



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Metodología para obtener distancias de viaje de movimientos en masa en la Región Andina Nororiental con base en métodos probabilísticos

Jose David Ortiz Contreras

Universidad Nacional de Colombia
Facultad de Ingeniería, Departamento de Ingeniería Civil y Agrícola
Bogotá D.C, Colombia

2024

Metodología para obtener distancias de viaje de movimientos en masa en la Región Andina Nororiental con base en métodos probabilísticos

Jose David Ortiz Contreras

Trabajo final de maestría presentado como requisito parcial para optar al título de:

Magister en Geotecnia

Director:

Ph.D. Mauricio Alberto Tapias Camacho

Codirector:

Magister Daniel Camilo Román Quintero

Línea de Investigación:

Análisis de confiabilidad y riesgos asociados al entorno geotécnico.

Grupo de Investigación:

GENKI

Universidad Nacional de Colombia

Facultad de Ingeniería, Departamento de Ingeniería Civil y Agrícola

Bogotá D.C, Colombia

2024

A las cuatro mujeres que siempre me han apoyado:

Mi madre, la persona que me enseñó el valor del trabajo duro y a no desfallecer frente a la adversidad.

Mi abuela, aquella que con una sonrisa y un abrazo me mostro lo valioso de un corazón noble, puro y servicial con los demás.

Mi hermana, fuente de inspiración y energía para avanzar día a día y demostrar que vale la pena luchar

Mi amada, cómplice y amiga, que me sigue brindando de manera desinteresada su compañía y amor dándome el impulso que necesito para mejorar cada día.

“Somos lo que hacemos repetidamente. La excelencia, entonces, no es un acto, es un hábito”

Aristóteles

Declaración de obra original

Yo declaro lo siguiente:

He leído el Acuerdo 035 de 2003 del Consejo Académico de la Universidad Nacional. «Reglamento sobre propiedad intelectual» y la Normatividad Nacional relacionada al respeto de los derechos de autor. Esta disertación representa mi trabajo original, excepto donde he reconocido las ideas, las palabras, o materiales de otros autores.

Cuando se han presentado ideas o palabras de otros autores en esta disertación, he realizado su respectivo reconocimiento aplicando correctamente los esquemas de citas y referencias bibliográficas en el estilo requerido.

He obtenido el permiso del autor o editor para incluir cualquier material con derechos de autor (por ejemplo, tablas, figuras, instrumentos de encuesta o grandes porciones de texto).

Por último, he sometido esta disertación a la herramienta de integridad académica, definida por la universidad.

Jose David Ortiz Contreras

Fecha 22/10/2024

Agradecimientos

A mi alma mater, la Universidad Nacional de Colombia que, como a muchos otros, me dio la oportunidad de formarme como profesional y forjarme con los más altos estándares de calidad.

A mi director, el Dr. Mauricio Alberto Tapias Camacho y al Ing. Daniel Camilo Román Quintero quienes con su experiencia y dedicación supieron guiarme en la elaboración de este trabajo final de maestría.

Al proyecto de investigación “Método para la estimación de distancias de viaje de movimientos en masa con base en aproximaciones geométricas” en participación con la Universidad Nacional de Colombia, la Universidad Industrial de Santander, el Ministerio de Ciencias y el Servicio Geológico Colombiano, su labor fue fundamental al brindarme los recursos de información y conocimientos necesarios para el desarrollo de este trabajo final de maestría.

Al grupo de investigación Geotechnical Engineering Knowledge and Innovation (GENKI) donde se brindan espacios de discusión sobre la investigación y la importante labor de la geotecnia en el país.

Finalmente destaco también la labor docente del Ing. Julio Estaban Colmenares Montañez, del Ing. Félix Hernández Rodríguez y el Ing. Carlos Eduardo Rodríguez Pineda, quienes compartieron conmigo todo su conocimiento de la mejor manera posible, brindándome herramientas para solucionar los retos presentados en el desarrollo de este trabajo.

Resumen

La geografía colombiana presenta grandes desafíos debido a la variabilidad de su topografía y los materiales presentes, dando como resultado uno de los países con mayor densidad de procesos de remoción en masa en la región, siendo valioso cada avance en el entendimiento y el análisis de estos procesos. El presente documento busca dar un paso más en el cumplimiento de este objetivo, planteando una metodología que parte de la obtención y categorización de la información, pasando por un análisis de validación estadística y finalizando en un modelo de *Machine Learning* que sirve como alternativa a las curvas de regresión utilizadas actualmente para el cálculo de la distancia de viaje.

El resultado se compone de un algoritmo que puede usarse dependiendo del nivel de procesamiento de la información que tenga el usuario, un modelo de *Random Forest* elaborado en R para facilitar su uso y calibración junto con una curva de probabilidad de excedencia de distancia de viaje aplicada a un caso de estudio. Estos productos presentan una mejoría tomando como base los modelos utilizados actualmente, aumentando su precisión y exactitud al disminuir el error relativo dando una alternativa sólida de cálculo.

Palabras clave: Distancia de viaje, probabilidad de excedencia, Random Forest.

Abstract

Methodology to obtain landslide's run-out distances in Northeast Andean Region based in probabilistic methods

Colombia has one of the most challenger geography owing the variability of its topography and existing materials, becoming in one of the countries with the highest density of landslide processes in the region, each advance to understand and analyze these processes are extremely valuable. This document seeks to contribute another step to reach this objective, proposing a methodology that starts from obtaining and categorizing information, passing through a statistical validation analysis and finally using a machine learning model that serves as an alternative of regression curves currently used for travel distance calculation.

The result is made up of an algorithm that can be used depending on the user's level of information processing, a Random Forest model developed in R to make easier its use and calibration along with a run out distance exceedance probability curve applied directly to a study case. These products make an improvement based on the models currently used, increasing their precision and accuracy by reducing the relative error, providing a solid calculation alternative.

Keywords: Run out distance, exceedance probability, Random Forest.

Contenido

	Pág.
Resumen	IX
Abstract	X
Lista de figuras	XIV
Lista de tablas	XVII
1 Introducción	19
1.1 Planteamiento del problema	19
1.2 Objetivos	21
1.2.1 Objetivo General.....	21
1.2.2 Objetivos Específicos.....	21
1.3 Metodología de trabajo	21
1.4 Estructura del documento	23
2 Marco Teórico	27
2.1 Movimientos en masa y distancias de viaje	27
2.1.1 Definición y clasificación de los movimientos en masa	27
2.1.2 Modelo planta-perfil típico para un movimiento en masa.	30
2.1.3 Metodologías empíricas para obtener una única distancia de viaje.....	31
2.2 Elementos para el análisis de estadístico de bases de datos	35
2.2.1 Media o promedio muestral	35
2.2.2 Cuartil 1, cuartil 3 y mediana.....	35
2.2.3 Valor mínimo y valor máximo.....	35
2.2.4 Varianza o covarianza muestral.....	36
2.2.5 Vector de coeficientes de variación.....	36
2.2.6 Distancia de Mahalanobis.....	36
2.2.7 Análisis por Componentes Principales.....	37
2.3 Método de Random Forest	38
2.3.1 Árbol de decisión	38
2.3.2 Random Forest.....	40
2.3.3 Error en el Random Forest	41
3 Área de estudio	43

3.1	Delimitación de la zona de estudio	43
3.2	El Servicio de Información de Movimientos en Masa (SIMMA) como punto de partida	44
3.3	Subdivisión de la zona de estudio	45
3.4	Metodología para la elaboración de la base de datos.....	49
3.4.1	Definición de la muestra	49
3.4.2	Definición de la campaña exploratoria.....	53
3.4.3	Medición de variables morfométricas.	55
4	Estimación de la distancia de viaje de movimientos en masa con base en aproximaciones geométricas.....	65
4.1	Corte talud.....	65
4.2	Modelo de Scheidegger (1973)	67
4.3	Modelo de Nicoletti & Sorriso(1991)	68
4.4	Modelo de Corominas.	69
4.5	Ecuaciones propuestas por la Guía Metodológica.....	71
4.6	Román et al. (2024).....	72
5	Análisis estadístico de la base de datos	75
5.1	Análisis descriptivo de la base de datos	75
5.1.1	Valores estadísticos fundamentales	75
5.1.2	Matrices de correlación	77
5.1.3	Transformación haciendo uso de función logarítmica.....	81
5.2	Distancia de Mahalanobis	86
5.3	Análisis por Componentes Principales	89
5.4	Comentarios finales.....	91
6	Estimación de la distancia de viaje de movimientos en masa con técnicas machine learning.....	93
6.1	Modelo generado	94
6.2	Variación del error	95
6.3	Comentarios finales.....	99
7	Probabilidad de excedencia de la distancia de viaje	101
7.1	Modelo propuesto para el cálculo de excedencia de la distancia de viaje con base en un algoritmo de <i>Random Forest</i>	101
7.2	Comparación de metodologías para el cálculo de probabilidades de excedencia de distancias de viaje	105
8	Metodología para la estimación de distancias de viaje de movimientos en masa	107
8.1	Fase 1: Recolección, levantamiento y ensamblaje de información	107
8.2	Fase 2: Verificación de información en la zona de estudio	109
8.3	Fase 3: Validación de los modelos a utilizar en el cálculo de la distancia de viaje.....	111

8.4 Fase 4. Evaluación de la probabilidad de excedencia de la distancia de viaje .	113
9 Conclusiones y recomendaciones	117
9.1 Conclusiones	117
9.2 Recomendaciones	118
Bibliografía	119
Anexos	123
A. Resumen información primaria	123
B. Código en R utilizado para realizar el análisis descriptivo.....	131
C. Código en R para modelo de Random Forest y simulaciones	135
D. Recopilación de información primaria – Fichas técnicas.....	145

Lista de figuras

	Pág.
Figura 2-1. Modelo planta perfil asumido Tomado de la Guía Metodológica (Ávila, y otros, 2016).	30
Figura 2-2. Modelo de regresión lineal entre el volumen y el ángulo de eficiencia del movimiento.	33
Figura 2-3. Comparación de correlación entre variables y la distancia euclidiana.	37
Figura 2-4. Esquema del funcionamiento de los árboles de decisión.	38
Figura 3-1. Ubicación de la zona estudio. Los puntos resaltados indican los 466 eventos de movimientos en masa considerados luego del muestreo por conglomerados.	44
Figura 3-2. Zonificación de la zona estudio.	45
Figura 3-3. Identificación de las partes principales de la superficie de falla. Evento Z6_20_5 (Carcasí, Santander).....	56
Figura 3-4. Generación de las curvas de nivel. Evento Z6_20_5 (Carcasí, Santander)...	57
Figura 3-5. Identificación de las partes principales del movimiento, zona de superficie de falla (rojo), zona de depósito (morado). Evento Z6_20_5 (Carcasí, Santander)	58
Figura 3-6. Toma de las medidas del área de la superficie de falla (A_p) y su ancho (W_r) en m^2 y m respectivamente. Evento Z6_20_5 (Carcasí, Santander)	60
Figura 3-7. Toma de la longitud de la superficie de falla (L_r) y su profundidad (D_r). Evento Z6_20_5 (Carcasí, Santander).....	61
Figura 3-8. Toma de la altitud de la corona de la superficie de falla, ángulo de cambio de dirección y tipo de obstáculo. Evento Z6_20_5 (Carcasí, Santander)	63
Figura 3-9. Toma de la magnitud del ángulo de talud prefalla, altura máxima del movimiento y distancia de viaje. Evento Z6_20_5 (Carcasí, Santander)	64
Figura 4-1. Distribución del error relativo para los 50 movimientos en masa utilizando el Modelo Corte Talud.	66

Figura 4-2. Distribución del error relativo utilizando el modelo Sheidegger para los 50 movimientos en masa.....	68
Figura 4-3. Distribución del error relativo utilizando el modelo (Nicolett & Sorriso, 1991)para los 50 movimientos en masa.	69
Figura 4-4. Distribución del error relativo utilizando el modelo de Corominas (1996) para los 50 movimientos en masa.	70
Figura 4-5. Distribución del error relativo con el modelo de regresión lineal categorizado.	72
Figura 4-6. Distribución del error regresión lineal con el volumen utilizando el modelo de Román y otros (2024) para los 50 movimientos en masa.	73
Figura 5-1. Diagrama de cajas y bigotes con las variables transformadas haciendo uso de los 50 movimientos muestreados.	82
Figura 5-2. Ortofoto del movimiento Z7_21 con las zonas principales identificadas.	83
Figura 5-3. Correlación entre variables con la base de datos original mostrada de manera gráfica haciendo uso de los 50 movimientos muestrados.....	83
Figura 5-4. Correlación entre las variables mostrada de manera gráfica luego de aplicar una transformación utilizando la función Log+1.....	85
Figura 5-5. Distribución de la distancia de Mahalanobis comparada con una distribución Chi-cuadrado de la base de datos.....	87
Figura 5-6. Distribución de la distancia de Mahalanobis comparada con una distribución Chi-cuadrado con la base de datos transformada usando la función Log+1.....	88
Figura 5-7. Importancia de las componentes principales.	90
Figura 5-8. Biplot mostrando la importancia de cada una de las variables originales.	91
Figura 6-1. Variación del error con respecto al No. de árboles.	94
Figura 6-2. Variación del error con respecto a las variables tomadas en cada nodo.	96
Figura 6-3. Variación del error con respecto al número máximo de nodos.	97
Figura 6-4. Variación del error con respecto al tamaño de la muestra de entrenamiento.	98
Figura 6-5. Error relativo del calculado con el Random Forest. haciendo uso de los 50 movimientos muestrados.....	99
Figura 7-1. Medición de Lr, Dr y Wr en el tercio medio de la superficie de falla. Movimiento Z11_37 (Girón, Santander).....	103
Figura 7-2. Probabilidad de Excedencia de distancia de Viaje. Movimiento Z11_CE1. (Tona, Santander)	105

Figura 7-3. Comparativa de la probabilidad de excedencia para los diferentes modelos.	106
Figura 8-1. Fase 1. Diagrama de flujo. Recolección de información.....	109
Figura 8-2. Fase 2. Validación de la información de la zona de estudio.	111
Figura 8-3. Fase 3. Validación del modelo.	113
Figura 8-4. Zonas de amenaza de movimientos en masa basados en la probabilidad de excedencia de la distancia de viaje.....	114

Lista de tablas

	Pág.
Tabla 2-1. Coeficientes de regresión lineal para cada tipo de movimiento.....	32
Tabla 3-1. Subzonas de la zona de estudio por conglomerados.....	46
Tabla 4-1. Media y Rango Intercuartílico (RI) aplicando el modelo Corte Talud a los 50 movimientos en masa del estudio.....	66
Tabla 4-2. Modelo de regresión lineal categorizado por tipo de movimiento para el cálculo de la distancia de viaje.	71
Tabla 5-1. Propiedades estadísticas para cada variable haciendo uso de los 50 movimientos muestrados.....	76
Tabla 5-2. Matriz de covarianzas de las variables haciendo uso de los 50 movimientos muestrados.	78
Tabla 5-3. Coeficientes de variación calculados para cada una de las variables geométricas haciendo uso de los 50 movimientos muestrados.	78
Tabla 5-4. Matriz de correlación de las variables haciendo uso de los 50 movimientos muestrados.	80
Tabla 5-5. Coeficientes de variación luego de transformar las variables independientes con la función Log+1 haciendo uso de los 50 movimientos muestrados.	81
Tabla 6-1. Muestra de los 10 modelos con el menor error.....	95
Tabla 8-1. Errores relativos del caso de estudio para cada modelo estudiado.	112
Tabla 8-2. Categorías de amenaza establecidas para el ejemplo de acuerdo con las...	115

1 Introducción

Este trabajo final de maestría estará enmarcado dentro del proyecto de investigación “Método para la estimación de distancias de viaje de movimientos en masa con base en aproximaciones geométricas” aprobado en la convocatoria 852 de 2019 de MINCIENCIAS (antiguo COLCIENCIAS), ente financiador, y ejecutado por la Universidad Nacional de Colombia (registrado bajo el código HERMES 46907), en compañía de la Universidad Industrial de Santander y el Servicio Geológico Colombiano.

Durante el desarrollo de este documento se busca plantear una metodología que permita al lector calcular la probabilidad de excedencia para una distancia de viaje dada, haciendo uso de métodos y técnicas de toma de información, análisis estadístico, modelos de regresión lineal, *machine learning* y simulación de casos, logrando dar una alternativa adicional y fortaleciendo las metodologías existentes. El desarrollo de este trabajo se basa en una información obtenida en campo a través de levantamientos topográficos con vuelos no tripulados que permitieron cubrir grandes áreas con una alta precisión abordando principalmente propiedades geométricas de los movimientos estudiados.

En este capítulo se presenta el planteamiento del problema, los objetivos del trabajo de grado, una metodología para cumplir dichos objetivos y una estructura del documento que permite al lector tener una guía para facilitar su lectura.

1.1 Planteamiento del problema

Debido a la alta densidad de fenómenos de remoción en masa que se presentan anualmente en el territorio colombiano donde aproximadamente el 80% de la población se encuentra expuesto a una amenaza alta o muy alta UNGRD(2020) es necesario que la comunidad científica desarrolle constantemente técnicas que permitan conocer más estos eventos, utilizando herramientas sustentadas teóricamente y con uso práctico para los

análisis de estabilidad de los suelos y macizos rocosos, pero esto es solo una parte del desafío, también se requiere estimar la distancia de viaje del material movilizado con el más alto nivel de certidumbre posible.

Las propuestas existentes para el cálculo de distancias de viaje están basadas en eventos de otras latitudes y aún teniendo metodologías empíricas como las propuestas por la Guía Metodológica del Servicio Geológico Colombiano (Ávila, y otros, 2016), se ha comprobado que no solamente el volumen desplazado es la única variable que afecta las distancias de viaje y que las expresiones propuestas hasta la fecha basadas en regresiones lineales podrían ser mejoradas utilizando bases de datos locales, adicionalmente, al haber incertidumbre al establecer las magnitudes para diferentes variables en un movimiento no se puede obtener un único valor para la distancia de viaje.

Otra dificultad que se debe superar es que los modelos basados en correlaciones no calculan el valor aislado de la distancia de viaje L , en su gran mayoría el resultado es una relación conocida como ángulo de eficiencia del movimiento (definido como la relación entre la altura máxima del movimiento H y L) y el volumen, pero H es desconocida hasta tanto se dé el desplazamiento del material, lo que hace depender a L de un valor desconocido ya que para un mismo ángulo de eficiencia del movimiento existen infinitos valores de H y L , estas dimensiones se encuentran representadas con mayor detalle en el capítulo 2.1.2.

El SIMMA (Sistema de Información de Movimientos en Masa) tiene registrados 1503 movimientos catalogados como flujos y deslizamientos únicamente en Santander, Norte de Santander y Boyacá, lo que muestra la alta densidad de eventos que pueden causar riesgos a la población residente en estos departamentos, causando impactos económicos y sociales, no solamente en el instante en el que se da el proceso de remoción, también afecta a largo plazo por consecuencias secundarias como cierres temporales o pérdidas de vías de comunicación, obstrucción de cuerpos de agua almacenando energía que pueden desencadenar en una segunda amenaza para la población, familias económicamente vulnerables, etc.

