY), la cual proporciona datos diarios desde 1981 con resolución espacial de aproximadamente 5 km.

Para cada punto, se acumuló la precipitación de los 7 días anteriores al evento, utilizando una ventana móvil basada en la fecha del deslizamiento. Se aplicó un filtrado temporal dinámico y la suma de la colección filtrada mediante `ee.Reducer.sum()`.

4. NDVI (Índice de Vegetación Normalizado)

El cálculo del NDVI se realizó utilizando una jerarquía de fuentes satelitales, según la disponibilidad temporal de cada misión:

- a. Sentinel-2 SR (COPERNICUS/S2_SR) — prioridad principal, resolución 10 m
- b. Landsat 8 SR (LANDSAT/LC08/C02/T1_L2) — respaldo para fechas anteriores
- c. Landsat 7 SR (LANDSAT/LE07/C02/T1_L2) — respaldo histórico adicional

Para cada misión se definió una ventana de ± 15 días alrededor de la fecha del evento. Se aplicó cloud masking de la siguiente manera:

- d. Sentinel-2: Se utilizó la banda SCL (Scene Classification Layer) para enmascarar nubes, sombra de nubes y otros artefactos no deseados.
- e. Landsat 7 y 8: Se usó la banda QA_PIXEL, aplicando operaciones bitwise para identificar y remover píxeles con nubes y sombras.

El índice NDVI fue calculado usando la relación:

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

5. Velocidad del viento La velocidad del viento a 10 metros se extrajo desde la colección ERA5Land Daily Aggregated (ECMWF/ERA5_LAND/DAILY_AGGR), específicamente las bandas:

- a. `u_component_of_wind_10m`
- b. `v_component_of_wind_10m`

Se calculó la velocidad escalar del viento mediante la fórmula:

$$Vel_viento = v^2 + u^2$$

El valor fue extraído para el día del evento, representando una condición atmosférica puntual que puede influir como detonante.

H. Construcción de la base de datos completa

1. Cargar archivo .csv o shp (.zip) como un asset a GEE.
2. A través de código de GEE, descargar archivo delimitado por comas (csv) que incluya las variables indicadas, al igual que las columnas de LAT, LONG de manera que se pueda integrar a QGIS.
3. Abrir archivo en QGIS. Capa > Añadir capa > Añadir capa de texto delimitado.
4. Obtener valores de mapas de taxonomía, geología, series de suelos para los puntos del inventario.
 - a. Vector > Herramientas de gestión de datos > Unir atributos por geolocalización.
 - b. Repetir este proceso por cada mapa, hasta obtener el inventario con todas las variables de entrenamiento.
5. Exportar capa final a archivo delimitado por comas (csv).

I. Generación de puntos de estudio

Con el objetivo de realizar una predicción espacial continua o un muestreo sistemático sobre el área de estudio, se generó una grilla de puntos distribuidos a intervalos regulares de 500 metros. Este proceso permite crear una capa de puntos con cobertura homogénea, adecuada para aplicar modelos de clasificación, interpolaciones u otros análisis espaciales.

1. Exportar capa final a archivo delimitado por comas (csv).
 - a. Vector > Herramientas de investigación > Puntos regulares.
 - b. Seleccionar la extensión de entrada como la capa del límite departamental de San Marcos, o dibujarla manualmente en el mapa asegurándose que abarque todo el departamento.
 - c. Seleccionar dirección y guardar archivo.

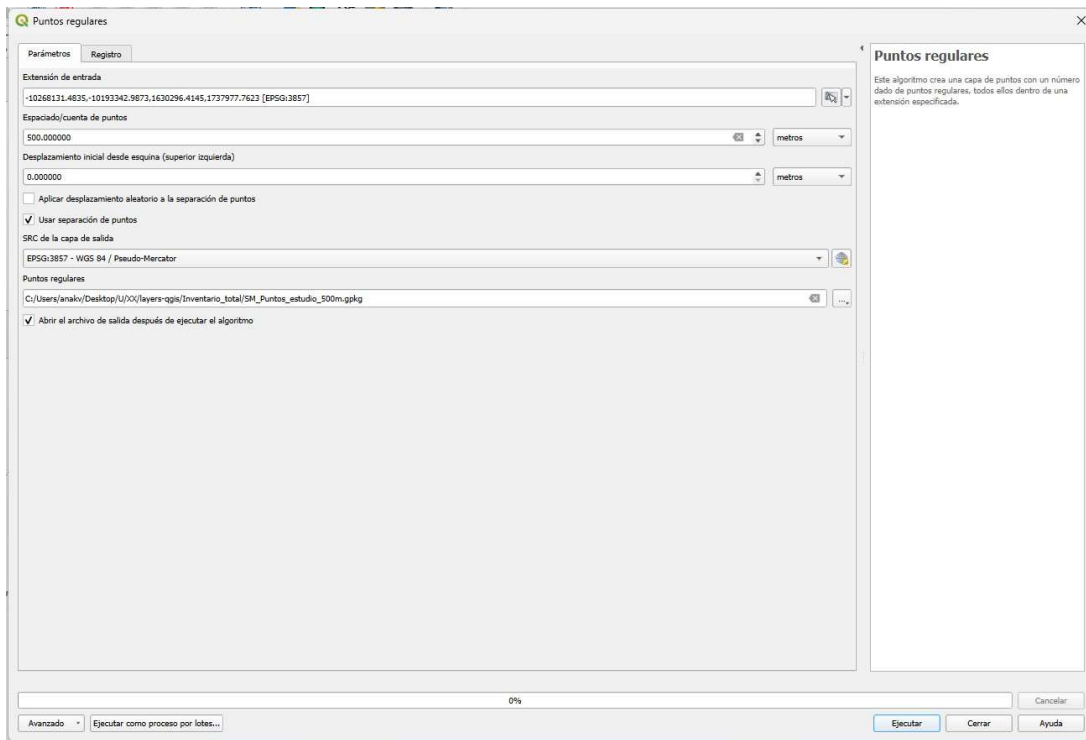


Figura 35. Generación de puntos regulares a una distancia de 500m entre cada uno en extensión del terreno de San Marcos.
(Fuente: Elaboración propia).

2. El paso anterior generará puntos regulares a lo largo de la extensión, como se observa en la Figura 28.

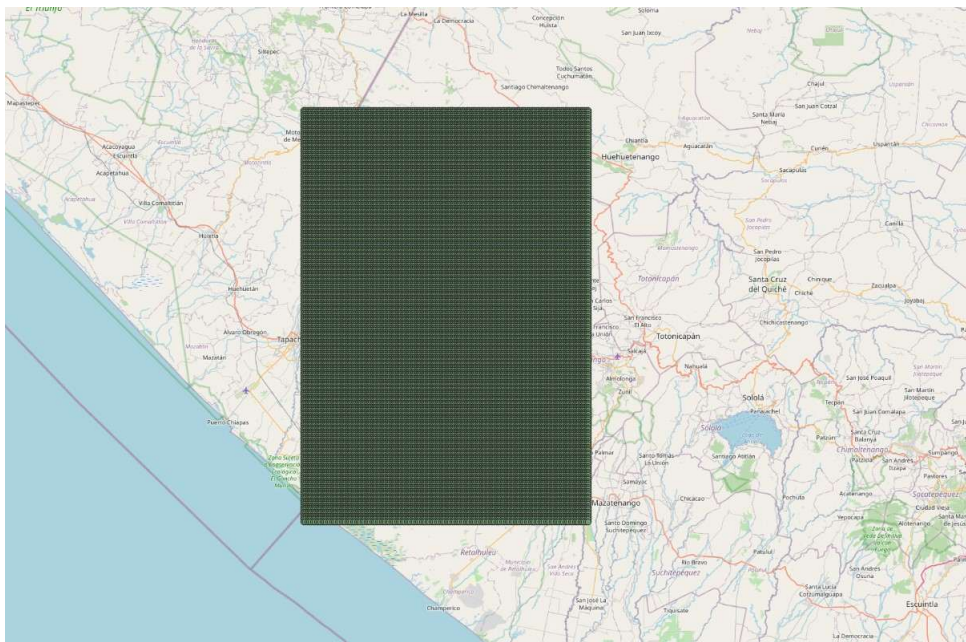


Figura 36. Visualización de puntos regulares a lo largo de la extensión.
(Fuente: Elaboración propia).

3. Recortar puntos generados a límite departamental.
 - a. Vector > Herramientas de geoprocreso > cortar.
 - b. Seleccionar capa de puntos regulares como entrada y capa de límite departamental como capa de superposición.
 - c. Seleccionar dirección de guardado y ejecutar proceso.

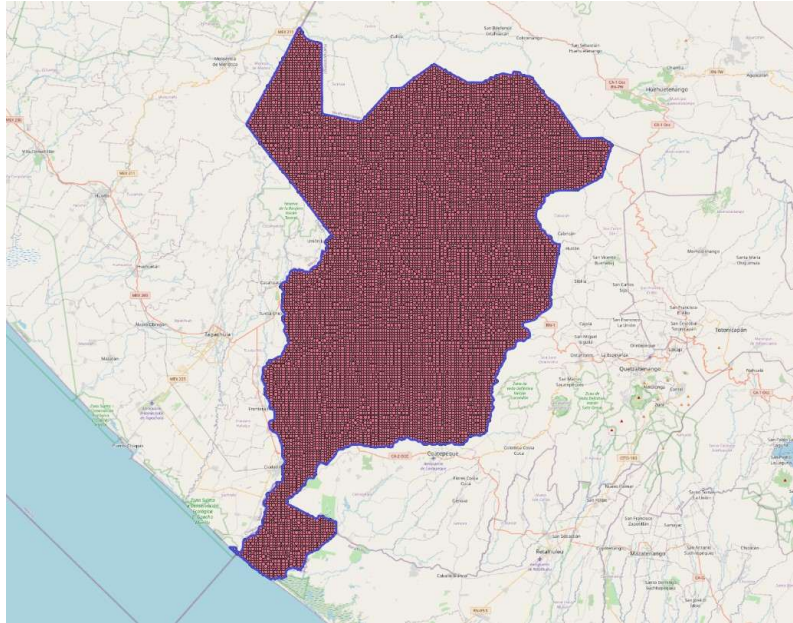


Figura 37. Visualización de puntos regulares cortados a lo largo del polígono delimitante.
(Fuente: Elaboración propia).

- d. Tabla de atributos > Crear nueva columna > “Fecha” > Calculadora de campo.
- e. Actualizar campo existente de fecha con la fecha de la que se busque realizar el estudio.
- f. Exportar capa a un archivo csv o archivo shapefiles guardado como una carpeta comprimida (.zip).
- g. Cargar archivo .csv o shp (.zip) como un asset a GEE.
- h. Cambiar dirección de asset en código GEE generado para obtener variables de inventario, y ejecutar las tareas.
- i. A través de GEE, descargar archivo delimitado por comas (csv) que incluya las variables indicadas, al igual que las columnas de LAT, LONG de manera que se pueda integrar a QGIS.

- j. Abrir archivo en QGIS. Capa > Añadir capa > Añadir capa de texto delimitado.
- k. Obtener valores de mapas de taxonomía, geología, series de suelos para los puntos del inventario.
- l. Vector > Herramientas de gestión de datos > Unir atributos por geolocalización.
- m. Repetir este proceso por cada mapa, hasta obtener el inventario con todas las variables de entrenamiento.
- n. Exportar capa final a archivo delimitado por comas (csv).

J. Generación de modelo – Random Forest:

1. Abrir ambiente de Python que se vaya a utilizar. Para este estudio se empleó el ambiente de Thonny.
2. Guardar el archivo de código en una carpeta fácil de ubicar.
3. Colocar archivos de comas delimitadas (csv) de inventario final y de puntos de predicción en la misma carpeta donde se encuentra el archivo del código. (Nota: Se debe verificar que el orden de las columnas de ambos archivos sea el mismo).
4. Seleccionar dirección y guardar archivo.
5. Instalar e importar las siguientes librerías:

Librería	Uso principal
pandas	Manipulación y análisis de datos estructurados en forma de tablas (DataFrames). Ideal para importar, limpiar y transformar datasets.
geopandas	Extiende las funcionalidades de pandas para manejar datos geoespaciales (shapefiles, GeoJSON). Permite operaciones como recorte, unión espacial y reproyección.
matplotlib.pyplot	Librería de visualización que permite generar gráficos estáticos como histogramas, curvas ROC, scatter plots, entre otros.
shapely.geometry	Herramientas para crear y manipular objetos geométricos como puntos, líneas y polígonos, esenciales en análisis espacial.

Librería	Uso principal
sklearn (Scikit-learn)	Biblioteca de aprendizaje automático que incluye algoritmos como Random Forest, XGBoost (vía wrapper), y métricas de evaluación (accuracy, f1-score, etc.).
joblib	Permite guardar y cargar modelos entrenados, optimizando su reutilización sin necesidad de retrain. Es eficiente para objetos grandes como modelos de ML.
seaborn	Librería de visualización basada en matplotlib, especializada en gráficos estadísticos como mapas de calor, diagramas de dispersión con densidad, y boxplots.