La metodología planteada en el presente trabajo final de maestría puede utilizarse como insumo para sistemas de alerta temprana y mapas de zonificación, los cuales son fundamentales en la toma de decisiones de autoridades locales ubicadas en la zona de estudio, destinando recursos a la evaluación y prevención de desastres, la evacuación de damnificados o en un caso más técnico, el diseño de obras lineales, lo que ayudaría a reducir los riesgos provocados por los fenómenos de remoción en masa.

1.2 Objetivos

1.2.1 Objetivo General

Evaluar la aplicación de una metodología para el cálculo de distancias de viaje de movimientos en masa a partir de análisis probabilístico.

1.2.2 Objetivos Específicos

1. Establecer una muestra de movimientos en masa tipo flujo y deslizamiento a partir de una base de datos de eventos ocurridos en una zona de estudio ubicada en la región Andina Nororiental de Colombia.
2. Evaluar la relevancia de las variables medibles en los movimientos en masa estudiados en campo sobre la distancia de viaje de los movimientos en masa en una zona de estudio.
3. Aplicar un modelo estadístico para la estimación de las distancias de viaje de movimientos en masa con base en una muestra de datos de una zona de estudio.
4. Validar la probabilidad de excedencia de la distancia de viaje de movimientos en masa en una zona de estudio.

1.3 Metodología de trabajo

Para cumplir con los objetivos propuestos y llegar a la evaluación de una metodología para el cálculo de distancias de viaje se llevó a cabo el siguiente procedimiento:

- a) Selección y definición de la zona de estudio y la muestra:

- i. Se toma como primera aproximación el área comprendida por los departamentos de Santander, Norte de Santander y Boyacá.
 - ii. Haciendo uso del registro disponible de movimientos en masa (Servicio Geológico Colombiano, 2024), se recopilan 1503 movimientos catalogados como flujos o deslizamientos.
 - iii. Bajo el uso de muestreo por conglomerados se establecen 12 subzonas de estudio reduciendo los casos a 466.
 - iv. Se recopila información secundaria de diferentes fuentes para establecer los casos a estudiar, reduciendo la muestra finalmente a 50 movimientos catalogados como flujos o deslizamientos.
- b) Producción de la base de datos
- i. Se establece una campaña exploratoria para recopilar información topográfica detallada haciendo uso de técnicas modernas de fotogrametría.
 - ii. Se procesa esta información para generar un modelo de elevación y establecer las diferentes dimensiones que componen cada uno de los movimientos.
 - iii. Se ensambla una base de datos que servirá como insumo para los procesos posteriores.
- c) Comparativa entre los modelos existentes para el cálculo de la distancia de viaje.
- i. A través de una recopilación bibliográfica se establecen diferentes modelos de cálculo de distancia de viaje a utilizar.
 - ii. Se utilizan los valores recopilados del área de estudio y se calculan las distancias de viaje.
 - iii. Se calcula el error para cada modelo comparando el valor real con el teórico.
- d) Análisis estadístico de la base de datos.
- i. Un primer análisis exploratorio haciendo uso de propiedades estadísticas básicas (media, mediana, desviación estándar, etc.) permite establecer el comportamiento general, dando indicativos de rango de variación y si existen correlación entre las variables.
 - ii. Haciendo uso de la distancia de Mahalanobis se establece la uniformidad de los datos y si es necesario realizar una transformación de estos, también permite establecer valores atípicos.

- iii. Se realiza una transformación $\log+1$ para las variables cuantitativas mejorando notablemente la uniformidad de los datos.
 - iv. Haciendo uso del análisis por componentes principales se establece la importancia de cada variable y si alguna de estas puede ser obviada en posteriores procedimientos debido a su baja importancia.
- e) Propuesta de un nuevo modelo de cálculo de distancia de viaje.
- i. Estableciendo la necesidad de aislar completamente la variable objetivo L e implementar en el mismo proceso de cálculo variables de diferentes naturalezas y comportamientos, se opta por utilizar técnicas de Machine Learning.
 - ii. Se establece la técnica de Random Forest para el cálculo de L .
 - iii. Se calibra el modelo buscando el menor error haciendo variar los hiperparámetros.
- f) Propuesta del cálculo probabilístico de la distancia de viaje.
- i. Se establece los coeficientes de variación para obtener valores aleatorios de entrada.
 - ii. Se recopila información de un movimiento en masa adicional que servirá como caso de estudio.
 - iii. Para cada modelo estudiado en el numeral c y e se ingresan los valores reales del caso de estudio y se afectan por su respectivo coeficiente de variación registrando las distancias de viaje calculadas.
 - iv. Se establecen curvas de probabilidad de excedencia para cada modelo permitiendo escoger el de mejor ajuste al valor real de la distancia de viaje medida para el caso de estudio.
- g) Finalmente se redacta una metodología de trabajo que permite replicar el procedimiento realizado en este documento a otras bases de datos o a otras zonas.

1.4 Estructura del documento

Este documento de trabajo final de maestría se divide en 9 capítulos contando la fase introductoria, el segundo busca relacionar al lector con los diferentes sistemas de clasificación de movimientos en masa y el modelo más común existente para el cálculo de la distancia de viaje donde relaciona en ángulo de fricción equivalente con el volumen. También recuerda conceptos básicos de propiedades estadísticas para un primer análisis

exploratorio de los datos recopilados y la definición de otro tipo de análisis más robustos como el Análisis por Componentes Principales (ACP) así como conceptos importantes de metodologías de cálculo descritos como *Machine Learning*, más específicamente el método de *Random Forest*.

El tercer capítulo establece el área de estudio, debido a que este trabajo se encuentra enmarcado en el proyecto de investigación “Método para la estimación de distancias de viaje de movimientos en masa con base en aproximaciones geométricas” el área de estudio este compuesto en una primera fase por Boyacá, Santander y Norte de Santander. También se describe la campaña exploratoria para la obtención de la información topográfica primaria, fundamental en la elaboración de la base de datos usada en capítulos posteriores solucionando el primer y segundo objetivo específico.

El cuarto capítulo recopila diferentes metodologías para el cálculo de la distancia de viaje, usando métodos principalmente basados en regresiones lineales con fuentes de información diversa y analizando los errores de cada uno de ellos con respecto al valor real junto con su distribución.

El quinto capítulo expone los resultados del análisis estadístico de los datos recopilados del área de estudio, iniciando con conceptos básicos que permiten realizar un análisis exploratorio entendiendo el comportamiento general para luego adentrarse en técnicas más robustas que permiten encontrar valores atípicos o variables con baja importancia.

El sexto capítulo presenta el montaje, la calibración y posterior validación del modelo de *Machine Learning*, siendo esta una herramienta poderosa a la hora de estimar valores dadas unas condiciones iniciales de diversas características.

El séptimo y octavo capítulo buscan plantear el cálculo probabilístico de la distancia de viaje, relacionando uno o varios parámetros aleatorios de entrada para lograr un rango de distancias de viaje, el séptimo capítulo presenta el método de cálculo, mientras que el octavo resume de manera detallada el proceso por el cual se obtienen los resultados del capítulo 7 dividiendo el proceso en cuatro fases dando como producto la metodología

descrita a lo largo del documento y una curva de probabilidad de excedencia que puede ser replicada con esta metodología.

Finalmente en el noveno capítulo se redactan algunas conclusiones y recomendaciones asociadas a los resultados y el trabajo realizado a lo largo del documento.

2 Marco Teórico

Para conocer y entender el estado actual de los análisis de distancias de viaje para los flujos de detritos y deslizamientos y cómo pueden ser implementados métodos probabilísticos en dichos análisis es necesario entender, en primera medida, qué es, cómo se define y cómo se compone un movimiento en masa. Así mismo, en este capítulo se presentan conceptos de la estadística para el manejo de bases de datos. Al final del capítulo se presenta el método de Random Forest con el fin de dar a conocer los principales aspectos que se tendrán en cuenta en la aplicación de esta técnica para la metodología propuesta para el cálculo de distancia de viaje durante el desarrollo de este trabajo.

2.1 Movimientos en masa y distancias de viaje

2.1.1 Definición y clasificación de los movimientos en masa

La clasificación de los movimientos en masa en Colombia está definida dentro de la Guía Metodológica para Estudios de Amenaza, Vulnerabilidad y Riesgo por Movimientos en Masa producida por el Servicio Geológico Colombiano (Ávila, y otros, 2016), para efectos prácticos únicamente se tomará la definición de flujo de detritos y deslizamientos de cada uno de estos sistemas. Dentro del documento anteriormente nombrado se resaltan cuatro sistemas de clasificación que son:

2.1.1.1 *Hutchinson (1988)*

Para este sistema los flujos de detritos se encuentran comprendido en un grupo más grande nombrado por el autor como movimientos de detritos con forma de flujo, dentro de los cuales se encuentran los flujos de tierras no periglaciales, flujos de tierras periglaciales, flujo en suelos colapsables, flujo de detritos y avalanchas.

La característica principal del movimiento definido como flujo de detritos es que poseen una alta velocidad (3m/s a 12m/s), lo que les atribuye un gran potencial de daño. En cuanto a su localización se presentan principalmente en zonas montañosas donde ocurran lluvias de alta intensidad (40 – 60 mm/h o 150 a 200 mm/día) lo cual permite obtener una relación de causalidad entre el detonante y el movimiento en masa. En cuanto a su composición posee un bajo contenido de finos (<3%) diferenciándolos de los flujos de tierras, donde el contenido de finos puede llegar al 65%, esto condiciona fuertemente la velocidad a la cual se produce el movimiento.

Se diferencia de las avalanchas por su alto peso unitario (18.6 – 19.6 kN/m³), valor cercano al de una mezcla de concreto fresco, mostrando así un alto contenido en la fase sólida lo que aumenta su capacidad de generar daños debido a la velocidad que puede alcanzar. El régimen de flujo depende de la profundidad, siendo laminar a bajas profundidades, los esfuerzos dentro del movimiento se producen en condiciones no drenadas y sus valores son cercanos a las presiones hidrostáticas lo que mantiene al material en condiciones de licuación constante dificultando así su sostenimiento y disminución en su velocidad. La forma del flujo se ve condicionada fuertemente por la morfometría de la zona, manteniendo su velocidad mientras este canalizado y formando un abanico de eyección cuando las condiciones morfométricas cambian ya sea por la inexistencia del canal o un cambio hacia una menor pendiente.

En cuanto a los movimientos definidos como deslizamientos los subdivide en fallas confinadas, deslizamientos rotacionales simples, sucesivos o múltiples, deslizamientos traslacionales y deslizamientos de roca llegando a velocidades relativamente altas donde su forma y velocidad dependerá principalmente de su forma (rotacional o traslacional) y tipo de material movilizado.

2.1.1.2 Varnes (1978)

En esta clasificación el autor enfoca su atención en la forma del movimiento y pasa a un segundo plano su origen, esto genera que los flujos de detritos se encuentren catalogados dentro del grupo “flujo de detritos y tierras”. Este tipo de movimientos se caracterizan principalmente porque, como su nombre lo indica, el movimiento se asemeja al de un fluido incluso sin poseer la fase líquida, además se encuentra comúnmente canalizado, con una

forma alargada y esbelta donde se puede diferenciar claramente la zona de origen del material y el cono de eyección, esto hace que posean una larga distancia de viaje. El autor también diferencia entre flujo de detritos, avalanchas de detritos y flujos de tierras donde, para las avalanchas la fase dominante es la líquida lo que le otorga su alta velocidad. Para el flujo de tierras domina la fase sólida pero el tamaño con mayor porcentaje dentro de la gradación es el correspondiente a arcillas o limos y no necesariamente poseen un alto contenido de agua.

Los deslizamientos los clasifica principalmente en dos grupos, rotacionales y traslacionales, esto dependerá de la forma de la superficie de falla que se genere debido a los esfuerzos de corte a los cuales esté sometido el material, son movimientos dados generalmente de manera progresiva indicando una propagación de la superficie de falla hasta generar la separación del volumen desplazado.

2.1.1.3 Cruden y Vanes (1996)

Dentro de esta clasificación se genera un complemento a lo establecido en Varnes (1978) donde se discretiza de manera más clara los diferentes tipos de movimientos en masa, tomando como características principales el tipo de movimiento (caídas, volteos, deslizamientos, propagaciones y flujos) y el material desplazado (roca, granular o fino).

Como descripción adicional se establece el concepto de frontera entre el material movilizado y el canal que lo contiene introduciendo el esfuerzo de corte entre estos dos elementos como variable a tener en cuenta. También da importancia a la evolución del movimiento, estableciendo que parámetros como el volumen desplazado o el contenido de agua modifican su magnitud a lo largo del recorrido.

Este documento es especialmente importante ya que establece por primera vez la forma del movimiento típico, definiendo las dimensiones geométricas que lo componen como la profundidad, el ancho, el área de la superficie de falla, la altura máxima del movimiento, una ecuación de cálculo para el volumen desplazado y finalmente la distancia de viaje, así como un formato de registro de reporte y una escala de velocidad.

2.1.1.4 Hungr (2014)

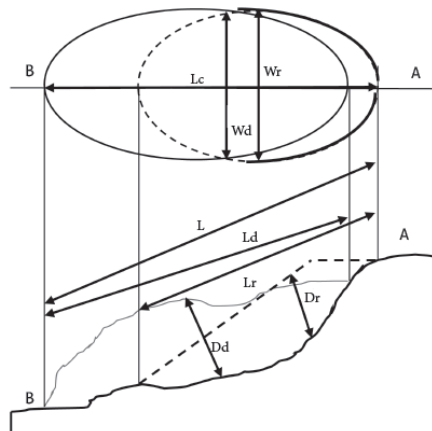
Para esta clasificación las características del movimiento como la velocidad, el material y la condición de saturación están lo suficientemente definidas por trabajos anteriores como las clasificaciones expuestas por Hutchinson, Cruden y Varnes, lo cual permite dar valores o rangos que permitan definir los flujos y deslizamientos los cuales corresponden a 19 de los 32 tipos de movimientos.

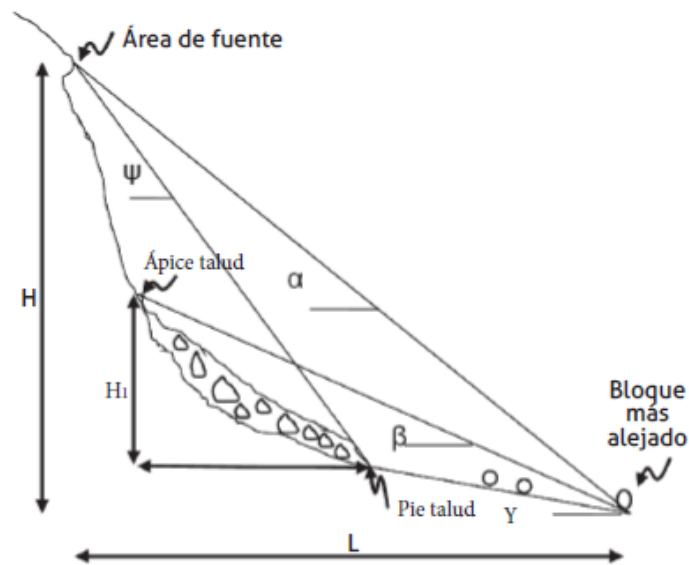
- Material: Si es roca, arcilla, arena o limo la composición principal.
- Velocidad: Dependiendo de diversos factores como la pendiente topográfica, el contenido de agua y el volumen desplazado su velocidad se clasifica entre muy rápido y extremadamente rápido (5m/s a 3m/min)
- Forma del movimiento: Si es rotacional, planar, compuesto, en cuña o irregular.

2.1.2 Modelo planta-perfil típico para un movimiento en masa.

Para definir las variables que se utilizan dentro de la estimación de la distancia viaje se toma como modelo el propuesto por la Guía Metodológica (Ávila, y otros, 2016) en el numeral 3-10 que se muestra en la figura 2-1, donde gráficamente se definen las características geométricas.

Figura 2-1. Modelo planta perfil asumido Tomado de la Guía Metodológica (Ávila, y otros, 2016).





Aunque sencillo, este paso permite estandarizar la información para poder organizarla y facilitar su uso ya que la cantidad de casos estudiados es considerable (se procesaron alrededor de 1503 eventos). Así mismo, con este procedimiento se permite tener una información base estableciendo unos parámetros mínimos que cualquier usuario puede utilizar para el cálculo de distancias de viaje usando diferentes modelos como los planteados en el capítulo 4 y 6.

Es importante resaltar que esta no es la única información tomada para el presente documento, existen valores como D_d , L_d , o β que no fueron medidos debido a la dificultad que presentaba establecer las características de la zona de depósito, así como la adición de otras medidas como δ relacionada con el cambio de dirección o DO relacionada con los obstáculos presentados en el recorrido de volumen desplazado. En el capítulo 3 se presentan las variables tenidas en cuenta en este trabajo.

2.1.3 Metodologías empíricas para obtener una única distancia de viaje

En la Guía Metodológica Para Estudios de Amenaza, Vulnerabilidad y Riesgo Por Movimientos en Masa (Ávila, y otros, 2016), se presenta un único proceso por el cual es posible estimar el ángulo de alcance, definido como la relación entre la altura total del movimiento H (medida desde la corona hasta el bloque más alejado) y L como la longitud

horizontal medida entre dichos puntos, si se conoce esta relación es posible obtener L. Para esto se utiliza la Ec. 2-1.

$$\log\left(\frac{H}{V}\right) = A + B * \log(V) \quad \text{Ec. 2-1}$$

Donde H es la altura total del movimiento, L la longitud total recorrida en planta, A y B son constantes obtenidas de la tabla 2-1.

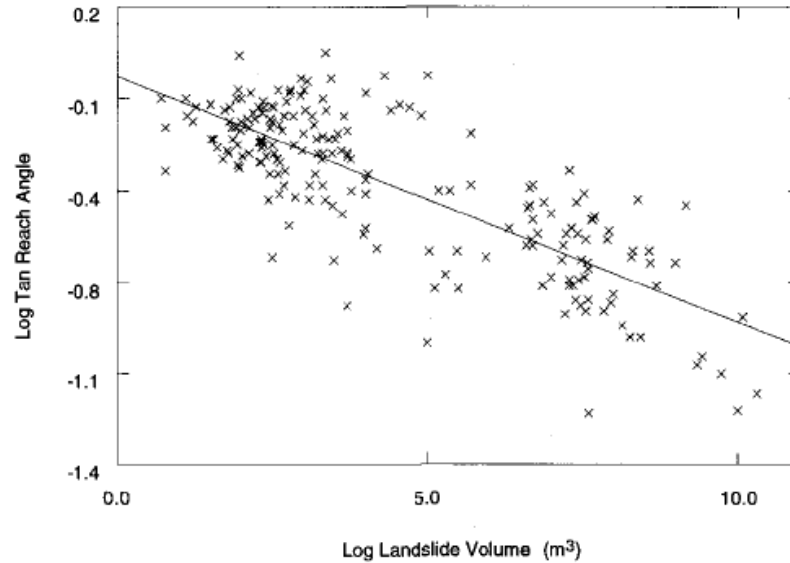
Tabla 2-1. Coeficientes de regresión lineal para cada tipo de movimiento.