Tabla 13. Librerías de Python utilizadas y su función principal para modelo Random Forest. (Fuente: Elaboración propia).

6. Instalar e importar las siguientes librerías:

- a. Tamaño de muestra de entrenamiento: 70% (`test_size=0.3`). Se utiliza este valor debido a que, Según James, et. al, 2013, particiones cercanas al 70/30 son apropiadas cuando se dispone de una cantidad moderada de datos, porque permiten una estimación robusta del error de generalización sin sacrificar demasiada información en la fase de entrenamiento.
- b. Se utilizaron 100 árboles de decisión. Un número óptimo es de 100 a 500 árboles, ya que de acuerdo con Breiman, L. (2001), la precisión del modelo se estabiliza después de un número de árboles, entre los 100 y 500.
- c. Usar un valor de semilla aleatorio, puede ser cualquier número, esto hará que los resultados sean reproducibles.
- d. Indicarle a Python que debe de entrenar el modelo usando la variable predictora "landslide_val" por lo que para las variables de entrenamiento se utilizan todas excepto esta.
- e. Entrenar el modelo de RF con las variables de entrenamiento `X_train` y `Y_train`.
- f. Evaluar el modelo con la muestra de prueba `X_test` y `Y_test`.
- g. Utilizar el archivo csv con los puntos de estudio para predecir sobre estos nuevos puntos.
- h. Imprimir el mapa de acuerdo con la clasificación indicada de probabilidad:

- Muy baja: probabilidad de deslizamiento < 0.25
- Baja: 0.25 < probabilidad de deslizamiento < 0.5
- Alta: 0.5 < probabilidad de deslizamiento < 0.75
- Muy alta: probabilidad de deslizamiento > 0.75

K. Generación de modelo – XGBoost

1. Abrir ambiente de Python que se vaya a utilizar. Para este estudio se empleó el ambiente de Thonny.
2. Guardar el archivo de código en una carpeta fácil de ubicar.
3. Colocar archivos de comas delimitadas (csv) de inventario final y de puntos de predicción en la misma carpeta donde se encuentra el archivo del código. (Nota: Se debe verificar que el orden de las columnas de ambos archivos sea el mismo).
4. Seleccionar dirección y guardar archivo.
5. Instalar e importar las siguientes librerías o herramientas:

Componente	Uso principal
xgboost	Implementación eficiente del algoritmo de <i>gradient boosting</i> sobre árboles de decisión. Ideal para clasificación y regresión con alto rendimiento computacional.
XGBClassifier()	Clase principal para construir modelos de clasificación. Permite configurar hiperparámetros como <code>n_estimators</code> , <code>max_depth</code> , <code>learning_rate</code> , entre otros.
fit()	Método utilizado para entrenar el modelo con datos de entrada (X) y etiquetas (y).
predict()	Realiza predicciones sobre nuevos datos una vez entrenado el modelo.
plot_importance()	Función que visualiza la importancia relativa de cada variable en el modelo, útil para la interpretación.
save_model() y load_model()	Métodos para guardar y cargar modelos entrenados, facilitando su reutilización sin necesidad de reentrenamiento.

Tabla 14. Librerías/componentes de Python utilizadas y su función principal para modelo XGBoost. (Fuente: Elaboración propia).

6. Generar código de XGBoost, con las siguientes especificaciones:
 - a. Tamaño de muestra de entrenamiento: 70% (James, et. al, 2013).

- b. Entrenar el modelo de RF con las variables de entrenamiento X_{train} y Y_{train} .
- c. Evaluar el modelo con la muestra de prueba X_{test} y Y_{test} .
- d. Utilizar el archivo csv con los puntos de estudio para predecir sobre estos nuevos puntos.
- e. Imprimir el mapa de acuerdo con la clasificación indicada de probabilidad:
 - Muy baja: probabilidad de deslizamiento < 0.25
 - Baja: $0.25 < \text{probabilidad de deslizamiento} < 0.5$
 - Alta: $0.5 < \text{probabilidad de deslizamiento} < 0.75$
 - Muy alta: probabilidad de deslizamiento > 0.75

L. Presentación de resultados

1. Tras correr ambos programas, guardar las imágenes de mapas de predicción a deslizamientos generadas, al igual que los gráficos de histograma de importancia de variables, el reporte de clasificación, la matriz de confusión y el valor de ROC AUC.
2. Presentar los resultados.
3. Analizar, comparar y discutir resultados.

VII. RESULTADOS Y DISCUSIÓN

La generación de un mapa de susceptibilidad a deslizamientos representa una herramienta esencial para la gestión territorial, en especial en regiones donde los fenómenos naturales como los movimientos en masa representan una amenaza significativa para la infraestructura, la seguridad de las personas y el desarrollo sostenible. La posibilidad de visualizar espacialmente las zonas con mayor propensión a deslizamientos permite priorizar intervenciones, delimitar áreas de uso restringido y fundamentar decisiones técnicas en planificación urbana, infraestructura vial y protección ambiental. En este estudio, se optó por aplicar los modelos de predicción únicamente al departamento de San Marcos, Guatemala. Esta decisión se sustenta en la necesidad de generar un modelo con mayor capacidad de generalización, aprovechando que dentro del mismo departamento existen condiciones meteorológicas, climáticas y geomorfológicas relativamente homogéneas, lo cual contribuye a que los datos de entrenamiento sean más representativos y pertinentes para identificar patrones reales de susceptibilidad.

A. Presentación de resultados – Mapa de predicción elaborado con Random Forest:

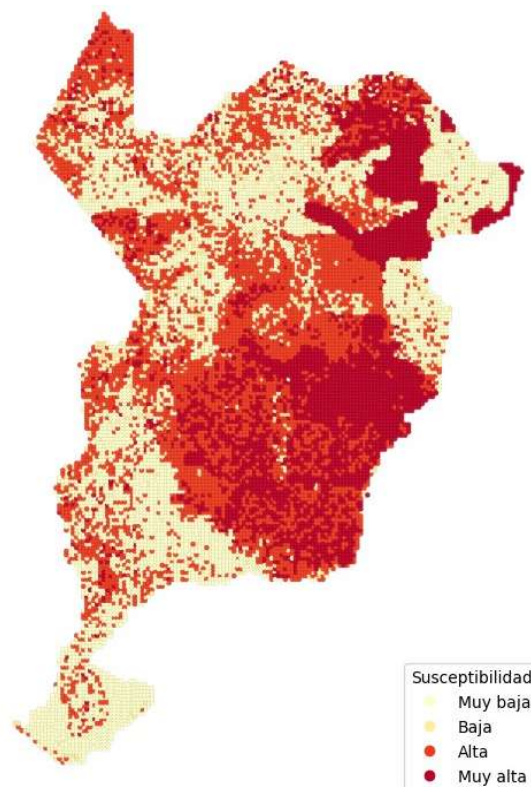


Figura 38. Mapa de probabilidad de deslizamientos empleando Random Forest.
(Fuente: Elaboración propia).

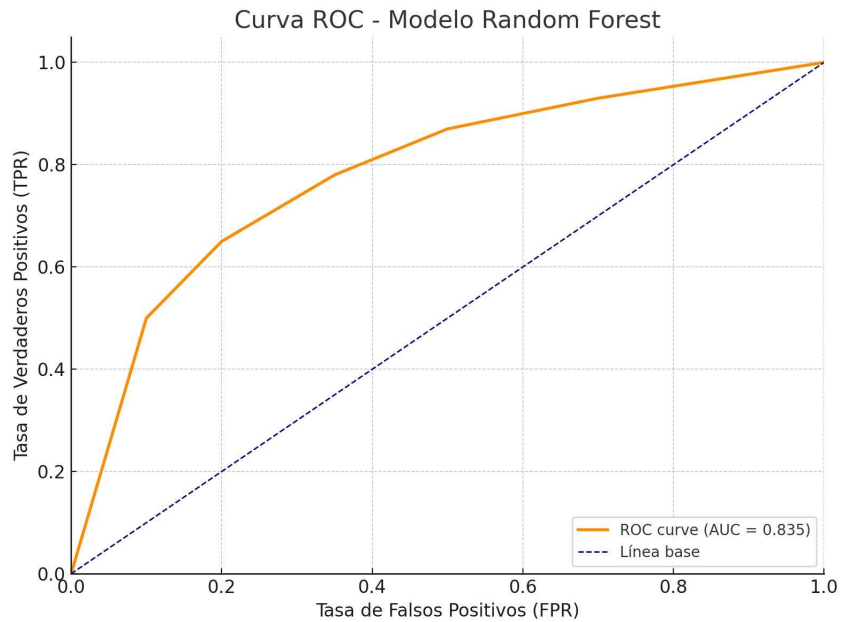


Figura 39. Gráfico de ROC-AUC para modelo de Random Forest.
(Fuente: Elaboración propia).

	Clase predicha	
	No deslizamiento (0)	Deslizamiento (1)
No deslizamiento (0)	58	27
Deslizamiento (1)	14	71

Tabla 15. Matriz de confusión - Modelo Random Forest.
(Fuente: Elaboración propia).

Clase	Precisión	Recall	F1-score	Soporte
0 (No deslizamiento)	0.81	0.68	0.74	85
1 (Deslizamiento)	0.72	0.84	0.78	85
Accuracy			0.76	170
Promedio macro	0.77	0.76	0.76	170
Promedio ponderado	0.77	0.76	0.76	170

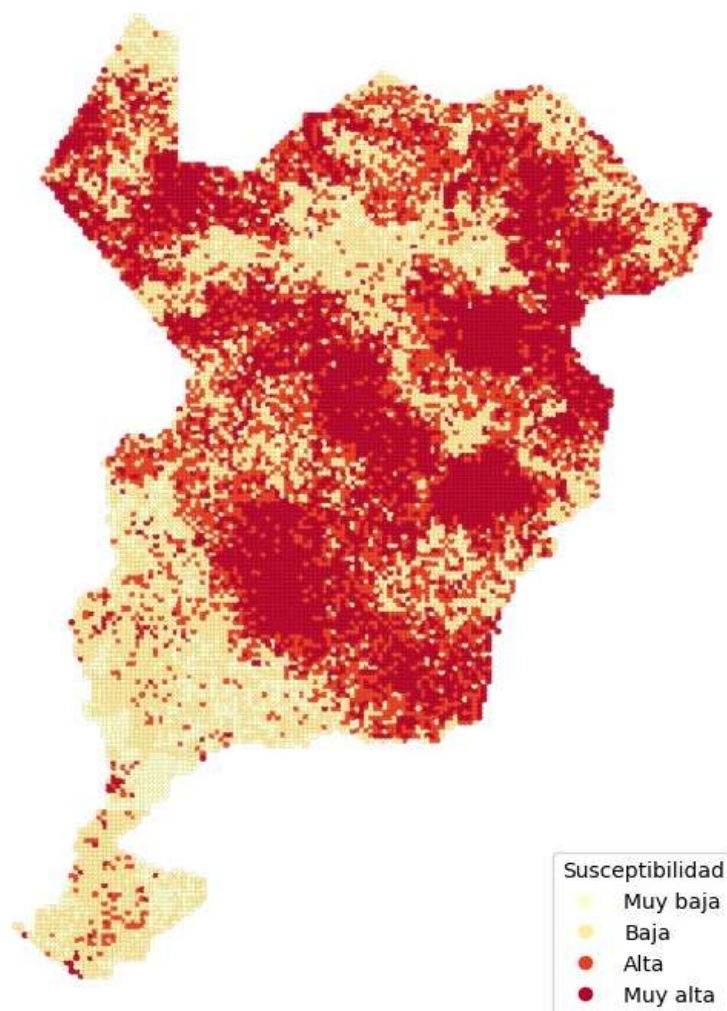
Tabla 16. Reporte de clasificación - Modelo Random Forest.
(Fuente: Elaboración propia).

Para la generación del mapa de susceptibilidad a deslizamientos, se utilizó el algoritmo de clasificación supervisada Random Forest, aplicando un enfoque basado en puntos georreferenciados dentro del área de estudio. Estos puntos incluyeron tanto eventos históricos de deslizamientos (clase positiva) como puntos de no deslizamiento generados aleatoriamente bajo criterios de separación espacial mínima. A cada punto se le asociaron variables predictoras extraídas de fuentes geográficas y satelitales, incluyendo factores topográficos, edafológicos, geológicos y climáticos. El modelo fue entrenado y validado utilizando una división estratificada del conjunto de datos, permitiendo evaluar su desempeño con métricas como precisión, F1-score y el área bajo la curva ROC (AUC). Este enfoque permitió la creación de una superficie continua que representa la probabilidad espacial de ocurrencia de deslizamientos dentro del área de estudio.

El rendimiento del modelo Random Forest fue evaluado mediante métricas de clasificación estándar, obteniendo un valor de precisión general (accuracy) de 0.76, lo cual indica que el 76% de los casos totales fueron correctamente clasificados como deslizamiento o no deslizamiento. Sin embargo, en contextos de análisis de riesgo, métricas como el recall y el F1-score adquieren mayor relevancia, dado que permiten evaluar la sensibilidad del modelo ante la clase positiva (eventos de deslizamiento) y su capacidad para balancear la precisión con la recuperación. En este sentido, el modelo alcanzó un recall de 0.84 para la clase de deslizamientos, lo que implica que identificó correctamente el 84% de los casos reales de deslizamiento, minimizando la ocurrencia de falsos negativos, un aspecto crucial en contextos de prevención de desastres (He & Garcia, 2009; Saito & Rehmsmeier, 2015).

El valor ROC AUC obtenido fue de 0.835, lo cual representa un rendimiento discriminativo alto. El área bajo la curva ROC (Receiver Operating Characteristic) mide la capacidad del modelo para distinguir entre clases, siendo 1.0 el valor ideal y 0.5 el equivalente al azar. De acuerdo con estudios previos, un AUC superior a 0.8 indica un rendimiento robusto para tareas de clasificación en entornos complejos y no lineales como la predicción de deslizamientos (Hong & Pradhan, 2019; Liaw & Wiener, 2002). En el contexto de este trabajo, un valor de 0.835 confirma que el modelo tiene una capacidad efectiva para priorizar áreas en riesgo, manteniendo un balance entre sensibilidad y especificidad.

B. Presentación de resultados – Mapa de predicción elaborado con XGBoost.



*Figura 40. Mapa de predicción de deslizamientos empleando XGBoost.
(Fuente: Elaboración propia).*

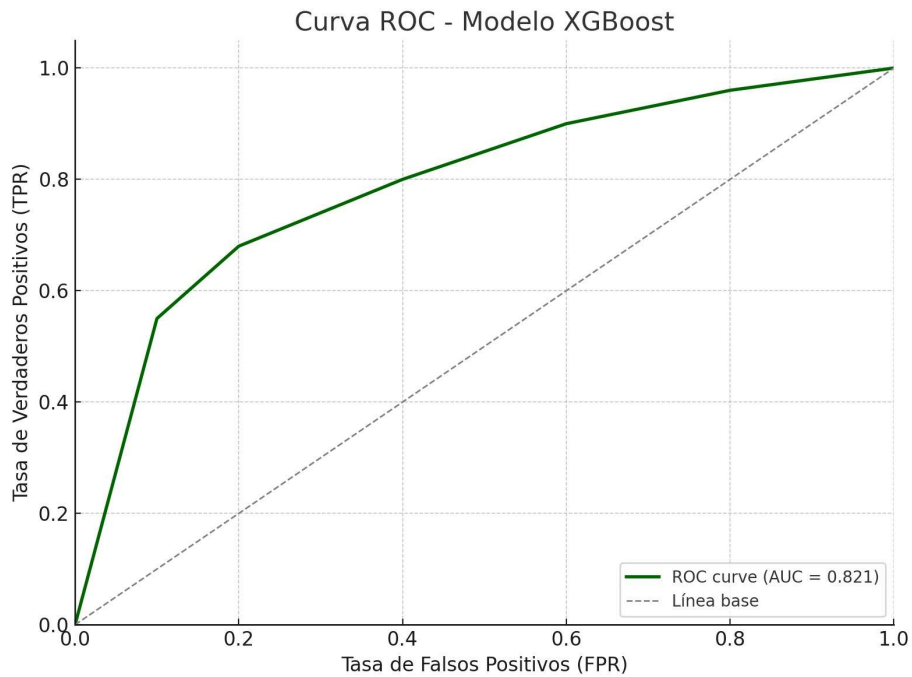


Figura 41. Gráfico de ROC-AUC para modelo de XGBoost.
(Fuente: elaboración propia).

	Clase predicha	
	No deslizamiento (0)	Deslizamiento (1)
No deslizamiento (0)	54	31
Deslizamiento (1)	14	71

Tabla 17. Matriz de confusión de modelo XGBoost.
(Fuente: elaboración propia).

Clase	Precisión	Recall	F1-score	Soporte
0 (No deslizamiento)	0.79	0.64	0.71	85
1 (Deslizamiento)	0.70	0.84	0.76	85
Accuracy			0.74	170
Promedio macro	0.75	0.74	0.73	170
Promedio ponderado	0.75	0.74	0.73	170

Tabla 18. Reporte de clasificación de modelo XGBoost.
(Fuente: elaboración propia).

El modelo XGBoost alcanzó un accuracy general del 74%, lo que indica que aproximadamente tres cuartas partes de las predicciones coinciden con las clases verdaderas. Esta cifra, aunque ligeramente inferior a la obtenida con Random Forest, sigue siendo representativa de un rendimiento aceptable, especialmente considerando la complejidad del fenómeno geoespacial modelado. En contextos de clasificación binaria para eventos naturales como deslizamientos, valores de exactitud superiores al 70% se consideran adecuados cuando hay un balance razonable entre clases y una alta dimensionalidad de datos (Pheng & Thapa, 2021; Vieira, Lacerda et al., 2022).

En cuanto a la sensibilidad o recall, el modelo mostró un valor de 0.84 para la clase positiva, es decir, logró identificar correctamente el 84% de los puntos etiquetados como zonas de deslizamiento. Este valor es especialmente relevante en la gestión de riesgos, ya que los falsos negativos (zonas peligrosas clasificadas como seguras) pueden tener consecuencias graves para la vida humana e infraestructura (Hong & Pradhan, 2019). La precisión fue de 0.70 para esa clase, lo que indica que el 70% de las predicciones positivas fueron efectivamente verdaderas. Esta relación equilibrada entre recall y precisión se traduce en un F1-score de 0.76, una medida útil cuando se busca compensar el impacto de ambos errores.

Por su parte, el área bajo la curva ROC (AUC = 0.8213) refuerza la eficacia general del modelo, ya que un valor superior a 0.8 es generalmente considerado como un indicador de buena capacidad discriminativa (Saito & Rehmsmeier, 2015). Esto implica que el modelo XGBoost puede distinguir con alta probabilidad entre zonas susceptibles y no susceptibles a deslizamientos, incluso si las clases no están perfectamente balanceadas. Aun cuando el modelo no alcance la máxima precisión posible, su capacidad de priorización espacial lo hace valioso en escenarios de toma de decisiones y asignación de recursos para prevención.

Finalmente, la matriz de confusión indica que el modelo cometió 14 falsos negativos (eventos de deslizamiento no detectados) y 31 falsos positivos (zonas seguras clasificadas como peligrosas). Esta distribución sugiere una ligera preferencia por minimizar los falsos negativos, lo cual es deseable en contextos de protección civil. En conjunto, los resultados confirman que XGBoost es un algoritmo competitivo y robusto para la modelación de susceptibilidad a deslizamientos, especialmente cuando se cuenta con un conjunto de variables diversas y categóricas.

C. Comparación de mapas de susceptibilidad.

Los mapas generados por los algoritmos Random Forest y XGBoost permiten visualizar espacialmente la clasificación de cada celda del territorio en una de las cuatro categorías de susceptibilidad: muy baja, baja, alta y muy alta. El modelo XGBoost muestra una alta concentración de celdas clasificadas como de alta y muy alta susceptibilidad en sectores del centro y noroccidente del departamento (Figura 28). Aunque no se dispone de verificación directa para estas zonas, el modelo aprendió que las combinaciones de variables en dichas regiones están asociadas a un alto riesgo histórico de deslizamientos, según el inventario empleado. Estas zonas

que el modelo identificó como de alta y muy alta susceptibilidad presentan características asociadas a altos valores de pendiente y precipitación, así como unidades geológicas que, según los datos de entrada, tienen mayor relación con la ocurrencia de deslizamientos, lo cual, válida, al menos de manera cualitativa, la consistencia del modelo.

Por su parte, el mapa generado mediante Random Forest (Figura 26) exhibe una distribución similar, aunque con una mayor proporción de áreas clasificadas como de “muy alta” susceptibilidad en el suroeste. Esta diferencia puede deberse a las características de cada modelo: mientras Random Forest tiende a ser más sensible a relaciones no lineales fuertes entre pocas variables, XGBoost es más eficaz capturando interacciones complejas y débiles entre múltiples variables (Chen & Guestrin, 2016).

Esta diferencia puede deberse a las características intrínsecas de cada algoritmo. Random Forest, al estar basado en la técnica de bagging (bootstrap aggregating), construye múltiples árboles de decisión entrenados sobre subconjuntos aleatorios del conjunto de datos, y realiza votación para la predicción final. Esta estructura permite que Random Forest sea particularmente efectivo en detectar relaciones no lineales fuertes entre un pequeño número de variables, ya que cada árbol puede especializarse en una combinación de atributos específicos, lo cual favorece la interpretación local de patrones bien definidos. Sin embargo, esta misma característica puede limitar su capacidad para identificar interacciones más sutiles o distribuidas entre múltiples variables con contribuciones marginales (Breiman, 2001).

Por otro lado, XGBoost, que implementa un enfoque de gradient boosting, construye árboles de manera secuencial, cada uno intentando corregir los errores del anterior. Esta metodología favorece una optimización progresiva y enfocada, permitiendo que el modelo aprenda de manera más eficiente las interacciones complejas, débiles o combinadas entre un gran número de variables, incluso cuando su efecto individual es mínimo. Gracias a su función objetivo regularizada y a su capacidad para manejar sparse data y regularización, XGBoost usualmente logra un mejor rendimiento en datasets con estructuras internas más complejas (Chen & Guestrin, 2016).

Esta diferencia metodológica explica por qué XGBoost puede extraer más información útil de variables categóricas con bajo poder predictivo individual, como algunas clases de suelos o unidades geológicas, las cuales tienden a ser desestimadas en modelos más orientados a interacciones fuertes como Random Forest. Adicionalmente, es importante aclarar que estas observaciones no implican que en dichas zonas haya necesariamente materiales geológicos o precipitaciones altas —sino que, en función del conjunto de entrenamiento y las variables disponibles, el modelo infirió una mayor probabilidad de deslizamiento en aquellas celdas donde se repiten patrones similares a los de eventos conocidos (los que tenían un valor de "1" en el atributo de "landslide_val"). Esta inferencia se basa en la lógica interna del modelo y no se puede considerar o interpretar como una validación de campo directa.

No obstante, las diferencias observadas entre los mapas generados por cada modelo, ambos coinciden en la identificación de patrones bien definidos de concentración de áreas con alta y muy alta susceptibilidad a deslizamientos. Es decir, a pesar de haber sido entrenados con estructuras internas distintas, tanto el modelo Random Forest como el XGBoost lograron destacar zonas donde, según el historial de eventos y las condiciones ambientales modeladas, existe una mayor probabilidad de ocurrencia de deslizamientos. Esta coincidencia no solo refuerza la validez de los resultados, sino que también indica que los modelos fueron capaces de capturar relaciones espaciales complejas entre las variables predictoras —como pendiente, precipitación, uso de suelo y geología— y los eventos históricos de deslizamiento.

Dicha capacidad es particularmente relevante porque los deslizamientos son fenómenos multicausales y no lineales, donde no existe una única variable que explique su ocurrencia, sino que dependen de la interacción simultánea de múltiples factores. A diferencia de los enfoques tradicionales, como los modelos determinísticos o simples análisis estadísticos bivariados, los algoritmos de aprendizaje automático utilizados en este estudio permiten identificar patrones no evidentes o relaciones que no siguen una lógica lineal directa, lo que mejora la calidad de las predicciones y aporta mayor profundidad a la comprensión del riesgo. En este sentido, el uso de modelos como Random Forest y XGBoost representa una ventaja metodológica importante, ya que ofrece una representación más realista del comportamiento del territorio frente a la amenaza de deslizamientos, lo cual es clave para tomar decisiones informadas en planificación y gestión del riesgo.