Tipo deslizamiento	de Trayectoria	A	B	R ²
Caída de rocas	Todo	0,210	- 0,109	0,76
	Obstruido	0,231	- 0,091	0,83
	Sin obstruir	0,167	- 0,119	0,92
Deslizamiento traslacional	Todo	- 0,159	- 0,068	0,67
	Obstruido	- 0,133	- 0,057	0,76
	Sin obstruir	- 0,143	- 0,080	0,80
Flujo de detritos	Todo	- 0,012	- 0,105	0,76
	Obstruido	- 0,049	- 0,108	0,85
	Sin obstruir	- 0,031	- 0,102	0,87
Flujo de tierra	Todo	- 0,214	- 0,070	0,65
	Sin obstruir	- 0,220	- 0,138	0,91

Fuente: Extraído de Guía Metodológica Para Estudios de Amenaza, Vulnerabilidad y Riesgo Por Movimientos en Masa (Ávila, y otros, 2016).

Estas funciones son producto del estudio realizado por (Corominas, 1996), el cual a través de la recopilación de 204 movimientos se logra, a través de regresiones lineales, obtener la forma y los valores descritos anteriormente mostrados en la figura 2-2.

Figura 2-2. Modelo de regresión lineal entre el volumen y el ángulo de eficiencia del movimiento.



Fuente: (Corominas, 1996).

Como se puede observar en la figura 2-2, de manera general, la relación que sostiene el logaritmo del ángulo de alcance (ordenada) y el logaritmo del volumen movilizado (abscisa) es inversa, lo que sugiere que, para volúmenes pequeños de $0.5 \times 10^3 \text{m}^3$ o menores, las distancias de viaje son similares a volúmenes mayores.

El autor es consciente de las limitaciones de este modelo, como la complejidad de cada caso a estudiar, la forma y longitud del canal sobre el cual se moviliza, el cambio de tipo de movimiento durante la trayectoria y si está o no obstruido el canal. Con el objetivo de reducir la incertidumbre debida a la gran cantidad de variables que participan en el movimiento, subdivide la información estudiada en el tipo de movimiento principal o inicial y si está obstruido o no, lo que permite mejorar el ajuste de las curvas.

Algo importante a destacar es que la información utilizada por Corominas (1996) es de distintas fuentes y estudios, bajo esta premisa es posible mejorar los resultados aplicando esta metodología, pero utilizando únicamente datos medidos para una única zona en específico donde se garantice uniformidad geológica y geomorfológica. Este trabajo ya ha sido adelantado por distintos autores como (Weiwei, y otros, 2017) con 55 de los más

catastróficos movimientos ocurridos en China, (Weiwei, y otros, 2017) con 38 avalanchas de roca canalizadas en la zona de Wenchuan, China y (Muhammad, 2015) con 106 movimientos en masa en Indonesia, realizando división por tipo de movimiento donde 8 de éstos son flujos de detritos, adicionalmente, utilizó distintas formas en la función de ajuste, no solamente relaciones lineales Log – Log, también hace uso de ecuaciones exponenciales.

Moncayo (2021); Moncayo y Ávila (2022)

Dentro de este trabajo final de maestría el autor hace uso de la base de datos proporcionada por el SGC a través del SIMMA, donde se registran 5994 eventos catalogados como flujo, caída y deslizamiento donde, luego de acotar el área de estudio a la zona Andina y pasar por una fase de depuración obtiene 199 deslizamientos con información confiable.

Posteriormente realiza un estudio detallado de las posibles variables dependientes e independientes y sus características estadísticas dando un total de 23 variables las cuales sirven como contexto para tener un punto de partida y saber qué información buscar en cada uno de los 50 casos de estudio facilitando la elaboración de la base de datos presente en el Anexo A. Algunas de estas variables fueron obviadas debido a la dificultad para rastrear información histórica como el contenido de agua, el factor detonante o el uso del suelo, tampoco se tuvieron en cuenta propiedades como la litología o el ambiente geomorfológico ya que toda esta información no era posible obtener debido a que gran parte de los movimientos estudiados no se encuentran registrados en el SIMMA pero fueron evidentes en el momento del levantamiento topográfico pero de manera asincrónica, lo cual imposibilita tomar esta información categórica en el momento en el que se da el desplazamiento del material.

También hace un importante aporte donde categoriza la probabilidad de no excedencia y realiza un mapa preliminar del deslizamiento Ricaurte-Altaquer, trazando áreas de influencia dependiendo del nivel de amenaza. Por lo tanto el trabajo realizado por Moncayo (2021) sirve como base importante para la elaboración del presente trabajo ya que da las

pautas de las variables que se pueden estudiar, la comparativa entre modelos de estimación y un primer acercamiento en el cálculo de la distancia de viaje. Debido a la diferencia en las variables utilizadas por los modelos en Moncayo (2021); Moncayo y Ávila (2022) no se tendrán en cuenta en el análisis debido a la falta de información como la litología o el factor detonante.

2.2 Elementos para el análisis de estadístico de bases de datos

Este subcapítulo presenta conceptos y propiedades estadísticas exploratorias que, aunque básicas, vale la pena recordarlas ya que permiten tener un primer acercamiento al comportamiento de los datos a estudiar.

2.2.1 Media o promedio muestral

Es la sumatoria las magnitudes de todos los individuos dada una variable dividido por el tamaño de la muestra, en este caso se habla de promedio muestral ya que no se está trabajando con la totalidad de los casos ocurridos en la zona de estudio, se puede calcular mediante la Ec. 2-2.

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad \text{Ec. 2-2}$$

2.2.2 Cuartil 1, cuartil 3 y mediana

En un conjunto de datos ordenados ascendentemente, el cuartil 1, cuartil 3 y la mediana corresponden a valores de la variable que dividen cuatro partes iguales dicho conjunto, cubriendo el 25, 50 y 75% de los datos respectivamente

2.2.3 Valor mínimo y valor máximo

En un conjunto de datos ordenado, son los valores extremos que toma la variable en estudio.

2.2.4 Varianza o covarianza muestral

Es definida como una medida de dispersión en torno a la media, es decir, permite establecer que tan dispersos se encuentra los datos de una variable ya sea comparada con si misma (varianza) o con otra variable (covarianza) y se puede calcular mediante la Ec. 2-3

$$Cov(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n} \quad \text{Ec. 2-3}$$

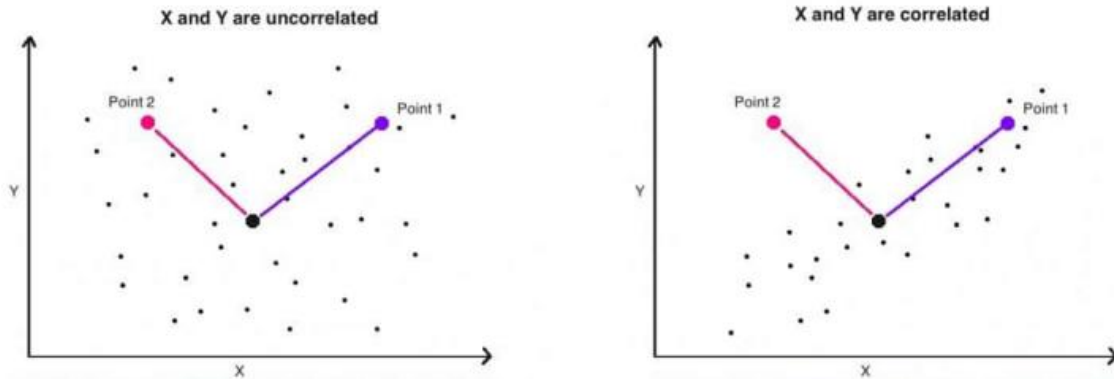
2.2.5 Vector de coeficientes de variación

Definido como la relación entre la raíz de la varianza (desviación estándar) y la media, permite conocer que tan uniforme es una muestra de datos en puntos porcentuales, interpretando un valor menor al 30% como una muestra homogénea, es decir, le media es un valor representativo del conjunto de datos, en caso contrario se cataloga como una muestra heterogénea y se calcula con la Ec. 2-4

$$C.V = \frac{S}{\bar{X}} \quad \text{Ec. 2-4}$$

2.2.6 Distancia de Mahalanobis

La distancia euclídea es utilizada para calcular la magnitud de la línea recta entre dos puntos en un espacio de n variables aleatorias y funciona bien si no existe correlación entre estas variables, buscando similitudes o diferencias entre individuos de una muestra o población, pero presenta deficiencias cuando estas variables cuentan con cierto grado de correlación como lo muestra la figura 2-3.

Figura 2-3. Comparación de correlación entre variables y la distancia euclidiana.

Fuente: (Kaya, 2020).

Es evidente la diferencia de la nube de puntos entre la figura de la derecha e izquierda donde ambos puntos conservan su distancia euclídea pero el punto dos en la figura de la derecha dista notablemente del conjunto de datos debido a la correlación entre las variables.

Para solucionar esto el científico indio Mahalanobis en 1936 propone una alternativa que incluye la matriz de covarianzas, resultando en un vector columna que posee la distancia de Mahalanobis de cada individuo la cual es adimensional positivo, esta distancia está dada por la Ec. 2.5

$$d_M(x, \bar{x}) = \sqrt{(x - \bar{x})S^{-1}(x - \bar{x})^T} \quad \text{Ec 2.5}$$

2.2.7 Análisis por Componentes Principales

Para simplificar la metodología de cálculo se debe realizar un análisis previo que permita concluir si todas las variables incluidas dentro de la base de datos presentan una importancia similar, esto se puede lograr haciendo un análisis de sensibilidad tomando la varianza como indicador.

El Análisis por Componentes Principales (PCA) es un algoritmo que permite reducir la cantidad de variables originales a un grupo de variables virtuales o equivalentes organizadas de mayor a menor varianza, siendo la primera la componente más representativa del conjunto de datos.

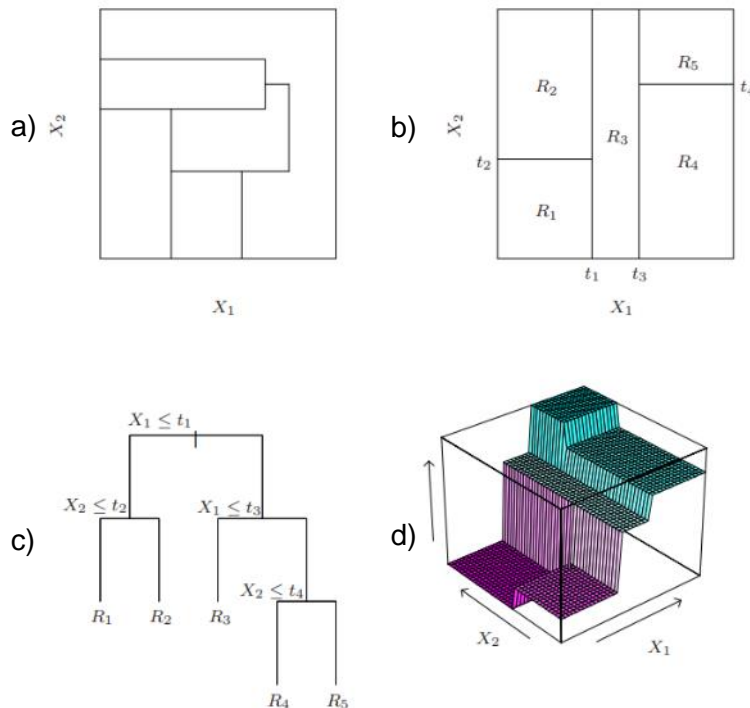
2.3 Método de Random Forest

La mayoría de los modelos de regresión utilizados comúnmente en el cálculo de la distancia de viaje usan el ángulo de eficiencia de movimiento H/L y el volumen de la masa desplazada V, esto imposibilita la opción del cálculo de L sin tener H, limitando la aplicación a movimientos ya ocurridos. Para solucionar esto es necesario aislar completamente la variable de interés L dentro de una técnica que permita ingresar más información.

2.3.1 Árbol de decisión

Los árboles de decisión son estructuras estadísticas basadas en la partición de las variables de entrada dada una condición para realizar dicha división, llegando a diferentes nodos terminales u “hojas”. Existen diferentes algoritmos para elaborar un árbol de decisión, para el presente trabajo se utilizará el algoritmo CART (*classification and regression trees*) debido a la gran cantidad de herramientas que presenta el lenguaje R para su uso, este algoritmo está basado en particiones binarias permitiendo ingresar n variables de entrada de diferente tipo.

Figura 2-4. Esquema del funcionamiento de los árboles de decisión.



Fuente: (Hastie, Tibshirani, & Friedman, 2001)

Para facilitar su interpretación en la figura 2-4 se observan gráficamente la construcción de un árbol de decisión dadas X_1 y X_2 como condiciones iniciales, la figura Figura 2-4(a) corresponde a un caso general que no se obtendría con el algoritmo CART, ya que este algoritmo divide los posibles resultados en el orden en el que X_1 y X_2 fueron seleccionados.

En la Figura 2-4(b) se encuentra zonificada la variable de respuesta en 5 áreas, la primera variable que gobierna las condiciones iniciales es X_1 tomando valores de t_1 y t_3 dando como resultado dos zonas $X_1 < t_1$, $t_1 < X_1 < t_3$ y $X_1 > t_3$, luego X_2 entra a subdividir cada intervalo, para $X_1 < t_1$ $X_2 = t_2$ generando las áreas R_1 y R_2 . Este mismo proceso se realiza para los intervalos restantes de X_1 obteniendo R_3 , R_4 y R_5 .

El valor de la variable objetivo para unas condiciones iniciales de un caso externo a la base de datos con la que se construye el árbol de decisión se obtiene ubicando, de acuerdo con los intervalos realizados, el caso externo en alguna de las cinco áreas generadas y el valor se obtiene del promedio de sus similares dentro de esta área.

Otra interpretación gráfica se puede notar en la Figura 2-4(c) a través de un diagrama donde cada ramal corresponde a un intervalo dando como resultado cinco ramas finales, para casos de tres dimensiones la Figura 2-4(d) muestra la superficie dada el árbol de decisión elaborado, donde cada escalón posee un valor medio de la variable de respuesta, indicando que este algoritmo es escalable a n variables iniciales.

Estos árboles de decisión poseen una alta precisión, es decir, el valor medio de todas las observaciones se ajusta correctamente a los valores de las observaciones con las cuales fue entrenado, pero poseen una alta variabilidad al calcular observaciones de casos fuera del conjunto de datos de entrenamiento debido a que esta exactitud depende del orden en el cual se toman las condiciones iniciales para elaborar los intervalos.

2.3.2 Random Forest

El Random Forest es un algoritmo de aprendizaje automático que pertenece a la categoría de ensamblado de modelos. Combina múltiples árboles de decisión individuales en un "bosque", donde cada árbol se entrena con una muestra aleatoria y características seleccionadas al azar. Luego, las predicciones de cada árbol se promedian o ponderan para generar una predicción final más precisa y estable. Esta técnica es efectiva para la clasificación y la regresión, reduciendo el riesgo de sobreajuste y mejorando la capacidad del modelo para manejar una variedad de datos y patrones (Hastie, Tibshirani, & Friedman, 2001).

Dependiendo del tipo de variable de respuesta, el Random Forest es un método de clasificación (cualitativa) o regresión (cuantitativa) que hace uso de los árboles de decisión generados aleatoriamente y promedia su resultado. Esto permite obtener la distancia de viaje dadas unas condiciones iniciales permitiendo:

- Encontrar aquellas variables que influyen de manera importante en la distancia de viaje y comparar este resultado con el ACP.
- Tomar el rol de ecuación de trabajo, dando una alternativa a las regresiones lineales de una sola variable para aplicar metodologías probabilísticas como la de Montecarlo.
- Utilizar variables cuantitativas y cualitativas dentro del mismo modelo.

Para la generación aleatoria de los árboles de decisión es necesario establecer una serie de parámetros que rigen todo el bosque, estos reciben el nombre de hiperparámetros, los cuales también funcionan como puntos de calibración de modelo. Los más relevantes son:

- Número de árboles generados (N).
- Número de variables que se tienen en cuenta en cada nodo (mtry).
- Número de nodos máximo en cada árbol.
- Porcentaje de la muestra que se toma para entrenar el modelo.

2.3.3 Error en el Random Forest

Dentro del algoritmo utilizado para esta metodología es necesario separar la base de datos en dos, una será de entrenamiento o in-bag y otra será para calcular el error o out-of-bag, el error es la diferencia entre el valor real y el valor generado por el modelo con los casos que se encuentra en la parte del out-of-bag llamado out-of-bag-error o OOBError, siendo calculado como la sumatoria de los cuadrados de la diferencia entre el valor real y el valor calculado.

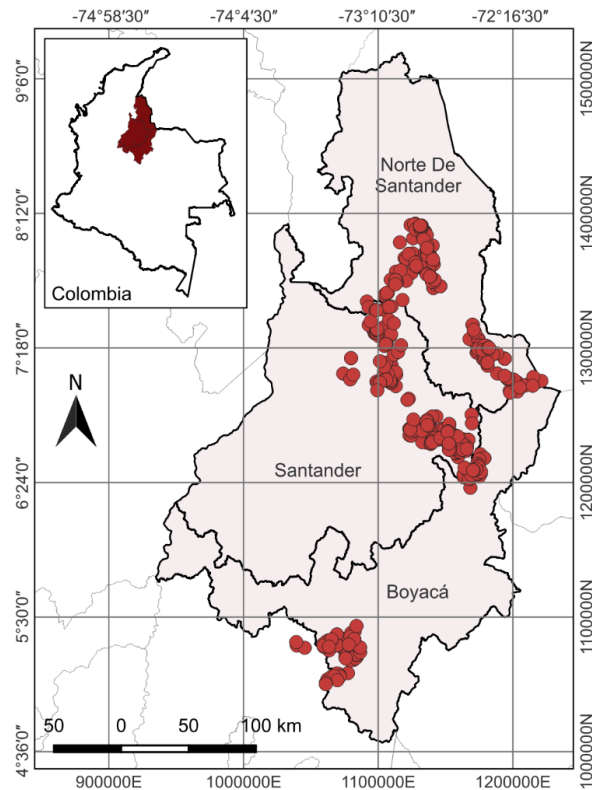
3 Área de estudio

Dentro de este capítulo se describe el proceso de delimitación de la zona de estudio, la información obtenida, las fuentes de información, la categorización y subdivisión de del área y una descripción detallada de la medición de las variables morfométricas que se utilizan como información de entrada para los capítulos 4, 5, 6 y 7.

3.1 Delimitación de la zona de estudio

De acuerdo con lo mencionado en la introducción, este trabajo final de maestría se encuentra enmarcado dentro del proyecto de investigación “Método para la estimación de distancias de viaje de movimientos en masa con base en aproximaciones geométricas” donde, por acuerdo entre las partes que componen el contrato, se define como zona de estudio los departamentos de Boyacá, Santander y Norte de Santander (figura 3-1).

Figura 3-1. Ubicación de la zona estudio. Los puntos resaltados indican los 466 eventos de movimientos en masa considerados luego del muestreo por conglomerados.



Fuente: Propia.

3.2 El Servicio de Información de Movimientos en Masa (SIMMA) como punto de partida

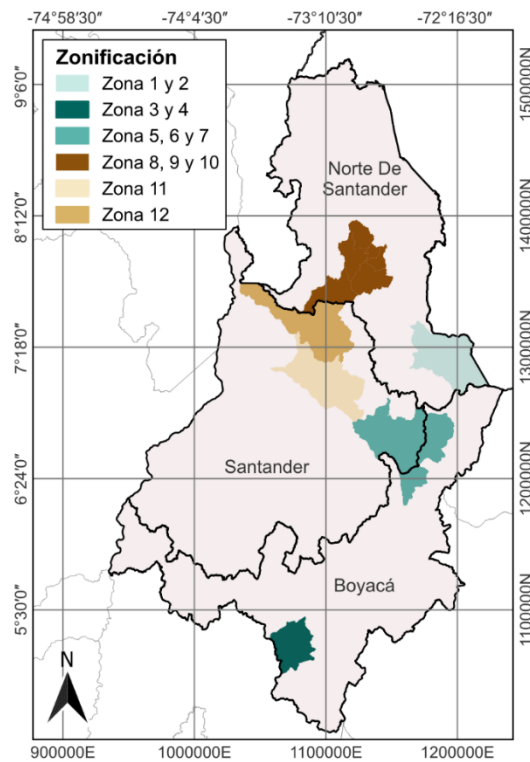
El Servicio de Información de Movimientos en Masa (SIMMA) es una base de datos de libre consulta creada y continuamente mejorada por el Servicio Geológico Colombiano, este sistema ofrece dos tipos de información, el primero obtenido a través de fuentes locales como periódicos relacionados con el evento de interés y otras entidades encargadas de la gestión del riesgo (fuentes secundarias), el segundo tipo son tomas exhaustivas de cada proceso de interés, registrando información como la ubicación, el tipo de proceso, dimensiones aproximadas, posibles causas, usos del suelo y características geológicas dando un mejor contexto en el que se desarrollan los casos de estudio.

Para el presente documento se limita el estudio a los tipos de movimientos definidos como deslizamientos y flujos dentro de la Guía Metodológica para Estudios de Amenaza, Vulnerabilidad y Riesgo por Movimientos en Masa debido a la relativa facilidad al obtener información de cada caso debido a las marcas que dejan estos procesos sobre el paisaje, contrario a las caídas de roca donde el paso del tiempo y la acción antrópica cubre o remueve estas marcas complicando realizar un proceso de ingeniería inversa.

3.3 Subdivisión de la zona de estudio

La zona definida para el desarrollo del proyecto cuenta con un área aproximada de 75.374 km² con 1503 movimientos registrados, esto convierte a la población geográficamente dispersa. Para reducir el tamaño de la muestra y, consecuentemente, los costos de toma de información sin que deje de ser representativa, se hace uso de la metodología de conglomerados, subdividiendo el área en pequeñas zonas con una alta densidad poblacional (figura 3-2).

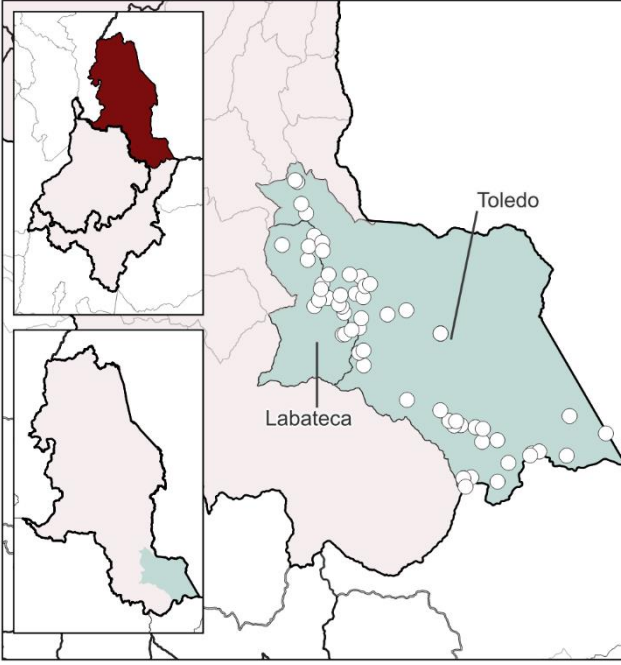
Figura 3-2. Zonificación de la zona estudio.



Fuente: Propia.

El muestreo por conglomerados plantea un equilibrio entre el área que cubre cada conglomerado y la cantidad de casos ya que, a mayor área, mayor será la confiabilidad de los datos, pero aumentarán los costos del proyecto dando como resultado 12 subzonas de estudio con un tamaño óptimo, tal como se muestra en la tabla 3-1.

Tabla 3-1. Subzonas de la zona de estudio por conglomerados.

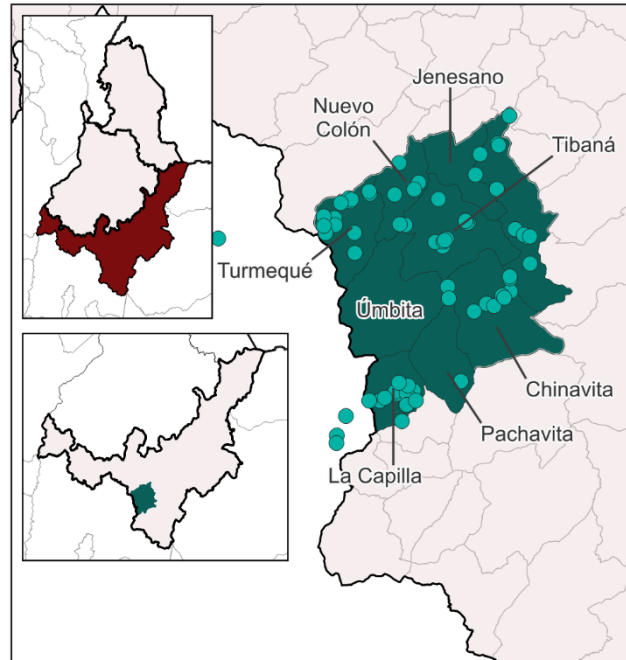
<p style="text-align: center;">Zona 1 y 2</p> <p>Ubicación: Toledo y Labateca, Norte de Santander.</p> <p>Cantidad de Movimientos: 62.</p> <p>Descripción: Comprende alturas entre los 400 y 3600 msnm, la temperatura fluctúa entre los 15 y 25°C y presenta una pluviosidad hasta los 3000 mm/año. El uso del suelo está destinado a la agricultura y ganadería intensiva y cubre mayoritariamente una zona carbonífera.</p>	 <p>The figure consists of two maps. The top-left map is a small-scale inset showing the location of the study area (highlighted in dark red) within the northern part of the Andes. The main map on the right shows a larger area with a light green shaded region representing the study zone. This region is bounded by Toledo to the north and Labateca to the south. Numerous white circles are scattered across the green shaded area, representing the locations of 62 mass movement events. Labels 'Toledo' and 'Labateca' are placed on the map with lines pointing to their respective locations.</p>
--	--

Zona 3 y 4

Ubicación: Sur de Boyacá – Noreste de Cundinamarca (Tibaná, Turmequé, Ramiriquí, Chinavita, Pachavita, Nuevo Colón, La Capilla-Lenguazaque, Tibirita).

Cantidad de Movimientos: 73.

Descripción: Comprende alturas entre los 1000 y 3600 msnm, la temperatura fluctúa entre los 15 y 22°C y presenta una pluviosidad entre los 600 a 1.500 mm/año. El uso del suelo está destinado al desarrollo de pastos y zonas de cultivo. Se presentan manifestaciones de carbón, fuentes de sal y rocas fosfóricas.

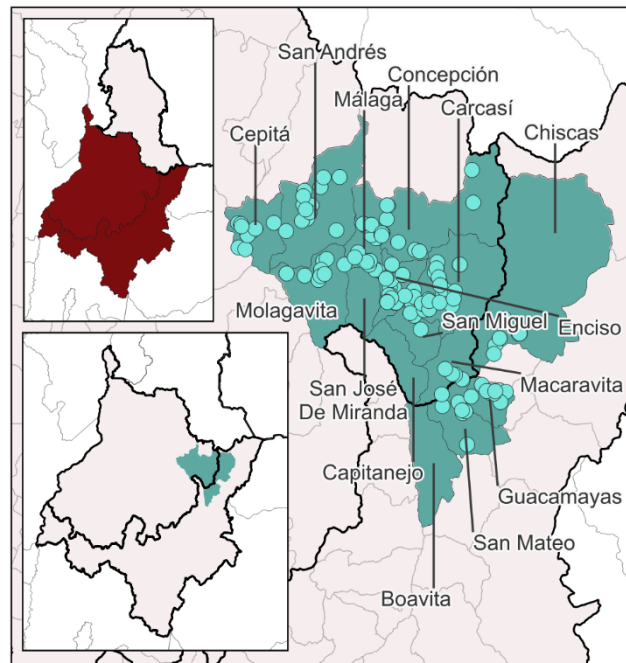


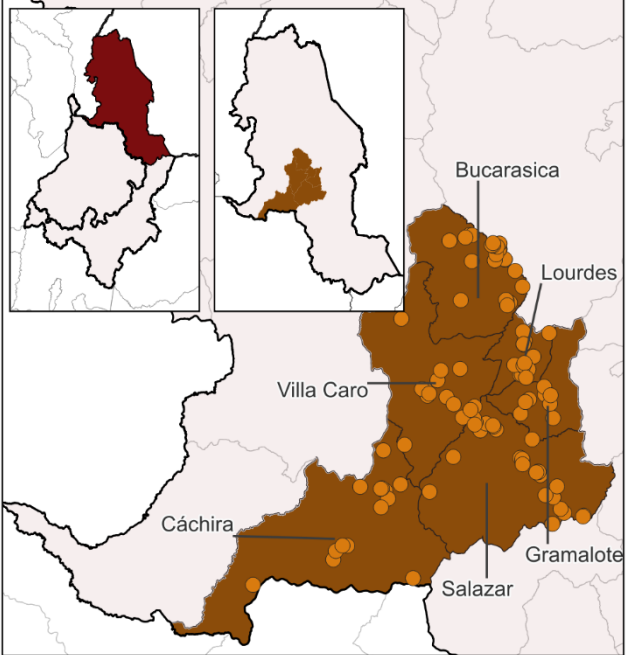
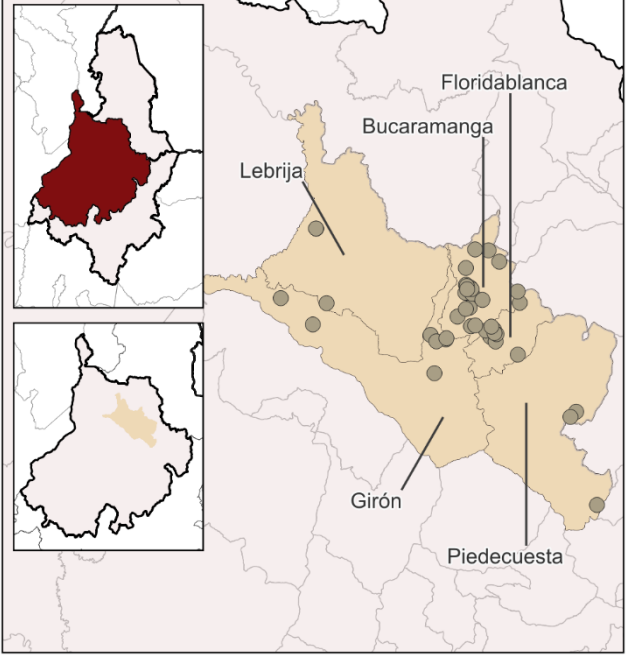
Zona 5, 6 y 7

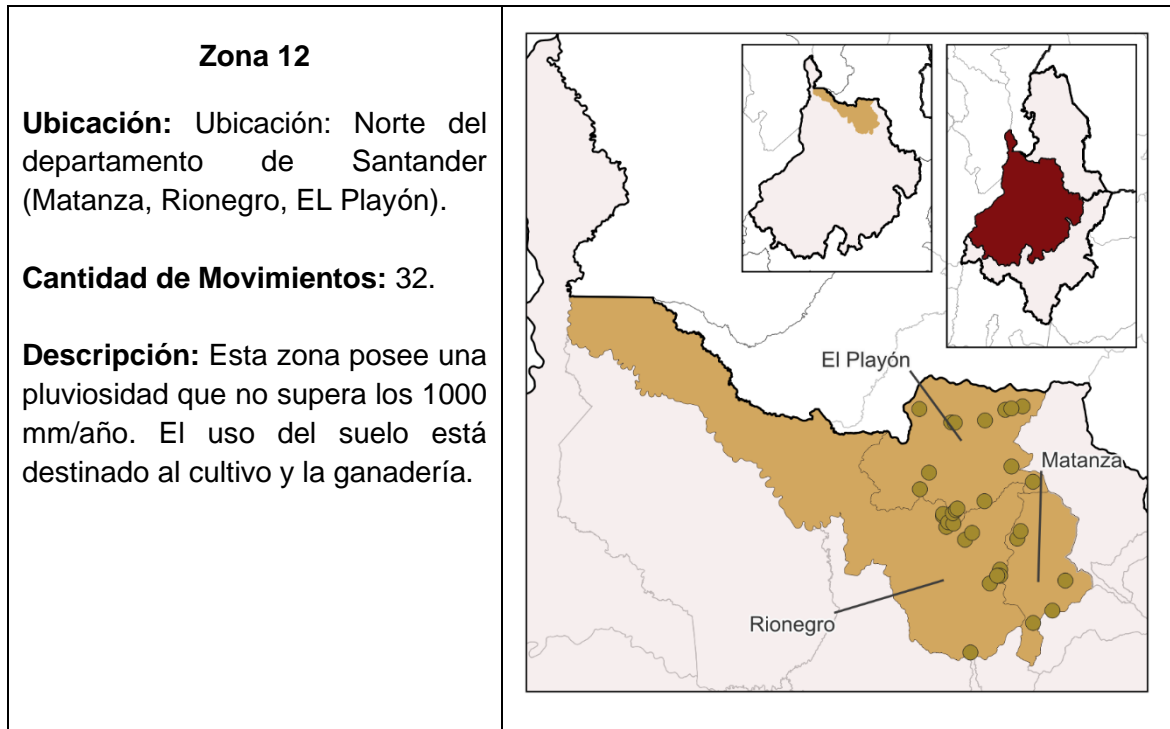
Ubicación: Sur de Santander - Norte de Boyacá (Málaga, Enciso, San Miguel, Concepción, Molagavita, Carcasí, Macaravita, Cepitá, San Andrés-San Mateo, Guacamayas, Chiscas, El Espino).

Cantidad de Movimientos: 140.

Descripción: Comprende alturas entre los 1800 y 3200 msnm. Presenta un potencial minero para la extracción de minerales metálicos, no metálicos y carbón, así como suelos aptos para cultivar.



<p style="text-align: center;">Zona 8, 9 y 10</p> <p>Ubicación: Parte central de Norte de Santander (Gramalote, Villa Caro, Salazar, Bucarasica, Cáchira, Lourdes).</p> <p>Cantidad de Movimientos: 117.</p> <p>Descripción: Posee una altura media de 1700 msnm, la temperatura fluctúa entre los 18 y 23°C. Presenta potencial minero y suelos aptos para el cultivo.</p>	
<p style="text-align: center;">Zona 11</p> <p>Ubicación: Parte norte del departamento de Santander (Bucaramanga, Floridablanca, Lebrija, Girón y Piedecuesta).</p> <p>Cantidad de Movimientos: 42.</p> <p>Descripción: Esta zona posee una pluviosidad que no supera los 1000 mm/año. El uso del suelo está destinado al cultivo y la ganadería.</p>	



Fuente: Propia.

3.4 Metodología para la elaboración de la base de datos

3.4.1 Definición de la muestra

Mediante el uso del SIMMA se registraron 1503 eventos catalogados como movimientos tipo flujo y deslizamiento dentro de los departamentos de Boyacá, Santander y Norte de Santander con una distribución de 524, 336 y 643 casos respectivamente.

Con el objeto de tener una muestra de eventos de movimientos en masa que tengan una información óptima y sean de fácil acceso para su respectivo estudio, se procede a seguir una serie de criterios para la selección de los sitios de estudio.

Haciendo uso de las 12 zonas de estudio con los tamaños óptimos de muestra, se logran reducir a 466 unidades, lo cual sigue siendo un alto número de casos a los cuales se debe extraer información primaria.

Posteriormente se buscan mantener todos aquellos casos que sean de fácil acceso en otras fuentes secundarias y se encuentren en las coordenadas indicadas dentro del reporte detallado, ya que existen movimientos donde no es evidente observar a través de fotos satelitales temporales las marcas que dejan en el paisaje estos procesos de remoción.

Otros aspectos importantes para tener en cuenta son existencia de acceso al caso en cuestión a través de, como mínimo, una ruta apta para vehículo y la cercanía a una cabecera municipal, así como la cercanía a otros casos de estudio. De acuerdo con los parámetros anteriormente mencionados y evaluando cada uno de ellos se logra reducir a 86 casos el tamaño de la muestra para la toma de información secundaria que de los cuales se escogerán los puntos de toma de información primaria. Para cada uno de ellos se genera una ficha técnica propia con la siguiente información:

1. Información general

- 1.1. Fecha de inspección: Fecha en la cual se revisó el estado actual de movimiento en la foto aérea más reciente.
- 1.2. Cod SIMMMA: Código preestablecido por el SIMMA dentro de su base de datos, en caso de no existir se da un valor de 0.
- 1.3. ID interno: Código asignado al caso dentro del proyecto bajo el formato ZXX-YY donde XX es una de las doce zonas dentro de la que se encuentra el movimiento y YY es el número secuencial dentro de la subzona, este último se reinicia para cada subzona.
- 1.4. Latitud: En sistema de coordenadas Magna Sirgas WGS 84 decimales
- 1.5. Longitud: En sistema de coordenadas Magna Sirgas WGS 84 decimales.
- 1.6. Altura: Metros sobre el nivel del mar a los cuales se encuentra aproximadamente la cresta del movimiento.
- 1.7. Zona de estudio: Zona a la cual corresponde entre las 12 posibles.
- 1.8. Municipio
- 1.9. Departamento

2. Información geológica

- 2.1. Fuente de información: En la mayoría de los casos fueron planchas a escalas 1:100000 otorgadas por el SGC.