D. Importancia de variables predictoras.

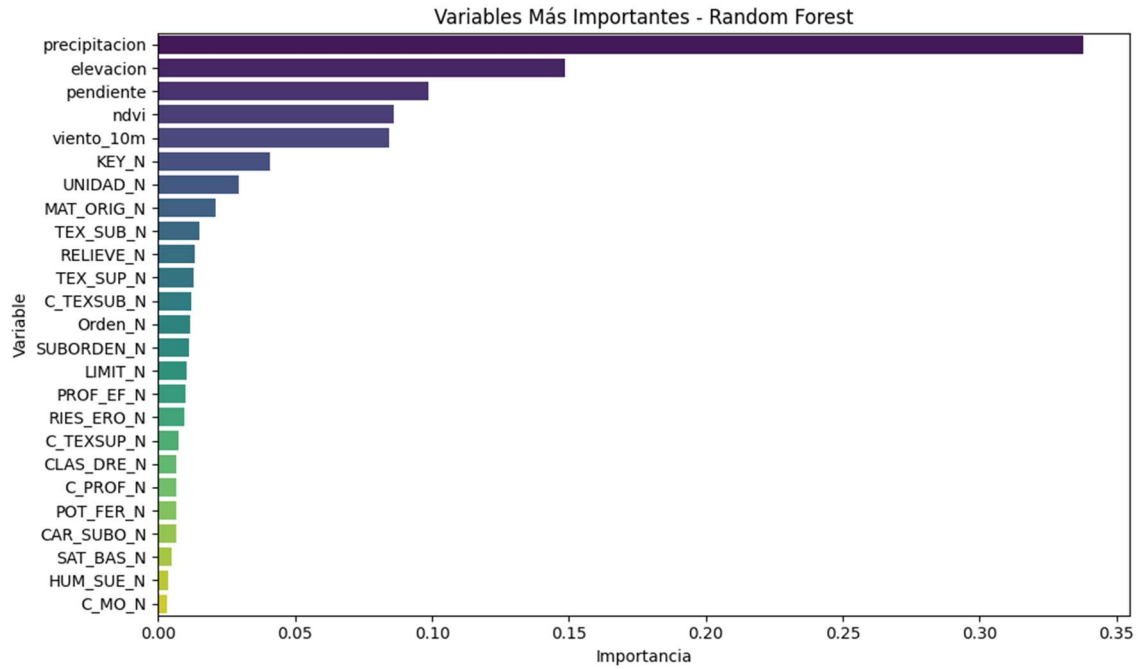


Figura 42. Histograma de importancias por variable - Modelo Random Forest.
(Fuente: elaboración propia).

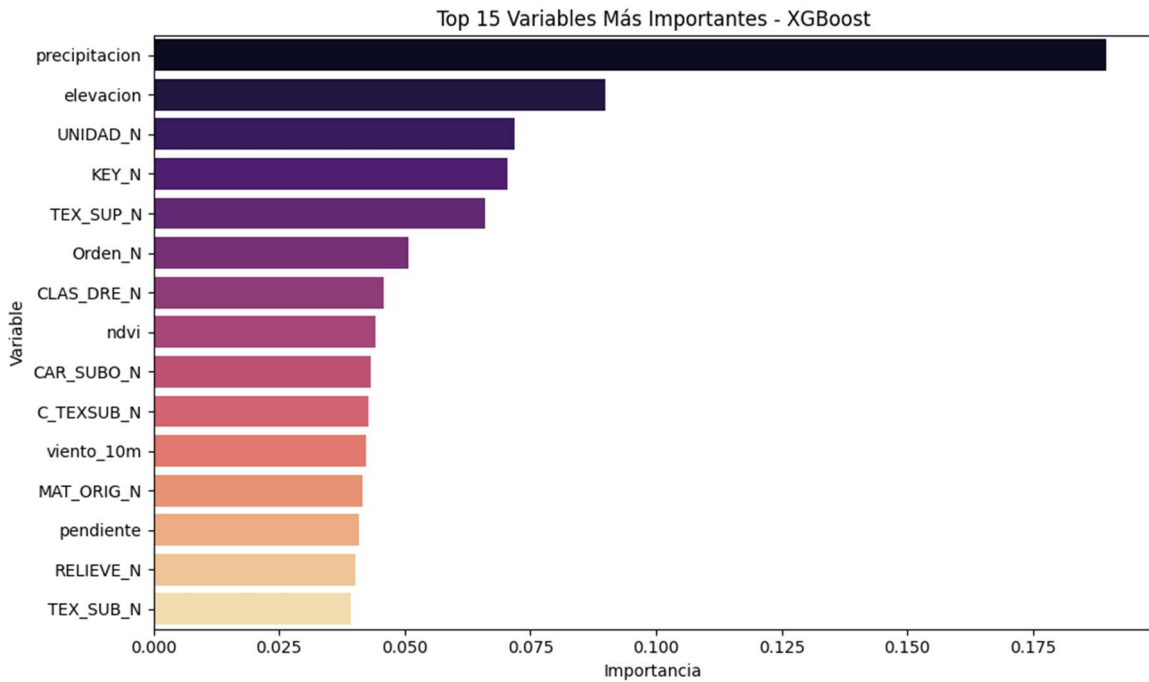


Figura 43. Histograma de importancias por variable - Modelo XGBoost.
(Fuente: elaboración propia).

El modelo XGBoost presenta una jerarquía de variables similar a la de Random Forest en términos generales, pero con una diferencia notable: otorga mayor peso a variables relacionadas con características edáficas y geológicas, tales como UNIDAD, KEY, TEX_SUP y CLAS_DRE. Estas variables, que corresponden a unidades geológicas, clasificación textural del suelo superficial y capacidad de drenaje del terreno, fueron identificadas como relevantes dentro del conjunto de predictores, lo que indica que el modelo logró establecer relaciones significativas entre ciertos tipos de suelo y la ocurrencia de deslizamientos.

Este comportamiento puede explicarse por la arquitectura interna de XGBoost, que optimiza sus predicciones mediante un proceso de gradient boosting y utiliza técnicas de regularización para reducir el sobreajuste, permitiéndole identificar patrones distribuidos entre variables incluso cuando su aporte individual es limitado (Chen & Guestrin, 2016). Esta capacidad le otorga una ventaja particular en el manejo de variables categóricas o discretas, las cuales, por su naturaleza, presentan múltiples clases con frecuencia no uniforme. Según Liu et al. (2021), XGBoost muestra un rendimiento superior al de otros modelos basados en árboles al trabajar con variables categóricas, ya que es capaz de optimizar de manera más precisa las divisiones dentro de cada árbol, mejorando así la sensibilidad del modelo frente a patrones más complejos y sutiles.

Asimismo, se destaca la posición del NDVI (Índice de Vegetación de Diferencia Normalizada) entre las variables más importantes en ambos modelos. Su relevancia como predictor puede atribuirse a que el NDVI actúa como un indicador indirecto del grado de cobertura vegetal, el cual está estrechamente vinculado con la estabilidad de las laderas. La vegetación contribuye al refuerzo del suelo mediante el anclaje de raíces, mejora la infiltración del agua y reduce el impacto directo de las lluvias sobre el terreno, lo cual significa una menor probabilidad de erosión y pérdida de masa superficial (Glade, 2003). La inclusión exitosa del NDVI entre las variables destacadas respalda hallazgos previos que demuestran su utilidad en estudios de susceptibilidad a deslizamientos en diversas regiones geográficas (Kavzoglu et al., 2014; Youssef et al., 2016). A partir de esto, es importante notar que, para poder aprovechar la información obtenida de los NDVI, se debe de contar con el conocimiento de las imágenes satelitales.

E. Análisis crítico de los resultados

El modelo Random Forest logró una precisión general (accuracy) del 76%, lo que significa que, en promedio, clasificó correctamente 76 de cada 100 puntos analizados. Además, alcanzó un F1score ponderado de 0.76, el cual combina la precisión (precision) y la sensibilidad (recall) en una sola métrica. Esta medida es especialmente útil cuando se trabaja con clases desbalanceadas, como es el caso de los deslizamientos, ya que penaliza los errores tanto por exceso como por defecto. Por otro lado, el valor del área bajo la curva ROC (AUC) fue de 0.835, lo que indica una muy buena capacidad para distinguir entre zonas con y sin deslizamiento. En general, valores de AUC mayores a 0.80 son considerados excelentes para tareas de clasificación

binaria, especialmente en contextos con alta variabilidad espacial y múltiples variables (Hosmer et al., 2013).

Con el uso de XGBoost, aunque mostró un desempeño ligeramente menor, también obtuvo resultados sólidos. Su precisión general fue del 74%, con un F1-score ponderado de 0.73 y un AUC de 0.821. Estos valores siguen estando dentro de un rango considerado adecuado para estudios de susceptibilidad geoespacial, y confirman que XGBoost es igualmente capaz de aprender patrones relevantes en los datos (Liu et al., 2021). La ligera diferencia respecto a Random Forest puede deberse a la forma en que cada modelo maneja la complejidad de los datos y las relaciones entre variables.

Una observación clave es que, en ambos modelos, el recall para la clase positiva (es decir, deslizamientos) fue superior a 0.84. Esto implica que ambos algoritmos fueron altamente sensibles para detectar correctamente la mayoría de las zonas donde sí ocurrieron deslizamientos en el pasado. Este resultado es particularmente relevante en el contexto de la gestión del riesgo, ya que cometer un error de tipo falso negativo (clasificar como segura una zona que en realidad es susceptible) puede tener consecuencias muy graves en términos de pérdidas humanas o materiales (Chung & Fabbri, 2003; Guzzetti et al., 2008).

Aunque la validación cruzada (cross-validation) es una técnica común para estimar el desempeño de modelos de clasificación, en este estudio se optó por una división estratificada de los datos en conjuntos de entrenamiento y prueba (70/30), dado que se contaba con una base de datos suficientemente representativa y balanceada entre clases. Esta decisión permitió simplificar el flujo de trabajo computacional y enfocarse en la comparación de modelos bajo un mismo esquema de evaluación. No obstante, se reconoce que el uso de validación cruzada podría aportar mayor robustez a las métricas, y se recomienda su aplicación en estudios futuros o en modelos a mayor escala.

En resumen, tanto Random Forest como XGBoost demostraron ser herramientas confiables para modelar la susceptibilidad a deslizamientos en un contexto geoespacial que toma en cuenta muchas consideraciones, ofreciendo un equilibrio razonable entre la identificación de zonas de riesgo y la minimización de errores de clasificación. Esta comparación funciona para evaluar qué modelo es más conservador en cuanto a la clasificación de áreas susceptibles, que puede promover una gestión territorial dentro del departamento o los municipios correspondientes de manera que se tomen las medidas necesarias con el fin de resguardar las vidas de las personas, edificios, infraestructura, etc.

F. Implicaciones prácticas

Los resultados obtenidos en este estudio permiten identificar las áreas críticas donde deben priorizarse intervenciones de mitigación de riesgo, tomando en cuenta tamaños geoespaciales más específicos o reducidos. En particular, las zonas clasificadas por los modelos como de muy alta susceptibilidad deben considerarse prioritarias no solo en términos de monitoreo constante, sino también para aplicar restricciones en el uso del suelo y normativas más estrictas para la

planificación de obras civiles. De igual forma, tomando en cuenta las formas de mitigar los deslizamientos, dichas normativas o regulaciones deberán especificar qué medidas preventivas se deberán considerar. Estas zonas suelen coincidir con laderas de fuerte pendiente, suelos de baja cohesión y antecedentes históricos de eventos, lo que justifica su tratamiento diferencial en políticas públicas (Guzzetti et al., 2005; Hong & Pradhan, 2019).

El hecho de que variables como la precipitación acumulada, la elevación, la pendiente del terreno y los parámetros edafológicos y geológicos dominen la explicación del modelo reafirma la importancia de adoptar un enfoque multidisciplinario en los sistemas de gestión de riesgos. La integración de datos geotécnicos, meteorológicos y topográficos no solo mejora la precisión de los modelos de susceptibilidad, sino que también fortalece los sistemas de alerta temprana, los cuales son fundamentales para reducir la exposición de las comunidades en riesgo (Pheng & Thapa, 2021; Vieira, Lacerda et al., 2022). Tomando esto en cuenta, es importante realizar registros geológicos, de suelos y de las formaciones de rocas en el sitio. Estos hallazgos resaltan la utilidad de mantener bases de datos actualizadas sobre precipitación y cobertura vegetal, así como realizar monitoreos periódicos en zonas con cambios en el uso del suelo o nuevas construcciones.

Por otro lado, los modelos desarrollados ofrecen la posibilidad de establecer umbrales de decisión con base en probabilidades objetivas de ocurrencia, los cuales pueden ser incorporados en herramientas de análisis multicriterio y en sistemas expertos. Esto permitiría alimentar de manera automática los planes de ordenamiento territorial a nivel municipal o departamental, fortaleciendo la gestión preventiva en lugar de reactiva. Esta vinculación entre modelos predictivos y planificación territorial representa una oportunidad clave para avanzar hacia políticas públicas basadas en evidencia, que prioricen la seguridad de las personas y la sostenibilidad del territorio (Guzzetti, 2012; Fernández & Mora, 2022).

Entre las limitaciones más relevantes se encuentra la resolución espacial de los datos, así como la disponibilidad de registros históricos de deslizamientos suficientemente representativos. Además, el uso de etiquetas binarias (deslizamiento / no deslizamiento) simplifica una realidad compleja que muchas veces presenta gradientes de susceptibilidad. Se debe verificar la viabilidad de tomar en cuenta el volumen de masa removida o desplazada.

También se reconoce que los modelos aún no incluyen factores antropogénicos como construcciones, deforestación reciente o alteraciones del drenaje, los cuales podrían aumentar la precisión del modelo si se incorporan adecuadamente.