- 2.2. Otras fuentes de información: En la mayoría de los casos era la información registrada en los Reportes Detallados dados por el SIMMA.
- 2.3. Escala de la información.
- 2.4. Unidad geológica.
- 2.5. Estructuras Cercanas
- 2.6. Estado del movimiento.
- 2.7. Tipo de movimiento.
- 2.8. Fecha del evento
- 2.9. Composición: Descripción de la roca o suelo presentado por la fuente principal de información.

3. Información Geomorfológica:

- 3.1. Fuente de información: En la mayoría de los casos fueron planchas a escalas 1:100000 otorgadas por el SGC.
- 3.2. Otras fuentes de información: En la mayoría de los casos era la información registrada en los Reportes Detallados dados por el SIMMA.
- 3.3. Escala de la información.
- 3.4. Detalle de la geoforma: Si es Ambiente, Unidad, Subunidad, Estructura o Subestructura.
- 3.5. Nomenclatura: Nomenclatura de la geoforma.
- 3.6. Ambiente
- 3.7. Nombre de la geoforma.
- 3.8. Descripción.

4. Información del uso del suelo:

- 4.1. Fuente de Información: Para la mayoría de los casos la única fuente existente es la registrada en los Reportes Detallados dados por el SIMMA
- 4.2. Otras fuentes de información.
- 4.3. Técnica: Técnica con la cual se obtuvo el uso del suelo.
- 4.4. Escala.
- 4.5. Tipo y porcentaje de cada uso del suelo.
- 4.6. Tipo y porcentaje de cada cobertura.

5. Información hidrológica:

- 5.1. Fuente de información: Generalmente es el IDEAM
- 5.2. Registro de las estaciones cercanas a 5km con el código asignado por el IDEAM, las coordenadas, la elevación sobre el nivel del mar y la categoría.

6. Información imágenes:

- 6.1. Fuente de información: Generalmente es Google Earth
- 6.2. Otras fuentes: La fuente de la cual extrae la información Google Earth
- 6.3. Escala.
- 6.4. Resolución: A cuantos metros corresponde cada píxel o spot
- 6.5. Información multitemporal: Si existen imágenes de la misma zona para diferentes años.
- 6.6. Tipo de Imagen.
- 6.7. Cantidad de imágenes.
- 6.8. Año de la primera imagen.
- 6.9. Año de la última imagen.

7. Información cartográfica:

- 7.1. Fuente de información: Generalmente es el IGAC
- 7.2. Otras fuentes.
- 7.3. Escala.
- 7.4. ID Plancha: ID establecido por la fuente de información
- 7.5. Información multitemporal: Si existe cartografía de la misma plancha para diferentes años.
- 7.6. Tipo de Imagen.
- 7.7. Cantidad de mapas.
- 7.8. Año del primer mapa.
- 7.9. Año del último mapa.

8. Información Sísmica

- 8.1. Fuente de información: Generalmente es el SGC
- 8.2. Otras fuentes.
- 8.3. Cantidad de eventos registrados.
- 8.4. Intensidad máxima observada

8.5. Intensidad sísmica esperada.

8.6. Zona de amenaza sísmica.

8.7. Registro de los eventos sísmicos: Donde se consigna la fecha del evento, las coordenadas, el código dado por la fuente de información y la magnitud.

La información registrada para cada uno de los 86 casos fue evaluada por expertos en distintas áreas del conocimiento buscando diversidad en las características generales y específicas a través del método del muestreo por conveniencia, donde se categorizan las diferentes variables a considerar y cada experto le asigna un valor ponderado donde la suma siempre debe dar 100, obteniendo las categorías principales a tener en cuenta a la hora de tomar decisiones en la definición de la muestra.

Esto reduce el área de estudio aún más, dando como resultado la elección de la zona 5, 7, 11 y 12 con 50 casos que se tendrán en cuenta en la campaña exploratoria, en el Anexo D se presentan las fichas con la información de los 50 eventos estudiados.

3.4.2 Definición de la campaña exploratoria

Dentro del planteamiento del proyecto de investigación no solamente se busca obtener información acerca de las variables geométricas como, sino también obtener la incidencia de variables morfológicas y geotécnicas que puedan condicionar la distancia de viaje.

Con este fin, la campaña de exploración plantea el levantamiento de información topográfica haciendo uso de Vehículos Aéreos no Tripulados (VANT) que permiten cubrir grandes áreas con una alta precisión a través de la reconstrucción del terreno haciendo uso de fotos aéreas y técnicas de fotogrametría, consolidando una nube de puntos como se verá más adelante.

Este levantamiento topográfico no solo aumenta la escala y la precisión de la toma de las dimensiones de las partes principales de los movimientos en masa, también permite realizar mapas geomorfológicos a escalas de detalle, así como la definición del uso del suelo y su recubrimiento.

Otra parte importante que compone la campaña es la extracción de información geotécnica, a través del uso de muestreo in-situ como el método del cono y arena para obtener densidades, y el penetrómetro de bolsillo como un ensayo índice para evaluar la dureza de la roca estudiada.

También se obtuvieron muestras alteradas logrando humedades y límites de consistencia. Se realizaron extracciones de muestras inalteradas para realizar ensayos de corte directo post falla obteniendo valores de resistencia residual. También se realizó un levantamiento geológico y geomorfológico de las zonas de estudio pero esta información no se incluye dentro del proceso y el análisis de este documento ya que no pertenece al alcance del mismo, en caso de requerir una consulta a esta información se puede encontrar en los productos del proyecto de investigación "Método para la estimación de distancias de viaje de movimientos en masa con base en aproximaciones geométricas" dentro del cual se encuentra enmarcado este trabajo final de maestría.

El recorrido realizado se subdividió en 7 estaciones así:

- Estación 1: Floridablanca, Bucaramanga y Girón
- Estación 2: Zona rural de Girón.
- Estación 3: Tona y Floridablanca en el nororiente de Bucaramanga.
- Estación 4: El Playón.
- Estación 5. Guaca.
- Estación 6: San Andrés
- Estación 7: Carcasí

Este plan de trabajo se estructura teniendo en cuenta dos factores fundamentales parcialmente ajenos al presente trabajo de grado, pero fundamentales en cualquier proyecto de ingeniería: El alcance financiero y el tiempo de ejecución, por lo tanto, se busca optimizar al máximo ambos recursos resultando en estaciones de muestreo cercanas entre sí, de fácil acceso pero que cubran la mayoría de las zonas con características similares.

Los volúmenes encontrados dentro de los casos estudiados rondan entre los 100 y los $1 \times 10^6 \text{m}^3$ de masa desplazada, y la dimensión máxima no supera los 300 m en la mayoría de los casos excepto uno en que se detallará más adelante. Esto permite establecer una

escala de trabajo general definida como escala de detalle, donde se usan levantamientos entre 1:1000 y 1:5000 (Ávila, y otros, 2016).

Por lo tanto, la base de datos de partida contiene información geométrica, morfológica, y geotécnica que van más de la mano con estas escalas de detalle. A continuación, se presenta una metodología de extracción de cada una de las características tomando como ejemplo el movimiento etiquetado como Z6_20_5 de la estación 7.1.

Esta campaña exploratoria fue realizada en colaboración con el equipo del proyecto de investigación de la Universidad Industrial de Santander, y jóvenes investigadores del proyecto, bajo la coordinación de la Universidad Nacional de Colombia. Este trabajo final de maestría toma los resultados de la campaña exploratoria para realizar el estudio objeto de este trabajo.

3.4.3 Medición de variables morfométricas.

Como se mencionó anteriormente, la información topográfica se obtiene a través de VANTs, los cuales, a través de una planificación de vuelo y una altura predeterminada para la zona, toman fotos secuenciales que permiten extraer la superficie del terreno con separaciones entre los puntos de incluso centímetros

Lo primero que se busca son las marcas generadas por la superficie de falla, principalmente la cresta y los flancos dejados al momento que se desprende el material haciendo uso de los ortomosaicos producidos anteriormente (Figura 3-3). Las figuras 3-2 a 3-8 presentan los análisis realizados a partir de la información topográfica utilizando como ejemplo el evento Z6_20_5 de la Zona 6 (Carcasí, Santander)

Figura 3-3. Identificación de las partes principales de la superficie de falla. Evento Z6_20_5
(Carcasí, Santander)



Fuente: Propia con base en ortomosaicos producidos dentro del proyecto “Método para la estimación de distancias de viaje de movimientos en masa con base en aproximaciones geométricas”.

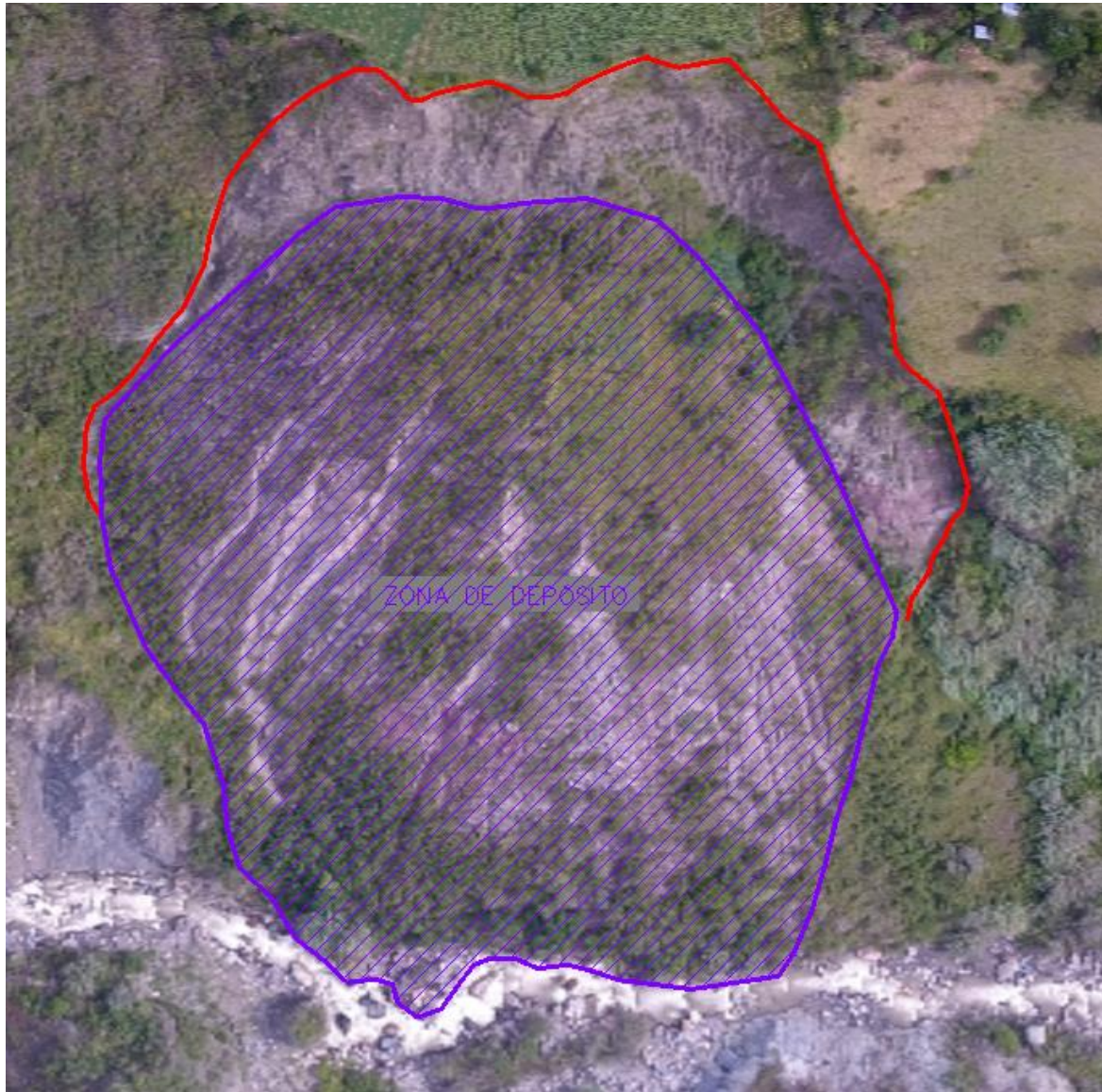
Un buen complemento en el ejercicio de encontrar la superficie de falla es la generación de las curvas de nivel de la zona, haciendo uso de las herramientas proporcionadas por AutoCAD Civil 3D y a través de la nube de puntos.

Figura 3-4. Generación de las curvas de nivel. Evento Z6_20_5 (Carcasí, Santander)

Fuente: Propia con base en ortomosaicos producidos dentro del proyecto “Método para la estimación de distancias de viaje de movimientos en masa con base en aproximaciones geométricas”.

Una vez identificada la superficie de falla, se limita la zona de flujo y depósito (Figura 3-5). Estas áreas no siempre son visibles con facilidad ya que, entre la ocurrencia del desprendimiento del material y la toma de información, pudieron haber ocurrido eventos como crecimiento de la vegetación, transporte del material de depósito por acción fluvial, eólica o antrópica, también se puede presentar mantenimiento y refuerzo del talud debido a la cercanía con una obra civil lo que dificulta en gran medida la toma de la distancia de viaje real de los diferentes movimientos en masa.

Figura 3-5. Identificación de las partes principales del movimiento, zona de superficie de falla (rojo), zona de depósito (morado). Evento Z6_20_5 (Carcasí, Santander)



Fuente: Propia con base en ortomosaicos producidos dentro del proyecto “Método para la estimación de distancias de viaje de movimientos en masa con base en aproximaciones geométricas”.

Una vez identificadas las áreas principales de afectación del movimiento (zona de superficie de falla, zona de transporte o flujo y zona de depósito) se procede a desglosar e identificar las diferentes variables que componen el movimiento. Estas variables se han dividido en tres grupos de acuerdo con la zona de localización dentro del movimiento.

3.4.3.1 Variables del primer grupo

VARIABLES relacionadas con la superficie de falla:

- Longitud de la superficie de falla (L_r): Medida desde la cresta del movimiento hasta la “pata” de la superficie de falla, paralela a la inclinación original del terreno.
- Ancho de la superficie de falla (W_r): Medida entre los escarpes laterales de la superficie de falla y en dirección perpendicular a L_r
- Profundidad de la superficie de falla (D_r): Medida perpendicular a la cota original del terreno, desde la superficie original hasta la superficie de falla.
- Área paramétrica (A_p): Correspondiente al área medida en planta de la superficie de falla.
- Volumen de la masa desplazada (V): Es la variable más común utilizada para describir la magnitud de los movimientos en masa, calculada mediante la ecuación propuesta por la Guía Metodológica (Ávila, y otros, 2016) y definida por la ecuación Ec. 3-1.

$$V = \frac{1}{6} \pi (D_r * W_r * L_r) * f_{exp} \quad \text{Ec 3-1}$$

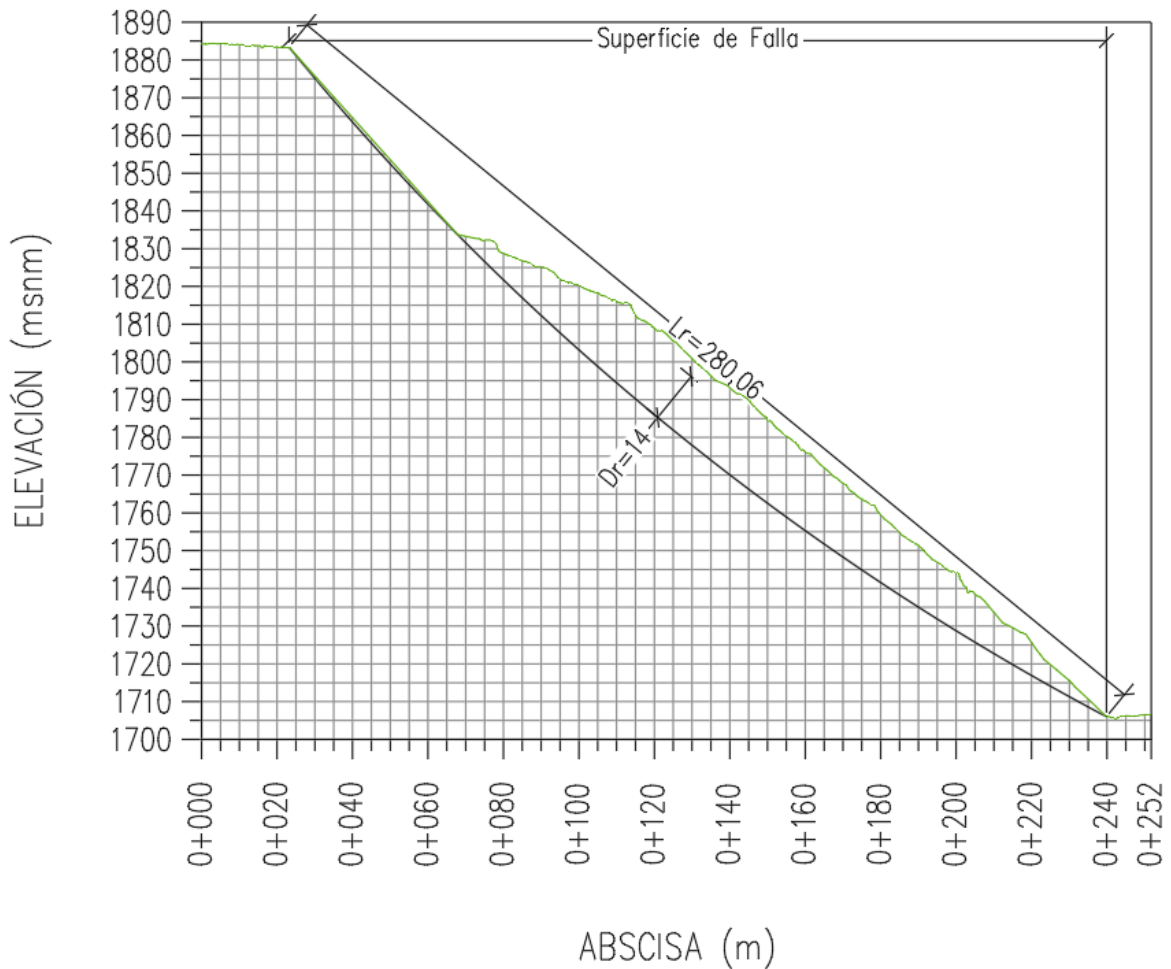
Donde f_{exp} es un factor de expansión del material, para este trabajo se considera un valor de 1,48 siendo un valor medio propuesto en la guía.

Figura 3-6. Toma de las medidas del área de la superficie de falla (A_p) y su ancho (W_r) en m^2 y m respectivamente. Evento Z6_20_5 (Carcasí, Santander)



Fuente: Propia con base en ortomosaicos producidos dentro del proyecto “Método para la estimación de distancias de viaje de movimientos en masa con base en aproximaciones geométricas”.

Figura 3-7. Toma de la longitud de la superficie de falla (L_r) y su profundidad (D_r). Evento Z6_20_5 (Carcasí, Santander)



Fuente: Propia con base en ortomosaicos producidos dentro del proyecto “Método para la estimación de distancias de viaje de movimientos en masa con base en aproximaciones geométricas”.

3.4.3.2 Variables del segundo grupo

El segundo grupo de información geométrica registrado es el relacionado con la zona de transporte, condicionado por la morfología de la zona y donde se encuentran dos principales dimensiones.

- Angulo del talud prefalla (α'): Como su nombre lo indica, es el ángulo medido entre la horizontal y la pendiente promedio del terreno previo al movimiento.

- Dirección del movimiento (δ): Medido en planta, es el ángulo relativo formado entre la dirección de salida del movimiento y la dirección de llegada del depósito, puede tomar valores negativos si el cambio se da en sentido antihorario.
- Elevación sobre el nivel del mar (msnm): Altitud de la corona de la superficie de falla.

3.4.3.3 Variables del tercer grupo

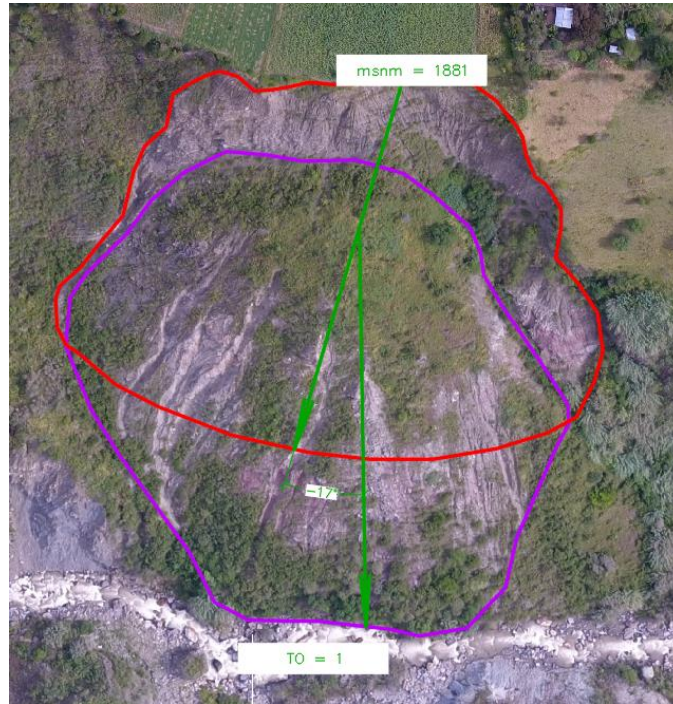
El tercer grupo está compuesto por variables que condicionan principalmente la zona de depósito, entre los que se encuentran:

- Tipo de obstáculo (TO): Variable categórica que permite establecer si el movimiento está obstaculizado y qué tipo de obstáculo presenta, se propone la siguiente organización.
 - No obstaculizado: 0
 - Cuerpo de agua: 1
 - Obra vial: 2

Se observan únicamente tres categorías para el tipo de obstáculo dentro de los casos registrados, pero puede ser ampliable a otro tipo de obstáculos como obras de contención o zonas urbanas.

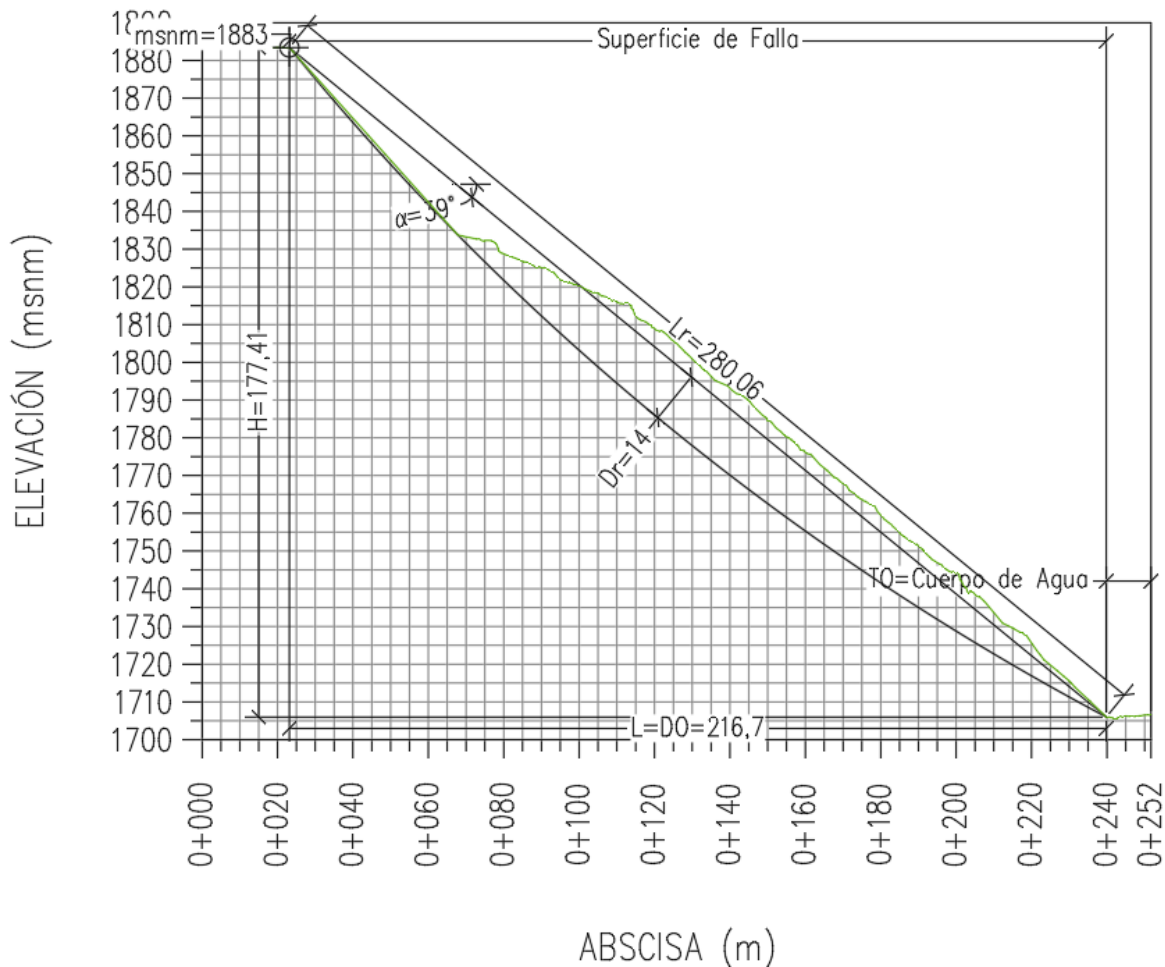
- Distancia al obstáculo (DO): Distancia medida en planta entre la corona de la superficie de falla y el obstáculo más cercano, esto permite conocer si el material movilizado superó el obstáculo o este es una limitante para la distancia de viaje.
- Distancia de viaje (L): Longitud medida en planta entre la corona del movimiento y la “pata” de la zona de depósito.
- Altura máxima (H): Relacionada directamente con la distancia de viaje, es la longitud vertical medida en corte entre la corona del movimiento y la “pata” de la zona de depósito.
- Coeficiente de fricción equivalente (H/L): Relación entre la altura máxima y la distancia de viaje.

Figura 3-8. Toma de la altitud de la corona de la superficie de falla, ángulo de cambio de dirección y tipo de obstáculo. Evento Z6_20_5 (Carcasí, Santander)



Fuente: Propia con base en ortomosaicos producidos dentro del proyecto “Método para la estimación de distancias de viaje de movimientos en masa con base en aproximaciones geométricas”.

Figura 3-9. Toma de la magnitud del ángulo de talud prefalla, altura máxima del movimiento y distancia de viaje. Evento Z6_20_5 (Carcasí, Santander)



Fuente: Propia con base en ortomosaicos producidos dentro del proyecto "Método para la estimación de distancias de viaje de movimientos en masa con base en aproximaciones geométricas".

Complementando la información geométrica, el tipo de movimiento proporciona información importante sobre la dinámica de este. Haciendo uso de las categorías presentadas por la Guía Metodológica, se dividen los casos en deslizamientos y flujos dando importancia a la velocidad que puede alcanzar el movimiento condicionando la distancia de viaje. Se asignaron las siguientes categorías:

- Deslizamiento rotacional (1)
- Deslizamiento traslacional (2)
- Flujo (3)

4 Estimación de la distancia de viaje de movimientos en masa con base en aproximaciones geométricas

Como punto de control dentro de la recopilación bibliográfica se tienen distintas curvas que permiten relacionar geoméricamente el volumen V , y en algunos casos el ángulo de talud prefalla α con la distancia de viaje L de manera directa o indirecta. En este capítulo se analizarán modelos que se aplicarán a los 50 movimientos de estudio: Corte talud ((Finlay & Mostyn, 1999), (Scheidegger, 1973), (Nicolett & Sorriso, 1991), (Corominas, 1996), (Ávila, y otros, 2016) y (Roman, Ortiz, Tapias, & Oviedo, 2024)

4.1 Corte talud

Este modelo relaciona el volumen de la masa desplazada V y el ángulo de talud prefalla α con la distancia de viaje L , calculando de manera directa el valor de interés.

Este modelo este compuesto por tres curvas, límite superior, límite inferior y promedio dadas por la Ec. 4.1

$$a. \text{Log}(L) = 0,062 + 0,965 * \text{Log}(H) - 0,558 * \text{Log}(\tan \alpha)$$

$$b. \text{Log}(L) = 0,109 + 1,01 * \text{Log}(H) - 0,506 * \text{Log}(\tan \alpha)$$

$$c. \text{Log}(L) = 0,156 + 1.055 * \text{Log}(H) - 0,454 * \text{Log}(\tan \alpha)$$

Ec. 4.1. Modelo Corte talud (Finlay & Mostyn, 1999), Tomado de (Ávila, y otros, 2016), (a) corresponde al límite superior, (b) al promedio y (c) al límite inferior.

Se realizo un análisis de las distancias de viaje L , calculadas para los 50 movimientos en masa de estudio. Al comparar L calculada con L medida en campo el error relativo

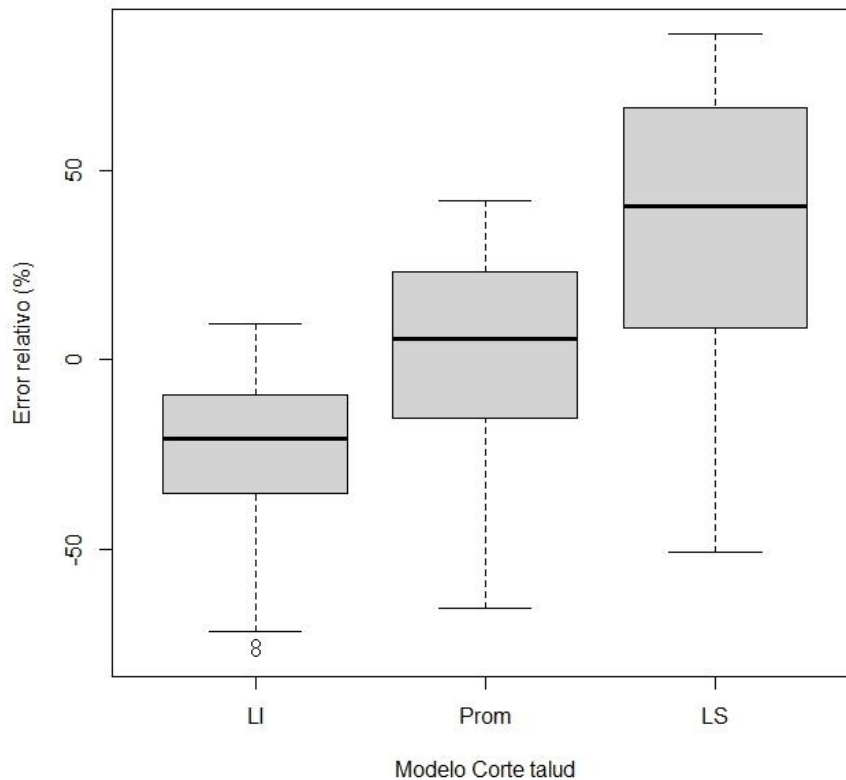
promedio y el rango intercuartílico (definidos en la sección 2.2.2) se pueden ver en la tabla 4-1.

Tabla 4-1. Media y Rango Intercuartílico (RI) aplicando el modelo Corte Talud a los 50 movimientos en masa del estudio.

	LI	Prom	LS
Med	-27,5%	-2,5%	31,3%
RI	25%	33%	50%

La distribución del error relativo se puede apreciar mejor en la figura 4-1.

Figura 4-1. Distribución del error relativo para los 50 movimientos en masa utilizando el Modelo Corte Talud.



Fuente: Propia.

Debido a la existencia de valores negativos en el cálculo del error relativo (valores negativos corresponden a una subestimación en la distancia de viaje) el promedio por sí solo puede dar una falsa idea de calidad, un claro ejemplo es la media del modelo promedio con un valor de aproximadamente -2.5% pero el 50% de los datos se encuentran distribuidos en un rango de 33% de error, y la mayoría se encuentran en un intervalo entre el -50% y 50% teniendo una variabilidad de 100% aproximadamente sin tener en cuenta valores atípicos.

Es evidente la variación de las distribuciones del error ya que cada ecuación corresponde a un límite distinto, pero también es notable el aumento de la variabilidad y como pasa el modelo de una predicción subestimada a una sobreestimada.

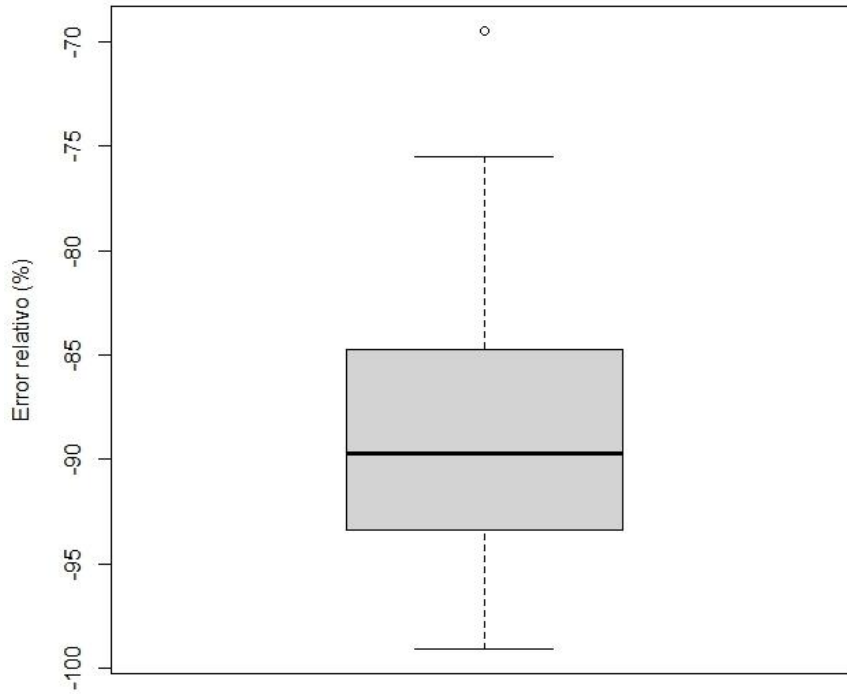
4.2 Modelo de Scheidegger (1973)

Este modelo al igual que las siguientes dos aproximaciones relacionan el logaritmo del volumen (v) con el logaritmo del ángulo de eficiencia del movimiento (H/L), por lo tanto, posee un fuerte condicionante al momento de calcular L ya que depende de H . La ecuación que plantea Scheidegger está dada por la Ec. 4-2:

$$\text{Log} \left(\frac{H}{L} \right) = 0,624 + 0,15666 * \text{Log}(V) \quad \text{Ec. 4-2}$$

Con V en 10^3 m^3 , en la figura 4-2 se observa la distribución del error relativo al comparar los resultados con los valores reales tomados en campo (para los 50 eventos), obteniendo valores con un promedio de -88.9% y un rango intercuartílico de 25%, la totalidad de los errores son negativos subestimando totalmente los valores reales.

Figura 4-2. Distribución del error relativo utilizando el modelo Scheidegger para los 50 movimientos en masa.



Fuente: Propia.

El hecho de que el rango intercuartílico sea menor que el modelo anterior no es un indicativo en la mejoría del cálculo de L, ya que, al no poder tomar valores negativos debido a el uso de una escala logarítmica, todos los resultados tienden a concentrarse en un mismo intervalo.

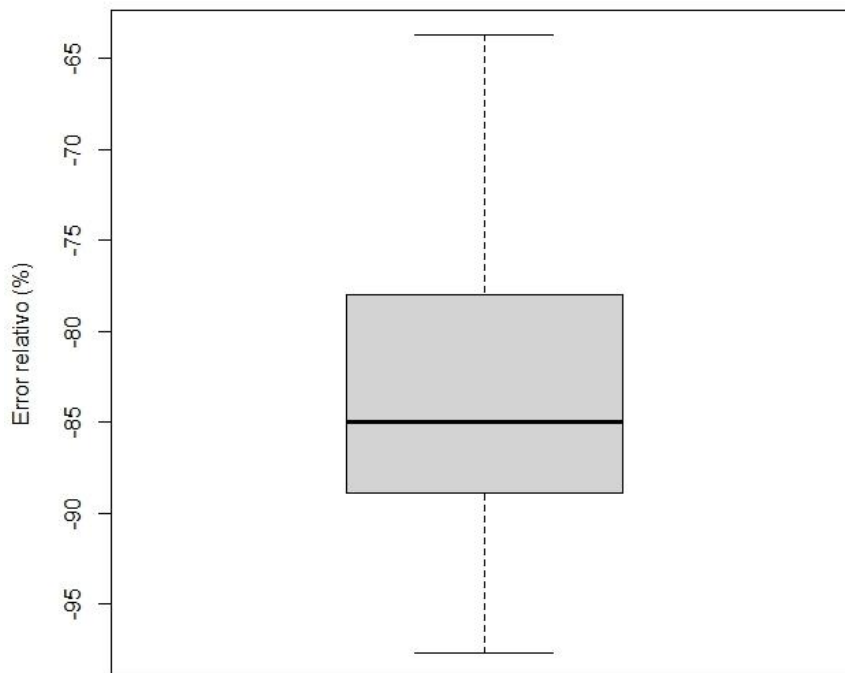
4.3 Modelo de Nicoletti & Sorriso(1991)

De manera similar al modelo de (Scheidegger, 1973), este modelo no calcula de manera directa L, dificultando su uso en laderas que no han presentado movimiento pero que pueden presentarlo, la regresión lineal propuesta esta expresada en la Ec. 4.3.

$$\text{Log}\left(\frac{H}{L}\right) = 0,527 - 0,0847 * \text{Log}(V) \quad \text{Ec. 4.3}$$

Al igual que el modelo propuesto por (Scheidegger, 1973), se presenta una relación directamente proporcional entre el logaritmo de H/L y el logaritmo de V y la totalidad de los errores siguen presentando valores negativos, obteniendo valores no conservativos al hacer uso de estos resultados (figura 4-3).