G. Comparación con mapas institucionales

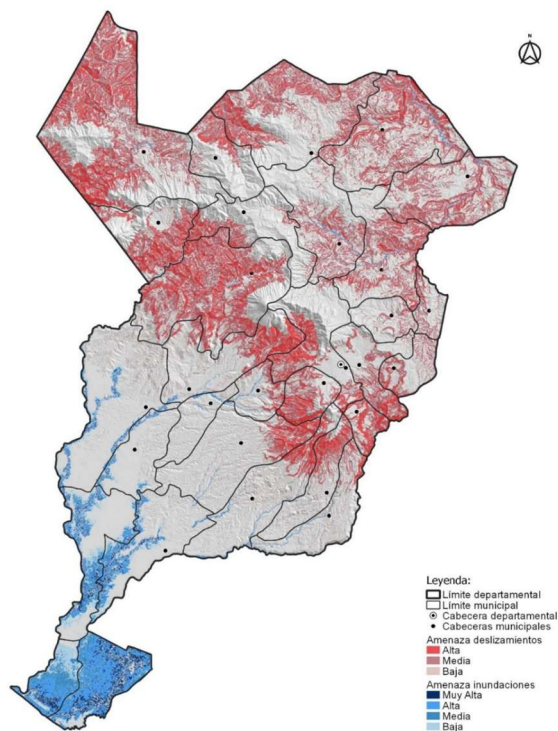


Figura 44. Mapa de vulnerabilidades principales del departamento de San Marcos, Guatemala (Fuente: SEGEPLAN, 2023).

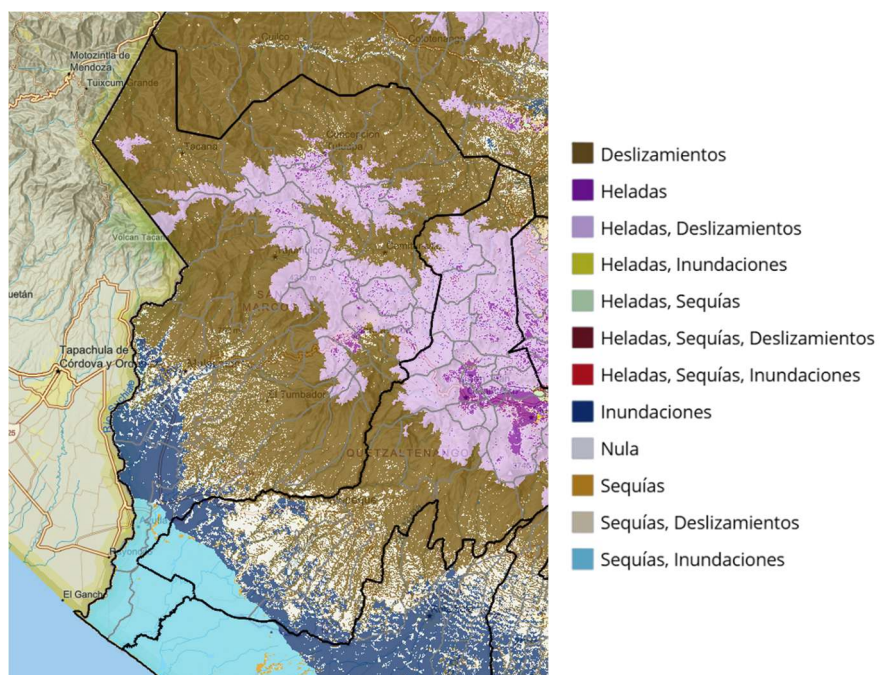


Figura 45. Mapa de amenazas de la República de Guatemala (enfocado a San Marcos). (Fuente: SIE – Universidad Landívar de Guatemala).

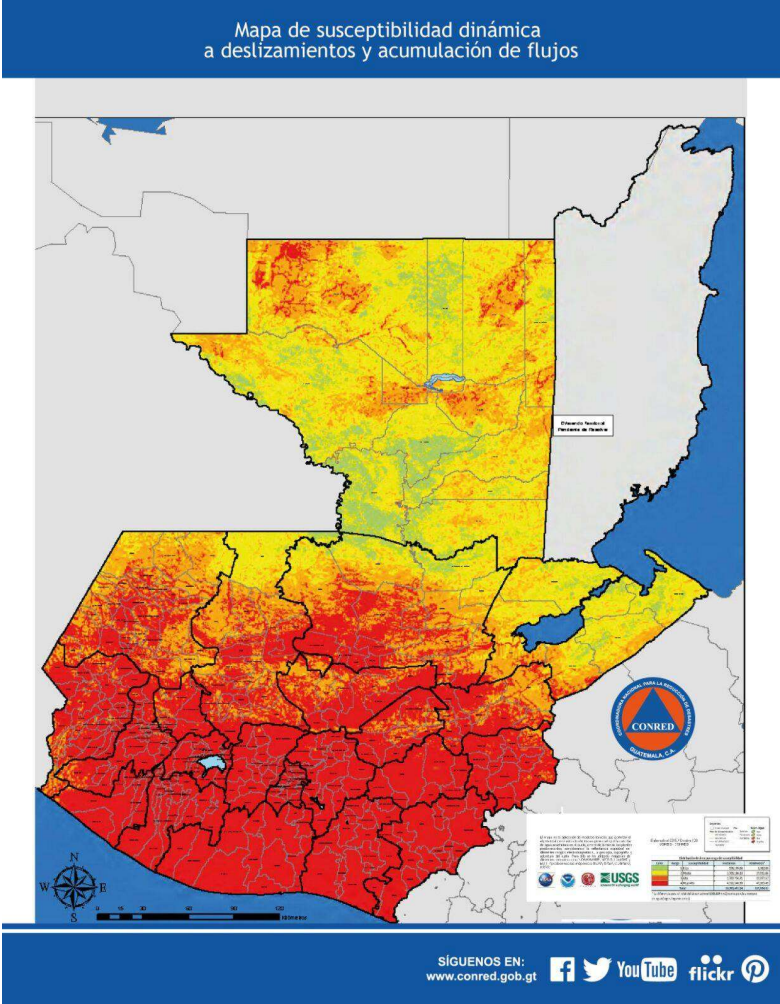


Figura 46. Mapa de susceptibilidad a deslizamientos del territorio de Guatemala (Fuente: CONRED, 2017).

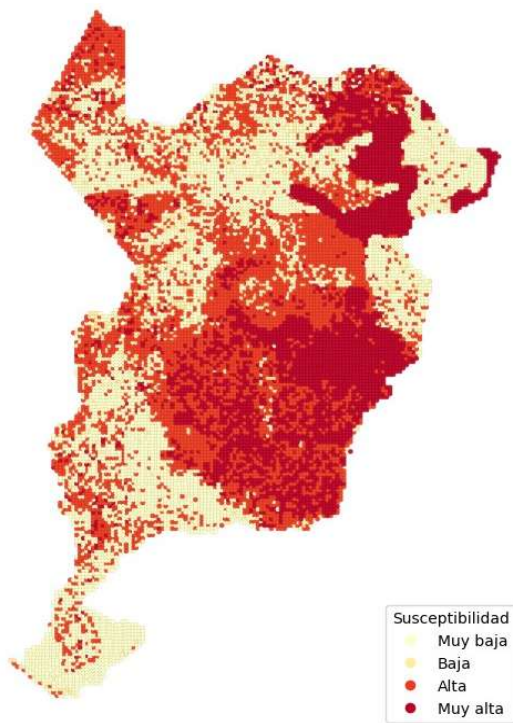


Figura 47. Mapa de probabilidad de deslizamientos empleando Random Forest.
(Fuente: Elaboración propia).

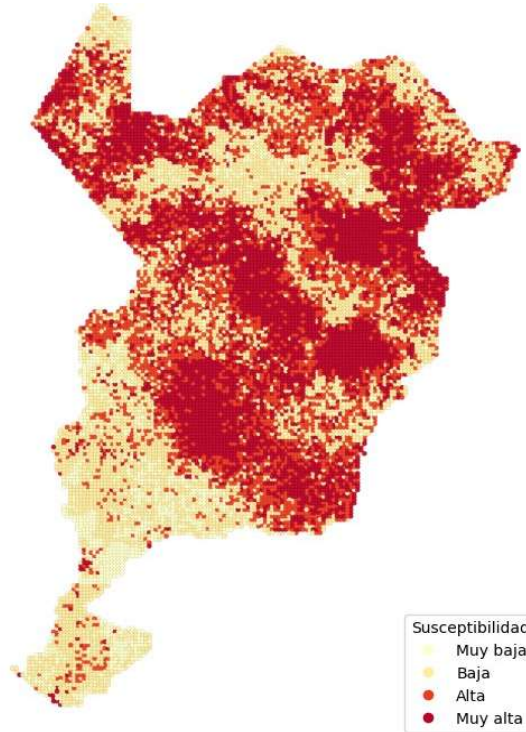


Figura 48. Mapa de predicción de deslizamientos empleando XGBoost.
(Fuente: Elaboración propia).

Como comparación, se está empleando el mapa de vulnerabilidades del departamento de San Marcos, Guatemala, elaborado por SEGEPLAN (Figura 44), y también se está utilizando el mapa de varias amenazas elaborado por el Sistema de Información Estratégica (SIE) de la Universidad Rafael Landívar (Figura X). Una revisión comparativa de los cinco mapas revela que existe un patrón espacial relativamente consistente respecto a las zonas de alta susceptibilidad a deslizamientos en la región centro-norte de San Marcos, particularmente en el corredor montañoso asociado a la Sierra Madre y a la cadena volcánica. Todos los mapas —incluidos los generados por modelos de aprendizaje automático (XGBoost y Random Forest)— destacan esa franja central como altamente vulnerable, lo que valida la relación entre elevaciones, pendientes abruptas y mayor probabilidad de remoción en masa. Esta coincidencia aporta robustez al análisis espacial y justifica la concentración de acciones de mitigación en dicha zona.

Sin embargo, al observar más detenidamente la esquina noroeste del departamento, donde se encuentra el municipio de Tacaná, se evidencian discrepancias importantes. El mapa de CONRED categoriza esa zona como de amenaza media, lo cual contrasta con los mapas de XGBoost, Random Forest, y el de SEGEPLAN, que la clasifican como de susceptibilidad alta o muy alta. Esta diferencia podría deberse a la resolución o antigüedad de los insumos de

CONRED, o bien a una subestimación del peligro en zonas de alta altitud donde históricamente ha habido deslizamientos, pero no han sido bien documentados.

Otra diferencia significativa se encuentra en la parte sur del departamento, correspondiente a la planicie costera del Pacífico. El mapa de CONRED muestra gran parte de esa zona con una susceptibilidad alta a deslizamientos, lo cual es inconsistente con todos los demás mapas. Los modelos de machine learning y el mapa de SEGEPLAN coinciden en clasificar esta región como de baja o muy baja susceptibilidad, lo cual es coherente desde un punto de vista geotécnico, ya que esta área se caracteriza por pendientes suaves, suelos aluviales y menor recurrencia histórica de deslizamientos. Esta sobreestimación por parte de CONRED podría explicarse por el uso de unidades administrativas grandes o criterios generales de clasificación a nivel nacional.

Respecto al mapa multi amenaza del SIE, si bien no detalla la intensidad de la amenaza, sí coincide con los otros mapas al indicar una alta presencia de eventos combinados (deslizamientos, heladas e inundaciones) en la parte central del departamento. Esta superposición de amenazas valida indirectamente los resultados de los modelos predictivos, que también identifican la región centro-oriental como de alta susceptibilidad. Sin embargo, el enfoque del SIE es más cualitativo y categórico, por lo que su utilidad es complementaria al enfoque probabilístico de los modelos de clasificación.

En resumen, los cinco mapas convergen en resaltar la vulnerabilidad de las regiones montañosas del centro y norte del departamento. No obstante, existen discrepancias importantes en áreas como el extremo noroeste (subestimado por CONRED) y el sur costero (posiblemente sobreestimado por CONRED). Estas diferencias ponen de manifiesto la necesidad de integrar enfoques complementarios: por un lado, los mapas institucionales respaldados por datos históricos y criterios normativos; y por otro, los modelos predictivos que aprovechan la precisión espacial y la capacidad analítica de los algoritmos de aprendizaje automático. Es importante resaltar que no se especifican los métodos empleados para la elaboración de dichos mapas, por lo que las discrepancias pueden ser provocadas por enfoques o métodos cualitativos en vez de cuantitativos.

VIII. CONCLUSIONES

- La recopilación e integración de datos geospaciales, ambientales y meteorológicos permitió construir un insumo robusto y coherente para el análisis territorial. Dicha estructuración es clave para modelar fenómenos complejos como los deslizamientos de tierra, resaltando la importancia de contar con información precisa y frecuentemente actualizada.
- La generación de un mapa de susceptibilidad a deslizamientos por medio de algoritmos de árboles de decisiones fue eficaz para identificar patrones entre múltiples variables y clasificar zonas de susceptibilidad en un territorio con alta heterogeneidad geográfica como el departamento de San Marcos.
- La comparación visual entre los resultados obtenidos y los mapas oficiales reveló coincidencias en áreas de alta susceptibilidad, pero también puso en evidencia diferencias relevantes. Estas discrepancias reflejan variaciones en métodos, escalas y enfoques conceptuales adoptados por distintas fuentes.
- El análisis de las variables empleadas en el modelo subrayó el papel determinante de factores como la pendiente, la geología y la cobertura vegetal en la ocurrencia de deslizamientos. Esto reafirma la necesidad de adoptar un enfoque multivariable para lograr una evaluación de riesgo más precisa y útil en la gestión territorial.
- La elección de un enfoque departamental permitió entrenar modelos con datos más homogéneos en términos geográficos y climáticos, lo cual favoreció la detección de patrones más representativos y específicos del territorio en estudio.
- Los modelos de machine learning representan una herramienta eficaz para la gestión del riesgo y la planificación territorial, especialmente en regiones montañosas como San Marcos, donde la ocurrencia de eventos de remoción en masa es frecuente y peligrosa.