Figura 4-3. Distribución del error relativo utilizando el modelo (Nicolett & Sorriso, 1991) para los 50 movimientos en masa.



Fuente: Propia.

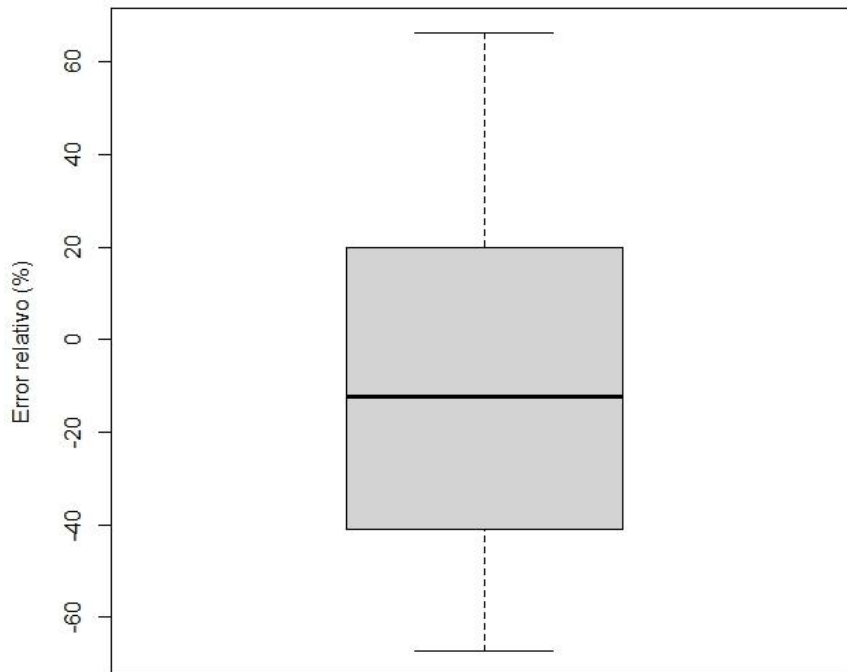
4.4 Modelo de Corominas.

Al igual que los modelos anteriores, se relacionan las mismas variables involucradas como lo muestra la Ec 4-4 y conservan la relación inversamente proporcional entre las variables

involucradas pero la distribución del error relativo exhibe una mejoría, ya que parte de los resultados pasan al lado conservativo, aunque estos no superan el 50% de los resultados manteniendo la media del error en la parte negativa (figura 4-4).

$$\text{Log}\left(\frac{H}{L}\right) = -0.085 * \text{Log}(V) - 0.047 \quad \text{Ec. 4-4.}$$

Figura 4-4. Distribución del error relativo utilizando el modelo de Corominas (1996) para los 50 movimientos en masa.



Fuente: Propia.

Al comparar estos resultados con los obtenidos en la sección 4.1, el Límite Superior presenta valores más conservadores ya que la totalidad del rango intercuartílico del error relativo se encuentra en la parte positiva, es decir, sobreestima la distancia de viaje. Si se compara las ecuaciones restantes del modelo de Corte Talud (Finlay & Mostyn, 1999) presenta comportamientos similares a el promedio y más conservadores que el límite inferior.

4.5 Ecuaciones propuestas por la Guía Metodológica

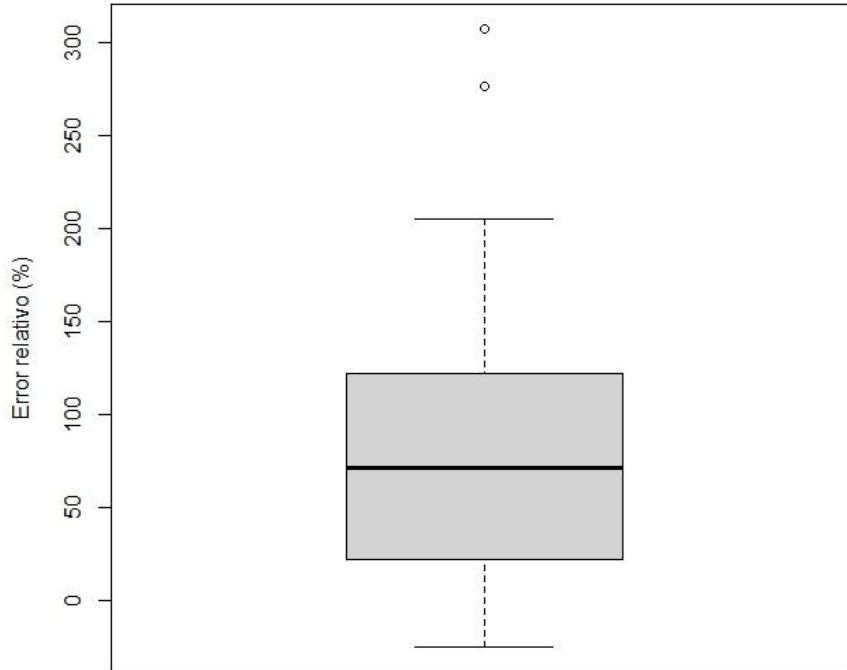
El avance propuesto en la Guía Metodológica (Ávila, y otros, 2016) está compuesto por la categorización del modelo lineal utilizado en las propuestas anteriores usando dos variables cuantitativas: el tipo de movimiento y la obstrucción de la trayectoria, como lo muestra la tabla 4-2, describiendo 4 tipos de movimiento, A representa el intercepto, B la pendiente de la recta y R2 es el coeficiente de correlación para dicha ecuación.

Tabla 4-2. Modelo de regresión lineal categorizado por tipo de movimiento para el cálculo de la distancia de viaje.

TIPO DE DESLIZAMIENTO	TRAYECTORIA	A	B	R2
Caída de rocas	Todo	0.210	-0.109	0.76
	Obstruido	0.231	-0.091	0.83
	Sin obstruir	0.167	-0.119	0.92
Deslizamiento traslacional	Todo	-0.159	-0.068	0.67
	Obstruido	-0.133	-0.057	0.76
	Sin obstruir	-0.143	-0.080	0.80
Flujo de detritos	Todo	-0.012	-0.105	0.76
	Obstruido	-0.049	-0.108	0.85
	Sin obstruir	-0.031	-0.102	0.87
Flujo de tierra	Todo	-0.214	-0.070	0.65
	Sin obstruir	-0.220	-0.138	0.91

Fuente: Tomado de (Ávila, y otros, 2016)

Al comparar los resultados calculados con los medidos en campo para los 50 movimientos de estudio, se obtiene la figura 4-5, mostrando una distribución conservadora, es decir, el 100% del rango intercuartílico se encuentra en el lado positivo, sobreestimando la distancia de viaje real, pero con alta variabilidad comparada con los modelos anteriores (Rango Intercuartílico. RI = 99.13%).

Figura 4-5. Distribución del error relativo con el modelo de regresión lineal categorizado.

Fuente: Propia.

4.6 Román et al. (2024)

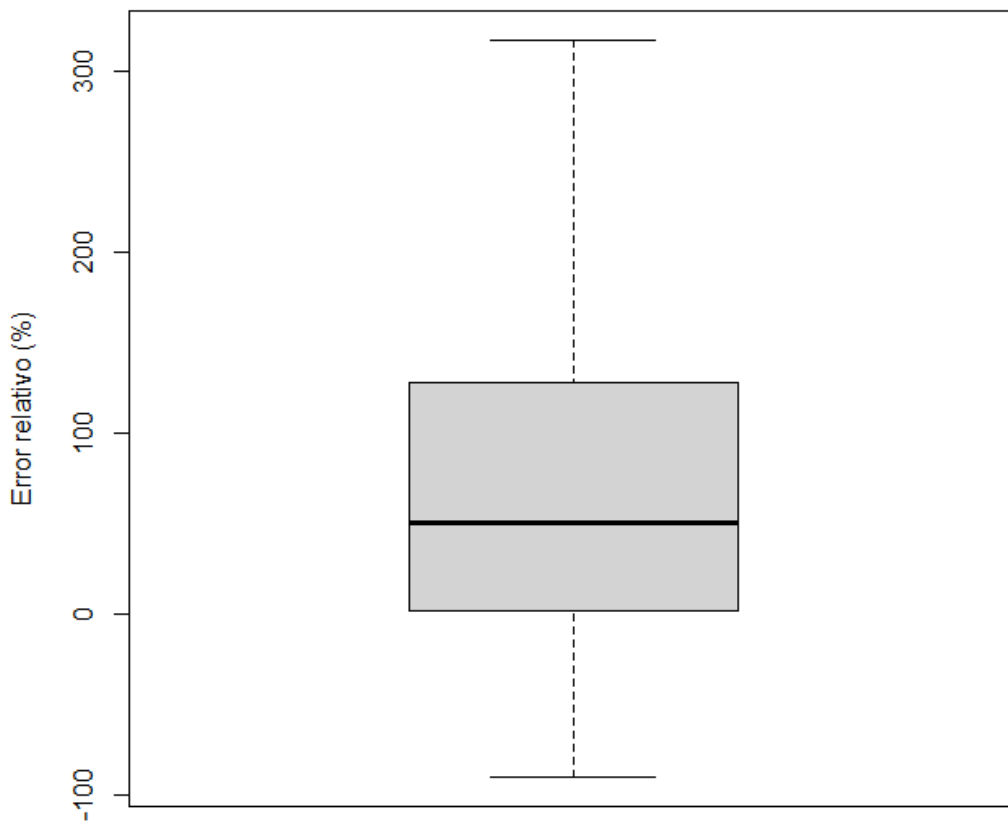
Este modelo para el cálculo de la distancia de viaje se encuentra explicado de manera detallada en (Roman, Ortiz, Tapias, & Oviedo, 2024), Este modelo está definido por la Ec 4-5 relacionando directamente la distancia de viaje L con el volumen de la masa desplazada V .

$$L = 0.0005V + 172.16 \quad \text{Ec. 4-5}$$

Esta expresión se presenta una regresión lineal haciendo uso del volumen como parámetro de entrada, la curva de ajuste es una recta de pendiente $m=0.0005$ y un intercepto $b=172.16$. La distribución del error relativo para los 50 eventos se muestra en la figura 4-6, en esta expresión V está dado en m^3 y L en metros. El *boxplot* posee un rango

intercuartílico entre 22.12% y 121.25%, esto se debe al coeficiente constante de la ecuación, donde para $V=0$ la distancia de viaje es 172.16 superior a la mediana de L presentada en la fase de exploración estadística de la base de datos en la sección 5.1.1, lo que indica que más del 50% de las predicciones están sobreestimando los valores reales, Es este el primer modelo donde el 75% de las estimaciones son conservadoras, es decir, el 75% de las estimaciones se encuentran sobre el rango positivo de los errores relativos.

Figura 4-6. Distribución del error regresión lineal con el volumen utilizando el modelo de Román y otros (2024) para los 50 movimientos en masa.



Fuente: Propia.

5 Análisis estadístico de la base de datos

En este capítulo se busca contextualizar y analizar la relación entre la distancia de viaje y las variables independientes obtenidas en el capítulo 3, tales como la profundidad de la superficie de falla (D_r), el volumen de la masa desplazada (V), etc. También verificar una posible correlación entre ellas y su importancia haciendo uso de la técnica de análisis por componentes principales (ACP) con base en la información de los 50 movimientos en masa.

La base de datos realizada a partir de información primaria tiene un mayor control sobre el inventario de los valores registrados para cada movimiento, por lo tanto, no existen individuos con características faltantes lo cual permite omitir un primer preproceso fundamental en todo análisis, que es la gestión de los datos faltantes, continuando así con el análisis exploratorio o descriptivo de los datos.

5.1 Análisis descriptivo de la base de datos

Todos los procesos aquí mostrados se realizaron en su mayoría haciendo uso del software libre R, más específicamente en RStudio, lo cual facilita al usuario la reproducibilidad de los cálculos realizados. En el Anexo B se presenta el código programado para la realización de estos cálculos.

5.1.1 Valores estadísticos fundamentales

Como primer acercamiento se busca calcular los valores de media o promedio muestral, mediana, primer y tercer cuartil, valor máximo y mínimo presentados para cada una de las variables. Estos indicadores permiten al analista dar una valoración previa de la distribución de las variables o si se encuentra la muestra sesgada positiva no negativamente y si existen posibles valores atípicos.

Haciendo uso de la rutina *summary* es posible obtener cada uno de estos valores para cada una de las doce variables cuantitativas enlistadas a continuación:

- Longitud de la superficie de falla, Lr
- Ancho de la superficie de falla, Wr
- Profundidad de la superficie de falla, Dr
- Área paramétrica, Ap
- Volumen de la masa desplazada, V
- Ángulo de talud prefalla, α
- Cambio de dirección, δ
- Altura sobre el nivel del mar, msnm
- Distancia al obstáculo, DO
- Distancia de viaje, L
- Altura máxima del movimiento, H
- Ángulo de eficiencia del movimiento, H/L

En el Anexo A se presenta la base de datos de información primaria utilizada durante el desarrollo de este trabajo final de maestría donde se registran 50 movimientos y 17 variables. Cabe aclarar que de las 17 columnas se eliminaron las correspondientes a variables cuantitativas como TM, TO o ID, también se omitieron los valores de coordenadas ya que su única función es ubicar espacialmente los movimientos dando como resultado la tabla 5-1.

Tabla 5-1. Propiedades estadísticas para cada variable haciendo uso de los 50 movimientos muestrados.

	Lr(m)	Wr(m)	Dr(m)	Ap(m ²)	V(m ³)	$\alpha(^{\circ})$
Min.	13,38	10,36	0,91	93,11	150	16
1 Qu	44,35	24,5	2743	668,81	1792	32,25
Mediana	76,19	43,15	4480	2495,22	8553	37
Media	124,3	75,08	7087	12064,74	412699	36,86
3 Qu	155,91	104,99	9080	6210,52	73494	43
Max	814,94	434,34	58150	280409,99	15950177	55

	$\delta(^{\circ})$	msnm(m)	DO(m)	L(m)	H(m)	H/L
Min.	0	783	0	21,97	25	0,1798
1 Qu	0	935,2	52,02	87,22	48,85	0,4692
Mediana	17	1777,5	114,53	154,36	106,12	0,6441
Media	22,46	1600	255,91	419,01	147,03	0,6383
3 Qu	33	1927,5	218,66	245,1	167,13	0,8171
Max	133	2757	1913,33	7670	1379	1,1502

Lo primero que se nota de estos valores es la diferencia en las escalas en cada una de las variables donde el volumen puede llegar a tomar valores de decenas de millones de metros cúbicos a diferencia de H/L que no supera las 1.5 unidades debido a la diferencia en las unidades de medida y la naturaleza de cada una, esto presenta un primer reto a la hora de aplicar los diferentes modelos de predicción existentes.

Otra característica por resaltar es que en la mayoría de los casos la media es mayor o mucho mayor que la mediana, esto indica una fuerte tendencia al sesgo positivo en la frecuencia de los datos, es decir, el grueso de la población oscila alrededor del valor de la mediana sin tomar valores negativos.

La distribución del área paramétrica A_p y del volumen de la masa desplazada V poseen un rango de variación que cubren 4 y 5 órdenes de magnitud respectivamente, dificultando ser comparadas con la distancia de viaje L la cual varían en apenas dos órdenes de magnitud.

5.1.2 Matrices de correlación

Otra forma de explorar la base de datos es ver si existe alguna relación entre las variables, esto se puede lograr a través de dos indicadores principalmente, la matriz de covarianzas o la matriz de correlaciones.

Varianza o covarianza muestral

La varianza es un caso particular de la covarianza donde $X_i = Y_i$ y conserva las dimensiones de las variables (la definición de la varianza y covarianza está dada por la Ecuación 2-3). Una matriz de covarianzas correlaciona la totalidad de las parejas de variables posibles con el fin de visualizar si existe una correlación directa (covarianza

positiva) o inversa (covarianza negativa) aplicando la Ecuación 2-3 a cada par de variables posible. Esta matriz de covarianzas está representada en la tabla 5-2.

Tabla 5-2. Matriz de covarianzas de las variables haciendo uso de los 50 movimientos muestrados.

	Lr	Wr	Dr	Ap	V	α	δ	msnm	DO	L	H	H/L
Lr	1,76E+04	9,21E+03	1,04E+03	4,60E+06	2,37E+08	-2,59E+02	5,66E+02	3,21E+04	2,43E+04	1,20E+05	2,36E+04	-9,50E+00
Wr	9,21E+03	6,62E+03	5,47E+02	2,56E+06	1,26E+08	-1,29E+02	5,05E+00	2,15E+04	7,88E+03	5,79E+04	1,14E+04	-2,47E+00
Dr	1,04E+03	5,47E+02	7,60E+01	3,16E+05	1,72E+07	-2,16E+01	4,23E+01	1,28E+03	1,67E+03	8,84E+03	1,63E+03	-9,25E-01
Ap	4,60E+06	2,56E+06	3,16E+05	1,60E+09	8,85E+10	-6,03E+04	1,17E+05	8,09E+06	4,04E+06	4,16E+07	7,38E+06	-2,59E+03
V	2,37E+08	1,26E+08	1,72E+07	8,85E+10	5,06E+12	-3,12E+06	7,74E+06	3,91E+08	1,92E+08	2,37E+09	4,08E+08	-1,53E+05
α	-2,59E+02	-1,29E+02	-2,16E+01	-6,03E+04	-3,12E+06	6,82E+01	-2,16E+01	8,66E+02	-4,65E+02	-1,90E+03	-2,98E+02	1,37E+00
δ	5,66E+02	5,05E+00	4,23E+01	1,17E+05	7,74E+06	-2,16E+01	6,80E+02	-6,01E+03	4,84E+03	7,99E+03	1,39E+03	-3,56E+00
msnm	3,21E+04	2,15E+04	1,28E+03	8,09E+06	3,91E+08	8,66E+02	-6,01E+03	3,16E+05	-4,76E+04	1,14E+05	3,06E+04	7,12E+01
DO	2,43E+04	7,88E+03	1,67E+03	4,04E+06	1,92E+08	-4,65E+02	4,84E+03	-4,76E+04	1,62E+05	2,28E+05	4,79E+04	-5,63E+01
L	1,20E+05	5,79E+04	8,84E+03	4,16E+07	2,37E+09	-1,90E+03	7,99E+03	1,14E+05	2,28E+05	1,25E+06	2,21E+05	-1,23E+02
H	2,36E+04	1,14E+04	1,63E+03	7,38E+06	4,08E+08	-2,98E+02	1,39E+03	3,06E+04	4,79E+04	2,21E+05	4,11E+04	-2,05E+01
HL	-9,50E+00	-2,47E+00	-9,25E-01	-2,59E+03	-1,53E+05	1,37E+00	-3,56E+00	7,12E+01	-5,63E+01	-1,23E+02	-2,05E+01	6,25E-02

Por ejemplo se puede apreciar que el ángulo de fricción equivalente H/L posee una relación inversa con el volumen debido al signo negativo que toma, similar a lo encontrado en la literatura, pero es difícil interpretar esta información ya que la diferencia entre dimensiones y unidades de medida no permite una interpretación más profunda de los datos, pero funciona como paso intermedio para otras técnicas exploratorias como la matriz de coeficientes de correlación.