IX. RECOMENDACIONES

- Es fundamental mantener actualizado y ampliar el inventario de deslizamientos, incorporando información detallada sobre el tipo de movimiento, profundidad, volumen y causas asociadas. Esta actualización periódica permitirá entrenar modelos más robustos y mantener vigente el mapa de susceptibilidad como herramienta de gestión.
- Incorporar variables antropogénicas como el tipo de cobertura urbana, cambios recientes en el uso de suelo y trazado de infraestructura podría mejorar significativamente la precisión de los modelos.
- La implementación de umbrales de decisión basados en la probabilidad de deslizamiento puede fortalecer los sistemas de alerta temprana y alimentar de forma automatizada los sistemas de información territorial.
- La metodología utilizada podría aplicarse a otros departamentos de Guatemala con características topográficas, climáticas y edafológicas comparables a las de San Marcos, como Quetzaltenango, Huehuetenango o Alta Verapaz. No obstante, se recomienda ajustar el modelo al contexto local, asegurando que las variables y pesos sigan siendo representativos.
- Dado que los patrones de deslizamiento pueden cambiar con el tiempo debido a variaciones en el uso del suelo, el clima o las intervenciones humanas, es recomendable actualizar periódicamente el inventario de deslizamientos y reentrenar el modelo con datos recientes. Esta estrategia garantizaría que el mapa de susceptibilidad se mantenga vigente como herramienta de gestión.
- Para capturar variaciones del terreno que influyen en la ocurrencia de deslizamientos, podría ser útil el uso de imágenes satelitales de muy alta resolución o productos derivados de sensores LiDAR, cuando estén disponibles. Esto permitiría afinar la escala de análisis y hacer evaluaciones a nivel parcelario o comunitario.
- Más allá de divisiones político-administrativas, una línea prometedora consiste en aplicar el modelo a regiones homogéneas en términos climáticos o fisiográficos. Esto permitiría mejorar la generalización del modelo, al entrenarlo en patrones que comparten dinámicas ambientales similares, facilitando así su aplicabilidad en zonas con escasos datos históricos.
- Promover la capacitación de técnicos municipales y personal de instituciones de gestión del riesgo en el uso de herramientas de análisis geoespacial y machine learning, fomentando una cultura de prevención y uso de evidencia técnica.
- Desarrollar protocolos de interoperabilidad entre el modelo de susceptibilidad generado y los sistemas de gestión de información de CONRED, como el SIGRID (Sistema de Información Geográfica para la Reducción del Riesgo de Desastres). Esto permitiría integrar automáticamente los mapas de susceptibilidad en los flujos de análisis institucionales,

facilitando la toma de decisiones en tiempo real, la priorización de zonas críticas y la activación de alertas tempranas basadas en umbrales probabilísticos.

X. BIBLIOGRAFÍA

- Aggarwal, C. C. (2014). *Data Classification: Algorithms and Applications*. CRC Press.
- Aliaga, G. (2017). Cartografía digital: fundamentos y aplicaciones con SIG.
- Argentina, T. (2023). Mapas vectoriales y ráster: diferencias y aplicaciones en SIG.
- Berman, E., & Popov, I. (2020). *Soil Mechanics Fundamentals*. CRC Press.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- Blight, G. E. (2013). *Geotechnical Engineering for Mine Waste Storage Facilities*. CRC Press.
- Brady, N. C., & Weil, R. R. (2008). *The Nature and Properties of Soils* (14th). Pearson Prentice
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Burrough, P. A., & McDonnell, R. A. (1998). *Principles of Geographical Information Systems*. Oxford University Press.
- Center for Disease Control, C., & (CDC), P. (2022). Landslides and Mudslides <https://www.cdc.gov/disasters/landslides.html>
- Chacón Montero, J. (2003). Riesgos de origen geológico y geomorfológico: deslizamientos de tierras, identificación, análisis y prevención de sus consecuencias. *Revista de Ciencias Sociales*, (23), 33-64.
- Chang, K.-t. (2018). *Introduction to Geographic Information Systems* (9.^a ed.). McGraw-Hill Education.
- Chang, K.-T. (2019). *GIS Fundamentals: A First Text on Geographic Information Systems* (6th). XanEdu Publishing.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system, 785-794.
- Chen, W., & Zhao, R. (2020). Comparative analysis of landslide susceptibility models using GIS and machine learning in China. *Remote Sensing*.
- Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1), 6. <https://doi.org/10.1186/s12864-019-6413-7>
- Coduto, D. P., Yeung, M.-c. R., & Kitch, W. A. (2011). *Geotechnical Engineering: Principles and Practices* (2.^a ed.). Prentice Hall.
- Comisión Económica para América Latina y el Caribe (CEPAL), C. E. (1991). *Manual para la evaluación de los efectos socioeconómicos de los desastres naturales*. CEPAL.
- CONRED. (2022). Acuerdo Número 06-2022: Actualización del protocolo para evaluación de daños y necesidades (EDAN) y zonas de riesgo [Guatemala, 18 de mayo de 2011 (actualización 2022)].
- Cruden, D. M., & Varnes, D. J. (1996). *Landslide types and processes* (A. K. Turner & R. L. Schuster, Eds.). National Academy Press.
- Das, B. M., & Sobhan, K. (2016). *Principles of Geotechnical Engineering* (9.^a ed.). Cengage Learning.

- Dávila Martínez, F., & Camacho Arranz, E. (2012). Georreferenciación de documentos cartográficos para la gestión de Archivos y Cartotecas. <https://doi.org/https://www.ign.es/web/resources/docs/IGNCnig/CTC-Ibercarto-V-Georreferenciacion.pdf#:~:text=1.,diferente%20al%20que%20se%20encuentra>.
- Denyer, P. (2009). *Geología y geotectónica de América Central y el Caribe*. Recuperado de <https://app.ingemmet.gob.pe/biblioteca/pdf/CPG14-010.pdf>
- Environmental Systems Research Institute (ESRI). (2022). What is GIS? <https://www.esri.com/en-us/what-is-gis/overview>
- EPSG Geodetic Parameter Registry. (2024). EPSG Geodetic Parameter Dataset. <https://epsg.org>
- ESRI. (2021). GIS for Transportation. <https://www.esri.com/en->
- Fell, R., Ho, K. K. S., Lacasse, S., & Leroi, E. (2008). A framework for landslide risk assessment and management. En O. Hungr, R. Fell, R. Couture & E. Eberhardt (Eds.), *Landslides: Risk Analysis and Management* (pp. 3-25). CRC Press.
- Fernández, L., & Mora, K. (2022). Evaluación de impactos económicos de deslizamientos en comunidades rurales de Centroamérica. *Revista Centroamericana de Geografía*, 18(1), 45-61.
- for Disaster Risk Reduction (UNDRR), U. N. O. (2015). Sendai Framework for Disaster Risk Reduction 2015–2030. <https://www.undrr.org/publication/sendai-framework-disaster-riskreduction-2015-2030>.
- Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (2nd). O'Reilly Media.
- Gijón Ferrer, M. (1988). Deslizamientos, Desprendimientos, Flujos y Avalanchas. *Riesgos Geológicos*, 175-185.
- González Bolaños, F., Olguín López, J. L., Guevara Gutiérrez, R. D., Palomera García, C., Ramírez Romero, J. M., & Villalvazo López, V. M. (2023). Identificación espacial del riesgo por deslizamiento de tierra en el municipio de Autlán de Navarro, Jalisco, México. *Revista Geográfica de América Central*, (70), 349-372. <https://doi.org/10.15359/rgac.70-1.13>
- Goodchild, M. F. (1993). The State of GIS for Environmental Problem-Solving. En *Environmental Modeling with GIS* (pp. 8-15). Oxford University Press.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT press.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18-27. <https://doi.org/10.1016/j.rse.2017.06.031>
- Gui, J., Alejano, L. R., Yao, M., Zhao, F., & Chen, W. (2023). GIS-Based Landslide Susceptibility Modeling: A Comparison between Best-First Decision Tree and Its Two Ensembles (BagBFT and RFBFT). *Remote Sensing*, 15(4), 1007. <https://doi.org/10.3390/rs15041007>
- Guzzetti, F., Cardinali, M., Reichenbach, P., & Carrara, A. (2005). The impact of landslides in the central Apennines, Italy. *Natural Hazards and Earth System Sciences*, 5(1), 123-135. <https://doi.org/10.5194/nhess-5-123-2005>

- Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M., & Ardizzone, F. (2005a). Landslide hazard assessment and risk evaluation: limits and prospects. *Natural Hazards and Earth System Sciences*, 5(6), 833-845.
- Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M., & Ardizzone, F. (2005b). Probabilistic landslide hazard assessment at the basin scale. *Geomorphology*, 72(1-4), 272-299. <https://doi.org/10.1016/j.geomorph.2005.06.002>
- Hall.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd). Springer.
- He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263-1284.
- Ho, K., & Barratt, D. (2005). *Soil nailing: Best practice guidance*. CIRIA.
- Hofmann-Wellenhof, B., Lichtenegger, H., & Wasle, E. (2011). *Introduction to Geospatial Technologies*. Springer.
- Hong, H., & Pradhan, B. (2019). Landslide susceptibility mapping using random forest and XGBoost models in the Andes of Colombia. *Catena*, 182, 104231. <https://www.ogc.org/standards/geopackage>
- Huabin, W., Gangjun, L., Weiya, X., & Gonghui, W. (2005). Landslide hazard and risk assessment using GIS and remote sensing. *Engineering Geology*, 76(1-2), 1-12. <https://doi.org/10.1016/j.enggeo.2004.06.006>
- INSIVUMEH. (2020). *Climatología general de Guatemala*. Instituto Nacional de Sismología, Vulcanología, Meteorología e Hidrología. Recuperado de: <https://www.insivumeh.gob.gt/meteorologia/climatologia.htm>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer.
- Keefer, D. K. (1984). Landslides caused by earthquakes. *Geological Society of America Bulletin*, 95(4), 406-421.
- Koerner, R. M. (2012). *Designing with Geosynthetics* (6.^a ed.). Xlibris Corporation.
- Kramer, S. L. (1996). *Geotechnical Earthquake Engineering*.
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3), 18-22.
- Lo, C. P., & Yeung, A. K. W. (2007). *Concepts and Techniques of Geographic Information Systems* (2.^a ed.). Prentice Hall.
- Londoño, R. D. (1999). Los sistemas de información georeferenciada - SIG. <https://doi.org/https://ciencia.lasalle.edu.co/cgi/viewcontent.cgi?article=1637&context=ruls>
- Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2015). *Geographic Information Science and Systems* (4.^a ed.). Wiley.
- Maguire, D. J., Goodchild, M. F., & Rhind, D. W. (2005). *Geographic Information Systems and Science* (2.^a ed.). John Wiley & Sons.

- Ministerio de Agricultura, Ganadería y Alimentación (MAGA). (2011). *Mapa de fisiografía de Guatemala*. <https://www.maga.gob.gt/download/fisiografia.pdf>
- Ministerio de Energía y Minas. (2015). *Catálogo de recursos geotérmicos de Guatemala*. <https://www.mem.gob.gt/wp-content/uploads/2012/04/Catalogo-final...versi%C3%B3n-12032015.pdf>
- Ministerio de Medio Ambiente y Recursos Naturales (MARN). (2004). *Memoria técnica para el mapa de susceptibilidad de deslizamientos de tierra en El Salvador* (inf. téc.).
- Mora, S., & Vahrson, W. (1994). Metodología de Macrozonificación para Deslizamientos de Tierra Determinación de peligro. <https://doi.org/10.2113/gseegeosci.xxxi.1.49>
- Moya, J., & Corominas, J. (1997). Landslides: ¿deslizamientos o movimientos del terreno? Definición, clasificaciones y terminología. *Investigaciones geográficas. IV Simposio Nacional sobre Taludes y Laderas Inestables. Granada. Vol. 1*, 199-212.
- Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.
- Negrete, R. A. (2023). Curso de Sistemas de Información Geográfica en línea (básico, intermedio y avanzado) como parte de las actividades de apoyo a la investigación arqueológica en la Coordinación Nacional de Arqueología INAH. <https://ru.dgb.unam.mx/bitstream/20.500.14330/TES01000847620/3/0847620.pdf>
- O. G. C. (2021). OGC GeoPackage Encoding Standard. <https://docs.qgis.org>
- Olaya, V. (2014). *Sistemas de Información Geográfica: Análisis Espacial y Modelos en SIG*. Universitat de Girona.
- Pheng, K., & Thapa, R. (2021). Landslide susceptibility mapping using machine learning: A comparative study in the Himalayas. *Geoscience Frontiers*, 12(4), 101111.
- Prinz, D., Malik, R. S., & Gupta, A. (2020). Bioengineering measures for slope stabilization and erosion control in hill regions. *International Soil and Water Conservation Research*, 8(2), 153-161. <https://doi.org/10.1016/j.iswcr.2020.04.005>
- Project, T. S. (2018). *The Sphere Handbook: Humanitarian Charter and Minimum Standards in Humanitarian Response* (4.^a ed.). Practical Action Publishing.
- Pucha, J. (2021). Modelos de datos espaciales: ráster y vectorial.
- Quesada Román, A., & Boraschi, S. (2018). Comparación de la Metodología Mora-Vahrson y el Método Morfométrico para Determinar Áreas Susceptibles a Deslizamientos en la Microcuenca del Río Macho, Costa Rica. <https://doi.org/dx.doi.org/10.15359/rgac.61-2.1>
- Reis, S., Huebl, J., & Sassa, K. (2008). *Rainfall-induced landslides: Mechanisms, monitoring techniques and nowcasting models for early warning systems*. CRC Press.
- República de Guatemala, C. (1996). Decreto Legislativo 109-96: Ley de la Coordinadora Nacional para la Reducción de Desastres [Publicado en el Diario de Centro América].
- Rivas, V., & Remondo, J., J and Bonachea. (2016). Análisis de las relaciones entre precipitación y actividad de deslizamientos en el pasado reciente (NO Guipúzcoa). *Geo-Temas*, 633-636.
- Robinson, R. G. (2022). *Slope Stability and Stabilization Methods* (3.^a ed.). John Wiley & Sons.
- Rodríguez, J. (2006). *Mecánica de suelos en la ingeniería práctica*. Limusa.

- Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, *10*(3), e0118432.
- Santos Preciado, J. (2010). *Geografía, espacio y poder: Introducción a los Sistemas de Información Geográfica*. Universidad de Guadalajara.
- Sarría, F. A. (2014). Sistemas de Información Geográfica, 239. <https://www.um.es/geograf/sigmur/sigpdf/temario.pdf>
- SEGEPLAN. (2018). Guía para el análisis y gestión del riesgo en proyectos de inversión pública. <https://www.segeplan.gob.gt>
- SEGEPLAN. (2021). *Informe Guatemala 2021: Cambio Climático*. Secretaría de Planificación y Programación de la Presidencia. Recuperado de https://portal.segeplan.gob.gt/segeplan/wp-content/uploads/2023/08/Informe_Guatemala2021_CC.pdf
- Servicio Nacional de Estudios Territoriales. San Salvador. <https://app.ingemmet.gob.pe/biblioteca/pdf/Amb-21.pdf>
- Smith, I., & Fleming, K. S. (2014). *Site Investigation: A Handbook for Engineers* (4.^a ed.). Thomas Telford Publishing.
- Smith, K., & Mullins, C. E. (2013). *Soil Survey and Land Evaluation*. Longman Scientific Technical.
- Suarez Días, J. (1998). Deslizamientos y Estabilidad de Taludes en Zonas Tropicales. <https://doi.org/https://desastres.medicina.usac.edu.gt/documentos/docgt/pdf/spa/doc0101/doc0101.pdf>
- Team, Q. D. (2023). QGIS User Guide: Working with Project Files. https://docs.qgis.org/3.40/en/docs/user_manual/introduction/project_files.html
- Tipos de Movimientos en Masa – Clasificaciones. (2007). *Proyecto Multinacional Andino: Geociencias para las Comunidades Andinas. 2007. Movimientos en Masa en la Región Andina: Una guía para la evaluación de amenazas.*, 1-31.
- UNDRR. (2019). Words into Action Guidelines: National Disaster Risk Assessment. <https://www.undrr.org/us/industries/transportation/overview>
- Van Westen, C. J., et al. (2006). Spatial data for landslide susceptibility. *Engineering Geology*, *102*(3–4), 112–131.
- Varnes, D. J. (1978). *Slope movement types and processes* (R. L. Schuster & R. J. Krizek, Eds.). National Academy of Sciences.
- Vieira, D. R., Lacerda, W. A., et al. (2022). A review of machine learning techniques applied to landslide susceptibility mapping. *Geomatics, Natural Hazards and Risk*, *13*(1), 123-150.
- Weather Spark. (s.f.). Clima promedio en San Marcos, Guatemala durante todo el año. <https://es.weatherspark.com/y/11157/Clima-promedio-en-SanMarcos-Guatemala-durante-todo-el-a%C3%B1o>
- Wen, B., Zhang, L., & Liang, G. (2020). Stability analysis and reinforcement measures of landslides in mountainous areas: A review. *Engineering Geology*, *270*, 105572. <https://doi.org/10.1016/j.enggeo.2020.105572>
- Wieczorek, G. F. (1996). Landslide triggering mechanisms (A. Turner & R. Schuster, Eds.), 76-90.

Zheng, Y., Liu, Y., Yuan, J., & Xie, X. (2011). Urban Computing with Taxicabs. *Proceedings of the 13th International Conference on Ubiquitous Computing (UbiComp)*, 89-98.

XI. APÉNDICES

A. Script para obtención de variables predictoras desde Google Earth Engine

```
// Cargar puntos desde Assets
var puntos = ee.FeatureCollection("projects/ee-anakv2001/assets

/GEE-Puntos_estudio");

// FUNCIONES AUXILIARES DE NDVI

// Sentinel-2
function getNDVI_S2(geom, fecha) {
  var start = fecha.advance(-15, 'day');
  var end = fecha.advance(15, 'day');
  var s2 = ee.ImageCollection('COPERNICUS/S2_SR')
    .filterBounds(geom)
    .filterDate(start, end)
    .filter(ee.Filter.lt('CLOUDY_PIXEL_PERCENTAGE', 20))
    .map(function(img) {
      var ndvi = img.normalizedDifference(['B8',
'B4']).rename('NDVI');
      var scl = img.select('SCL');
      var sclMask =
scl.eq(4).or(scl.eq(5)).or(scl.eq(6)).or(scl.eq(7));
      return ndvi.updateMask(sclMask);
    });
  var ndviImage = s2.mean();
  var ndviDict = ndviImage.reduceRegion({
    reducer: ee.Reducer.first(),
    geometry: geom,
    scale: 10,
    maxPixels: 1e13
  });
  return ee.Algorithms.If(ndviDict.contains('NDVI'),
ndviDict.get('NDVI'), null);
}

// Landsat 8
function getNDVI_L8(geom, fecha) {
  var start = fecha.advance(-15, 'day');
  var end = fecha.advance(15, 'day');
  var l8 = ee.ImageCollection('LANDSAT/LC08/C02/T1_L2')
    .filterBounds(geom)
    .filterDate(start, end)
    .map(function(img) {var ndvi = img.normalizedDifference(['SR_B5',
'SR_B4']).rename('NDVI');
      var qa = img.select('QA_PIXEL');
```

```

        var mask = qa.bitwiseAnd(1 << 3).eq(0);
        return ndvi.updateMask(mask);
    });
    var ndviImage = l8.mean();
    var ndviDict = ndviImage.reduceRegion({
        reducer: ee.Reducer.first(),
        geometry: geom,
        scale: 30,
        maxPixels: 1e13
    });
    return ee.Algorithms.If(ndviDict.contains('NDVI'),
    ndviDict.get('NDVI'), null);
}

// Landsat 9
function getNDVI_L9(geom, fecha) {
    var start = fecha.advance(-15, 'day');
    var end = fecha.advance(15, 'day');
    var l9 = ee.ImageCollection('LANDSAT/LC09/C02/T1_L2')
        .filterBounds(geom)
        .filterDate(start, end)
        .map(function(img) {
            var ndvi = img.normalizedDifference(['SR_B5',
'SR_B4']).rename('NDVI');
            var qa = img.select('QA_PIXEL');
            var mask = qa.bitwiseAnd(1 << 3).eq(0);
            return ndvi.updateMask(mask);
        });
    var ndviImage = l9.mean();
    var ndviDict = ndviImage.reduceRegion({
        reducer: ee.Reducer.first(),
        geometry: geom,
        scale: 30,
        maxPixels: 1e13
    });
    return ee.Algorithms.If(ndviDict.contains('NDVI'),
    ndviDict.get('NDVI'), null);
}

// Landsat 7
function getNDVI_L7(geom, fecha) {
    var start = fecha.advance(-15, 'day');
    var end = fecha.advance(15, 'day');
    var l7 = ee.ImageCollection('LANDSAT/LE07/C02/T1_L2')
        .filterBounds(geom)
        .filterDate(start, end)
        .map(function(img) {
            var ndvi = img.normalizedDifference(['SR_B4',
'SR_B3']).rename('NDVI');
            var qa = img.select('QA_PIXEL');

```

```

        var mask = qa.bitwiseAnd(1 << 3).eq(0);
        return ndvi.updateMask(mask);
    });
var ndviImage = l7.mean();
var ndviDict = ndviImage.reduceRegion({
    reducer: ee.Reducer.first(),
    geometry: geom,
    scale: 30,
    maxPixels: 1e13
});
return ee.Algorithms.If(ndviDict.contains('NDVI'),
ndviDict.get('NDVI'), null);
}

// MAPEO CON VARIABLES

var puntosProcesados = puntos
    .filter(ee.Filter.notNull(['Fecha']))
    .map(function(feature) {
        var rawFecha = feature.getString('Fecha');
        var fechaEvento = ee.Date(rawFecha);
        var geom = feature.geometry();
        var coords = geom.coordinates();

        var dem = ee.Image("USGS/SRTMGL1_003");
        var elev = dem.select("elevation");
        var slope = ee.Terrain.slope(dem);

        var elevVal = elev.reduceRegion({
            reducer: ee.Reducer.first(),
            geometry: geom,
            scale: 30,
            maxPixels: 1e13
        }).get("elevation");

        var slopeVal = slope.reduceRegion({
            reducer: ee.Reducer.first(),
            geometry: geom,
            scale: 30,
            maxPixels: 1e13
        }).get("slope");

        var lluviaCol = ee.ImageCollection("UCSB-CHG/CHIRPS/DAILY")
            .filterDate(fechaEvento.advance(-7, 'day'), fechaEvento)
            .select("precipitation");

        var lluvia = ee.Image(lluviaCol.reduce(ee.Reducer.sum()));
        var lluviaVal = lluvia.reduceRegion({
            reducer: ee.Reducer.first(),
            geometry: geom,

```

```

    scale: 5000,
    maxPixels: 1e13
  }).get("precipitation_sum");

var ndviS2 = getNDVI_S2(geom, fechaEvento);
var ndviL8 = getNDVI_L8(geom, fechaEvento);
var ndviL9 = getNDVI_L9(geom, fechaEvento);
var ndviL7 = getNDVI_L7(geom, fechaEvento);

var ndviRandom = ee.Number(Math.random() * 0.25 + 0.25);

var ndviVal = ee.Algorithms.If(
  ndviS2,
  ee.Algorithms.If(ee.Number(ndviS2).eq(0), ndviRandom, ndviS2),
  ee.Algorithms.If(
    ndviL8,
    ndviL8,
    ee.Algorithms.If(
      ndviL9,
      ndviL9,
      ee.Algorithms.If(
        ndviL7,
        ndviL7,
        ndviRandom
      )
    )
  )
);

var vientoCol = ee.ImageCollection("ECMWF/ERA5_LAND/DAILY_AGGR")
  .filterDate(fechaEvento, fechaEvento.advance(1, 'day'))
  .select(["u_component_of_wind_10m",
    "v_component_of_wind_10m"]);

var viento = vientoCol.mean();

var velocidad = viento.expression(
  'sqrt(pow(u, 2) + pow(v, 2))', {
    u: viento.select('u_component_of_wind_10m'),
    v: viento.select('v_component_of_wind_10m')
  }
).rename('velocidad_viento');

var vientoVal = velocidad.reduceRegion({
  reducer: ee.Reducer.first(),
  geometry: geom,
  scale: 1000,
  maxPixels: 1e13
}).get('velocidad_viento');

```

```

return feature.set({
    'LAT': coords.get(1),
    'LONG': coords.get(0),
    'fecha_str': fechaEvento.format('yyyy-MM-dd'),
    'elevacion': elevVal,
    'pendiente': slopeVal,
    'precipitacion_7d': lluviaVal,
    'ndvi': ndviVal,
    'viento_10m': vientoVal
});
});

//Exportación

Export.table.toDrive({
    collection: puntosProcesados,
    description: 'Puntos_Prediccion_ConVariables',
    fileFormat: 'CSV'
});

```

B. Código para generación de mapa de susceptibilidad usando Random Forest

Dicho código genera lo siguiente:

- 1) Modelo de predicción usando Random Forest.
- 2) Reporte de clasificación
- 3) Importancia por variable
- 4) Matriz de confusión
- 5) Mapa de calor de probabilidad de deslizamiento El código es el siguiente:

```

# 1. Importar librerías necesarias
import pandas as pd import geopandas as gpd import matplotlib.pyplot as
plt from shapely.geometry import Point from sklearn.ensemble import
RandomForestClassifier from sklearn.model_selection import
train_test_split from sklearn.metrics import classification_report,
confusion_matrix, roc_auc_score, RocCurveDisplay from matplotlib.lines
import Line2D

# 2. Cargar el dataset de entrenamiento
df = pd.read_csv("Inventario_final_2.csv") #Inventario con puntos de no
deslizamientos y deslizamientos

X = df.drop(columns=["landslide_val"]) #Se usan todas las columnas
excepto la de entrenamiento "landslide_val" y = df["landslide_val"] #Esta
es la variable de entrenamiento

```

```

#3. División en entrenamiento y prueba (70/30)
# Debido a que es una base de entrenamiento pequeño, no se usa 80/20 de
manera que el grupo de prueba sea más robusto.
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
random_state=42, stratify=y)

# 4. Entrenar modelo
Random Forest modelo = RandomForestClassifier(n_estimators=100,
random_state=85) #Se están utilizando 100 árboles (n_estimators=100), se
puede usar entre 100 y 500 recomendablemente.

#la semilla aleatoria es de 85 pero se puede usar cualquier número.
modelo.fit(X_train, y_train)

# 5. Evaluación
y_pred = modelo.predict(X_test) y_proba = modelo.predict_proba(X_test)[:,
1] print("\Matriz de Confusión:\n", confusion_matrix(y_test, y_pred))

#Obtener datos de matriz de confusión print
("Reporte de Clasificación:\n", classification_report(y_test, y_pred))

#Obtener rep print
"ROC AUC Score:", roc_auc_score(y_test, y_proba))

#Obtener ROC AUC Score

# 6. Guardar métricas
pd.DataFrame(classification_report(y_test, y_pred,
output_dict=True)).T.to_csv("Resultados_RF.csv", index=True)
#Generar reporte de clasificación en archivo csv.

# 7. Guardar modelo import joblib
joblib.dump(modelo, "Modelo_RF.pkl")

# 8. Predecir sobre nuevos puntos
df_pred = pd.read_csv("SM_Puntos_prediccion.csv")

#Puntos generados dentro polígono de San Marcos en grilla de 500m.
coords = df_pred[["LAT", "LONG"]]
X_pred = df_pred.drop(columns=["FID", "LAT", "LONG"],
errors="ignore") #Ignorar columnas de LAT y LONG, sirven para #poder
reproducir el mapa de resultado pero no para el análisis.

df_pred["probabilidad_deslizamiento"] = modelo.predict_proba(X_pred)[:,
1]

```

```

# 9. Crear GeoDataFrame geometry = [Point(xy) for xy in
zip(df_pred["LONG"], df_pred["LAT"])] gdf_pred =
gpd.GeoDataFrame(df_pred, geometry=geometry, crs="EPSG:4326")

# 10. Clasificación para visualización de acuerdo a probabilidad de
deslizamiento def clasificar(prob):
    if prob < 0.25: return "Muy baja"
    elif prob < 0.5: return "Baja" elif prob < 0.75:
        return "Alta"
    else: return "Muy alta" gdf_pred["clase"] =
gdf_pred["probabilidad_deslizamiento"].apply(clasificar)

colores = {
    "Muy baja": "#ffffcc",
    "Baja": "#ffeda0",
    "Alta": "#f03b20",
    "Muy alta": "#bd0026"
}
#Asignar colores dependiendo de probabilidad de deslizamiento, escala de
amarillo (baja) a rojo (alta).

# 11. Graficar el mapa fig, ax = plt.subplots(figsize=(10, 10))
gdf_pred.plot(ax=ax, color=gdf_pred["clase"].map(colores), markersize=10,
alpha=0.9, edgecolor='black', linewidth=0.05)

legend_elements = [Line2D([0], [0], marker='o', color='w', label=label,
markerfacecolor=color, markersize=8)
    for label, color in colores.items()]
ax.legend(handles=legend_elements, title="Susceptibilidad", loc="lower
right")

ax.set_title("Mapa de Susceptibilidad a Deslizamientos\n(Random Forest)",
fontsize=14) ax.axis("off") plt.tight_layout()
plt.savefig("Mapa_susceptibilidad_RF.png", dpi=300) plt.show()

# 12. Exportar resultados
gdf_pred.drop(columns="geometry").to_csv("Predicciones_RF.csv",
index=False)

# 13. Generar histograma de importancias por Random Forest
import seaborn as sns df = pd.read_csv("Importancia_Variables_RF.csv")

plt.figure(figsize=(10, 6)) sns.barplot(x="Importancia", y="Variable",
data=df, palette="viridis") plt.title("Variables Más Importantes - Random
Forest") plt.tight_layout() plt.savefig("Importancia_RF.png", dpi=300)
plt.show()

```

C. Código para generación de mapa de susceptibilidad usando XGBoost

Dicho código genera lo siguiente:

1. Modelo de predicción usando Random Forest.
2. Reporte de clasificación.
3. Histograma de importancia por variable.
4. Matriz de confusión.
5. Mapa de calor de probabilidad de deslizamiento.

El código es el siguiente:

```
# 1. Importar librerías necesarias
import pandas as pd
import geopandas as gpd
import matplotlib.pyplot as plt
from shapely.geometry import Point
from xgboost import XGBClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report, confusion_matrix,

roc_auc_score, RocCurveDisplay
from matplotlib.lines import Line2D
import seaborn as sns

# 2. Cargar el dataset de entrenamiento
df = pd.read_csv("SM_Inventario_final.csv")
X = df.drop(columns=["landslide_val"])
y = df["landslide_val"]

# 3. División en entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,

# 3. División en entrenamiento y prueba (70/30) - debido a que es una base
de entrenamiento pequeño, no se usa 80/20 de manera que el grupo de prueba
sea más robusto.

random_state=35, stratify=y)

# 4. Entrenar modelo XGBoost
model = XGBClassifier(n_estimators=100, use_label_encoder=False,

eval_metric="logloss", random_state=63)
model.fit(X_train, y_train)
#De igual forma, se utilizan 100 árboles y una semilla aleatoria.
```

```

# 5. Evaluación
y_pred = model.predict(X_test)
y_proba = model.predict_proba(X_test)[: , 1]

print("\nMatriz de Confusión:\n", confusion_matrix(y_test, y_pred))
print("\nReporte de Clasificación:\n", classification_report(y_test,
y_pred))
print("\nROC AUC Score:", roc_auc_score(y_test, y_proba))
#Esto dará en el Shell los valores buscados para validar el modelo.

# 6. Guardar métricas
pd.DataFrame(classification_report(y_test, y_pred,
output_dict=True)).T.to_csv("Resultados_XGBoost.csv", index=True)

# 7. Guardar modelo
model.save_model("Modelo_XGBoost.json")

# 8. Predecir sobre nuevos puntos
df_pred = pd.read_csv("SM_Puntos_prediccion.csv")
coords = df_pred[["LAT", "LONG"]]
X_pred = df_pred.drop(columns=["FID", "LAT", "LONG"], errors="ignore")

df_pred["probabilidad_deslizamiento"] = model.predict_proba(X_pred)[: , 1]
df_pred["prediccion_landslide"] = model.predict(X_pred)

# 9. Crear GeoDataFrame
geometry = [Point(xy) for xy in zip(df_pred["LONG"], df_pred["LAT"])]
gdf_pred = gpd.GeoDataFrame(df_pred, geometry=geometry, crs="EPSG:4326")

# 10. Clasificar para visualización
def clasificar(prob):
    if prob < 0.25:
        return "Muy baja"
    elif prob < 0.5:
        return "Baja"
    elif prob < 0.75:
        return "Alta" # Media se agrupa con Alta
    else:
        return "Muy alta"

gdf_pred["clase"] =
gdf_pred["probabilidad_deslizamiento"].apply(clasificar)

colores = {
    "Muy baja": "#ffffcc",
    "Baja": "#ffeda0",
    "Alta": "#f03b20",

```

```

    "Muy alta": "#bd0026"
}

# 11. Graficar
fig, ax = plt.subplots(figsize=(10, 10))
gdf_pred.plot(ax=ax, color=gdf_pred["clase"].map(colores),

markersize=10, alpha=0.9, edgecolor='black', linewidth=0.05)

legend_elements = [Line2D([0], [0], marker='o', color='w', label=label,
                           markerfacecolor=color, markersize=8)
                    for label, color in colores.items()]
ax.legend(handles=legend_elements, title="Susceptibilidad", loc="lower
right")

ax.set_title("Mapa de Susceptibilidad a Deslizamientos (XGBoost)",

fontsize=14)
ax.axis("off")
plt.tight_layout()
plt.savefig("Mapa_susceptibilidad_XGBoost.png", dpi=300)
plt.show()

# 12. Exportar resultados
gdf_pred.drop(columns="geometry").to_csv("Predicciones_XGBoost.csv",

index=False)

# Obtener importancias de las variables para determinar que variables
influyen más

importancias = model.feature_importances_
df_importancia = pd.DataFrame({
    "Variable": X.columns,
    "Importancia": importancias
}).sort_values(by="Importancia", ascending=False)

# Exportar como CSV
df_importancia.to_csv("Importancia_Variables_XGB.csv", index=False)
df_importancia.head(15).to_csv("Top15_Importancia_Variables_XGB.csv",

index=False)

# Graficar
plt.figure(figsize=(10, 6))
sns.barplot(x="Importancia",y="Variable", data=df_importancia.head(15),

```

```
palette="magma")
plt.title("Top 15 Variables Más Importantes - XGBoost")
plt.tight_layout()
plt.savefig("Importancia_XGB.png", dpi=300)
plt.show()
```


XII. GLOSARIO

- **Aprendizaje no supervisado:** Método en el que el modelo analiza los datos sin conocer previamente las salidas, buscando patrones ocultos.
- **Aprendizaje supervisado:** Tipo de Machine Learning donde el modelo se entrena con datos de entrada y salida ya conocidos.
- **Área bajo la curva ROC (AUC):** Indicador que evalúa qué tan bien un modelo distingue entre clases.
- **Coordenadas geográficas (SRC):** Sistema que permite ubicar cualquier punto en la Tierra mediante latitud y longitud.
- **Deslizamiento:** Movimiento descendente de una masa de tierra o roca por una pendiente, causado por factores naturales o humanos.
- **Evaluación de riesgos:** Análisis de la probabilidad y consecuencias de un evento peligroso, como un deslizamiento.
- **F1-score:** Métrica que combina precisión y sensibilidad en un solo valor equilibrado.
- **Factores detonantes:** Elementos que provocan o aceleran la ocurrencia de un deslizamiento, como lluvias intensas, sismos o actividad humana.
- **Factores permanentes:** Factores que son inherentes de algún punto, como pendiente, elevación, tipo de suelo, litografía, etc.
- **Georreferenciación:** Proceso de asignar coordenadas reales a un mapa o imagen para ubicarlo correctamente en el espacio geográfico.
- **Inventario de deslizamientos:** Registro geográfico que identifica y describe la ubicación, tipo, extensión y fecha (cuando se conoce) de los deslizamientos ocurridos en una zona determinada. Es una herramienta clave para el análisis del riesgo y la planificación territorial.
- **Machine Learning:** Técnica de inteligencia artificial que permite a los sistemas aprender automáticamente a partir de datos.
- **Mapa ráster:** Representación espacial basada en una cuadrícula de píxeles o celdas, donde cada celda contiene un valor que representa información, como color, elevación o temperatura. Es ideal para representar imágenes satelitales o datos continuos como el relieve del terreno.
- **Mapa vectorial:** Tipo de representación espacial que utiliza geometrías como puntos, líneas y polígonos para describir la ubicación y forma de elementos geográficos. Es útil para mostrar datos precisos como carreteras, límites políticos o redes de infraestructura.
- **Matriz de confusión:** Tabla que muestra los aciertos y errores de un modelo de clasificación, evaluando su rendimiento.
- **Modelos de datos espaciales:** Formas de organizar y representar datos geográficos, como modelos vectoriales (puntos, líneas, polígonos) o raster (celdas).
- **Precisión:** Medida que indica cuántas de las predicciones positivas del modelo fueron correctas.
- **Random Forest:** Algoritmo de aprendizaje automático basado en árboles de decisión que mejora la precisión mediante el uso de múltiples árboles.

- **Sensibilidad (Recall):** Porcentaje de verdaderos positivos identificados correctamente por el modelo.
- **Sistemas de Información Geográfica (SIG):** Herramientas digitales que permiten capturar, analizar y visualizar datos espaciales para la toma de decisiones.
- **XGBoost:** Algoritmo de aprendizaje automático que optimiza modelos mediante la técnica de gradiente reforzado, siendo eficaz y preciso.