Vector de coeficientes de variación

El vector de coeficientes de variación (definido mediante la Ecuación 2-5) está constituido por el arreglo de los coeficientes de variación (C.V) tal como se muestra en la tabla 5-3 donde S es la desviación estándar muestra, \bar{X} es el valor promedio y C.V es el coeficiente de variación.

Tabla 5-3. Coeficientes de variación calculados para cada una de las variables geométricas haciendo uso de los 50 movimientos muestrados.

	Lr	Wr	Dr	Ap	V	α
S	132,68	81,36	8,71	40004,97	2249692,00	8,26
\bar{X}	124,30	75,08	7087,00	12064,74	412699,00	36,86
C,V	107%	108%	0%	332%	545%	22%

	δ	msnm	DO	L	H	H/L
S	26,08	561,77	402,89	1116,04	202,78	0,25
\bar{X}	22,46	1600,00	255,91	419,01	147,03	0,64
C,V	116%	35%	157%	266%	138%	39%

Se puede observar una alta variabilidad en los datos, donde la pendiente del talud prefalla (α) es la única variable homogénea ya que tiene un valor de coeficiente de variación (C.V) del 22%, este resultado era algo esperado ya que la zona de estudio es de alta montaña y el valor medio oscila cercano a los valores comunes del ángulo de fricción que poseen los materiales encontrados en la zona.

Las siguientes dos variables con coeficientes de variación (C V) más cercanas al 30% son la altura sobre el nivel del mar (msnm) y el ángulo de fricción equivalente (H/L), siendo este último de mayor interés ya que indica una prevalencia en el recorrido del material movilizado en dirección horizontal, impulsado evidentemente por la pendiente promedio.

También se debe resaltar la alta variabilidad que presenta el volumen de la masa desplazada (V) y el área paramétrica (Ap) que, al ser magnitudes medidas en tres y dos dimensiones respectivamente, se ven afectadas por la variabilidad de las componentes básicas Dr, Lr y Wr, esto conlleva a tener la necesidad de transformar la distribución de los datos para mejorar los resultados.

Matriz de coeficientes de correlación.

Debido a la dificultad que presenta interpretar los resultados obtenidos en la matriz de covarianzas, es necesario buscar otros indicadores para visualizar cómo se relacionan entre sí las variables. Uno de los más utilizados es el coeficiente de correlación R de Pearson (dado por la Ec. 5-1) que, al calcularlo para una base de datos da como resultado una matriz cuadrada de dimensión igual al número de variables como se muestra en la tabla 5-4.

$$R^2(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X) * Var(Y)}} \quad \text{Ec 5-1}$$

Tabla 5-4. Matriz de correlación de las variables haciendo uso de los 50 movimientos muestrados.

	Lr	Wr	Dr	Ap	V	α	δ	msnm	DO	L	H	HL
Lr	1,00	0,85	0,90	0,87	0,79	-0,24	0,16	0,43	0,45	0,81	0,88	-0,29
Wr	0,85	1,00	0,77	0,79	0,69	-0,19	0,00	0,47	0,24	0,64	0,69	-0,12
Dr	0,90	0,77	1,00	0,91	0,88	-0,30	0,19	0,26	0,48	0,91	0,92	-0,42
Ap	0,87	0,79	0,91	1,00	0,98	-0,18	0,11	0,36	0,25	0,93	0,91	-0,26
V	0,79	0,69	0,88	0,98	1,00	-0,17	0,13	0,31	0,21	0,95	0,90	-0,27
α	-0,24	-0,19	-0,30	-0,18	-0,17	1,00	-0,10	0,19	-0,14	-0,21	-0,18	0,66
δ	0,16	0,00	0,19	0,11	0,13	-0,10	1,00	-0,41	0,46	0,27	0,26	-0,55
msnm	0,43	0,47	0,26	0,36	0,31	0,19	-0,41	1,00	-0,21	0,18	0,27	0,51
DO	0,45	0,24	0,48	0,25	0,21	-0,14	0,46	-0,21	1,00	0,51	0,59	-0,56
L	0,81	0,64	0,91	0,93	0,95	-0,21	0,27	0,18	0,51	1,00	0,97	-0,44
H	0,88	0,69	0,92	0,91	0,90	-0,18	0,26	0,27	0,59	0,97	1,00	-0,40
HL	-0,29	-0,12	-0,42	-0,26	-0,27	0,66	-0,55	0,51	-0,56	-0,44	-0,40	1,00

En esta matriz se ha colocado en verde los valores con correlación positiva, es decir, existe una relación directa entre ellas, y en rojo los que exhiben una relación inversa, dentro de esta escala de color se muestran valores en amarillo o verde claro (relación directa leve) y naranja (relación inversa leve), mientras el color sea más fuerte (verde o rojo) mayor será la correlación entre dichas variables.

A primera vista se puede apreciar que la pendiente del talud prefalla posee una relación inversamente proporcional con las demás variables, pero no supera en ninguno de estos casos el 30%, y solamente posee una relación directa con H/L y msnm, lo cual es lógico ya que corresponde a una zona de alta montaña, donde a mayor altura, mayor es la pendiente del escarpe estudiado.

Si se analizan las dos variables más comunes usadas en la literatura para realizar modelos predictivos que son H/L y V se encuentra un coeficiente de correlación negativo (relación inversa) pero su valor absoluto es relativamente bajo si se compara con otros valores de la tabla como la relación entre la distancia de viaje y el volumen, cabe aclarar que dentro de los modelos utilizados actualmente se hace una transformación de los valores usando la función logarítmica en base 10 para poder uniformizar la distribución de los datos, tratamiento que se hará posteriormente.

Las dos variables de interés que si presentan una fuerte correlación son la distancia de viaje con todas las variables geométricas (Lr, Wr, Dr, Ap, V) lo cual corrobora la fuerte influencia que tienen este tipo de condicionantes.

5.1.3 Transformación haciendo uso de función logarítmica.

Como se evidencia en la Tabla 5-3 , existe una alta variabilidad en los datos que dificultan tanto la visualización como el tratamiento de los datos, para solucionar esto se pueden emplear diferentes transformaciones de estandarización o normalización de las variables, algunas de estas buscan eliminar la dimensionalidad o acotar el rango entre ciertos valores, en este caso se busca continuar con la técnica más común utilizada para casos en el estudio de remoción en masa, transformar a escala logarítmica las diferentes variables.

Debido a que el resultado de calcular el logaritmo de cero no converge a un valor con el que se pueda trabajar, se debe adaptar esta transformación a aquellas bases de datos que posean valores iguales a cero.

Para aplicar esto, basta con calcular el logaritmo de cada uno de los valores de la base de datos. Al transformar los valores y recalculer los coeficientes de variación se obtiene la tabla 5-5.

Tabla 5-5. Coeficientes de variación luego de transformar las variables independientes con la función Log+1 haciendo uso de los 50 movimientos muestrados.

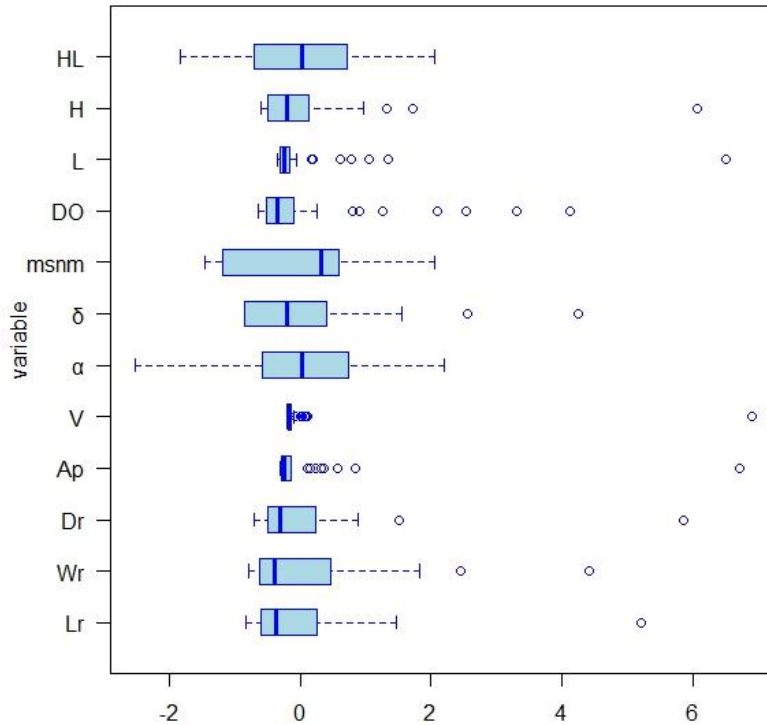
	Lr	Wr	Dr	Ap	V	α
S	0,39	0,39	0,33	0,75	1,06	0,1
\bar{X}	1,93	1,7	0,77	3,38	4,15	1,57
C,V	20%	23%	43%	22%	26%	6%

	δ	msnm	DO	L	H	H/L
S	0,73	0,17	0,74	0,48	0,36	0,07
\bar{X}	0,96	3,17	2	2,23	2	0,21
C,V	76%	5%	37%	22%	18%	33%

Es notable la mejoría en el coeficiente de variación, siendo Dr, δ , H/L y DO las únicas variables que aún no logran llegar a un valor igual o menor al 30% pero si logran una

reducción en este valor. Otro beneficio de transformar las variables es la facilidad al momento de graficar o generar una regresión entre dos variables ya que linealiza distribuciones similares a V como lo muestran la figura 5-1.

Figura 5-1. Diagrama de cajas y bigotes con las variables transformadas haciendo uso de los 50 movimientos muestreados.



Fuente: Propia.

Se puede apreciar que incluso al tener dos variables con magnitudes tan lejanas como lo son en ángulo de fricción equivalente (H/L) y el volumen (V) la transformación $\log+1$ permite agrupar todos los diagramas de caja y bigotes en una sola figura y tener una visión total de la base de datos, mostrando de manera simultánea la existencia de casos atípicos como en el ángulo de talud prefalla (α) donde, el valor que se aparta de la distribución común corresponde al movimiento Z7_21 con valor de 10° con la horizontal (Ver Anexo A y D). Visualizando en una toma aérea como lo muestra la figura 5-2, podemos notar que el material desprendido no corresponde a la totalidad de la superficie de falla, parte de este volumen permanece en la fuente con un aparente movimiento de reptación evidenciado por las grietas de tracción, este comportamiento es típico de suelos que tienen resistencias

pico similares a materiales ya movilizados, este es solo un ejemplo de las alertas y análisis que se puede obtener de una visualización general de los datos.

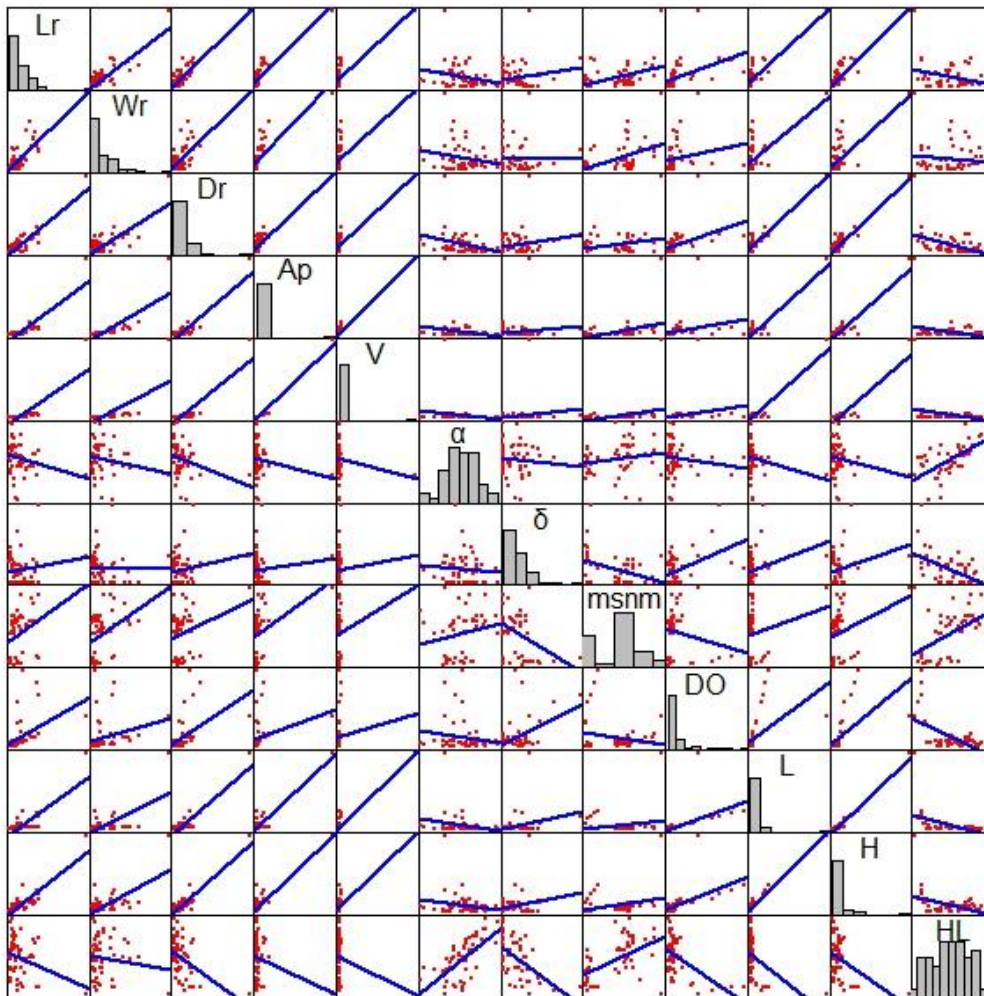
Figura 5-2. Ortofoto del movimiento Z7_21 con las zonas principales identificadas.



Fuente: Propia.

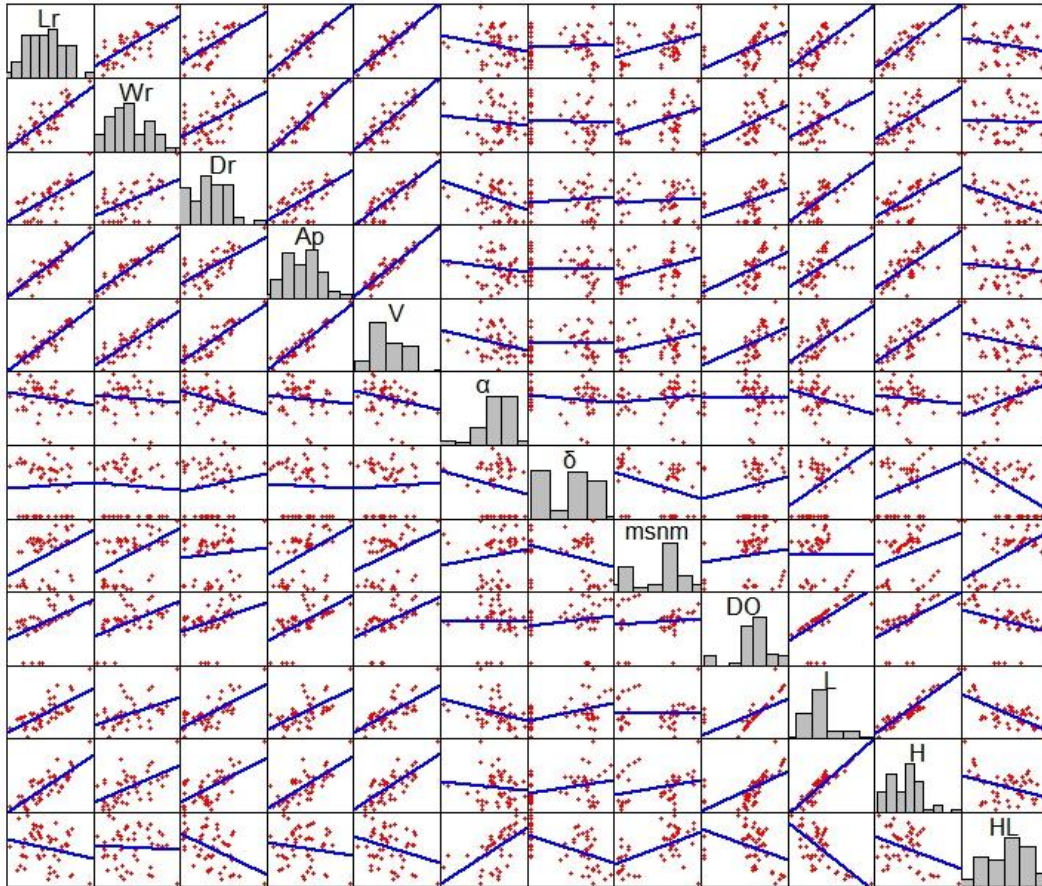
La linealización de las variables genera, como es de esperar, una visualización de las correlaciones y distribuciones más clara, expandiendo la nube de puntos y relacionando de mejor manera una curva de regresión, la Figura 5-3 y la Figura 5-4 muestran una comparativa en la visualización de la distribución de los datos. La diagonal principal posee el histograma de la variable que muestra la casilla, por ejemplo en la casilla 2,2 (fila 2, columna 2) se muestra el histograma de los 50 anchos de superficie de falla (W_r) medidos mostrando un valor sesgado positivamente, las casillas complementarias muestran gráficamente la correlación entre dos variables, donde la línea azul corresponde a la aproximación lineal de mejor ajuste y los puntos rojos a la distribución de los valores de dichas variables, como ejemplo se puede tomar la casilla 2,3 donde se comparan el ancho de la superficie de falla (W_r) con la profundidad de la superficie de falla (D_r).

Figura 5-3. Correlación entre variables con la base de datos original mostrada de manera gráfica haciendo uso de los 50 movimientos muestrados.



Fuente: Propia.

Figura 5-4. Correlación entre las variables mostrada de manera gráfica luego de aplicar una transformación utilizando la función Log+1.



Fuente: Propia.

Lo primero que resalta de la figura 5-3 y figura 5-4, es un cambio en el tipo de distribución, pasan de ser la mayoría con un sesgo positivo fuerte en los histogramas que se encuentran en la diagonal principal a acercarse a una distribución normal. Al transformar las dimensiones también se modifican las nubes de puntos, lo que facilita la construcción de un modelo lineal asignando una curva de regresión lineal. Nótese que las matrices no son simétricas, es decir, no poseen la misma forma sobre y bajo la diagonal principal, esto ocurre ya que las columnas se construyen con la variable que se muestra en la diagonal principal ubicada en el eje de las abscisas, si se toma en la figura 5-3 la columna correspondiente al volumen (quinta columna) todas las gráficas allí expuestas tienen en su

eje de ordenadas el volumen de la masa desplazada (V) y en el eje de las abscisas cada una de las diferentes variables restantes.

Luego de realizar el trabajo exploratorio y de reconocimiento de las diferentes variables y como se relacionan entre sí, es necesario identificar si son aplicables modelos predictivos como técnicas de machine learning, para esto se busca evaluar:

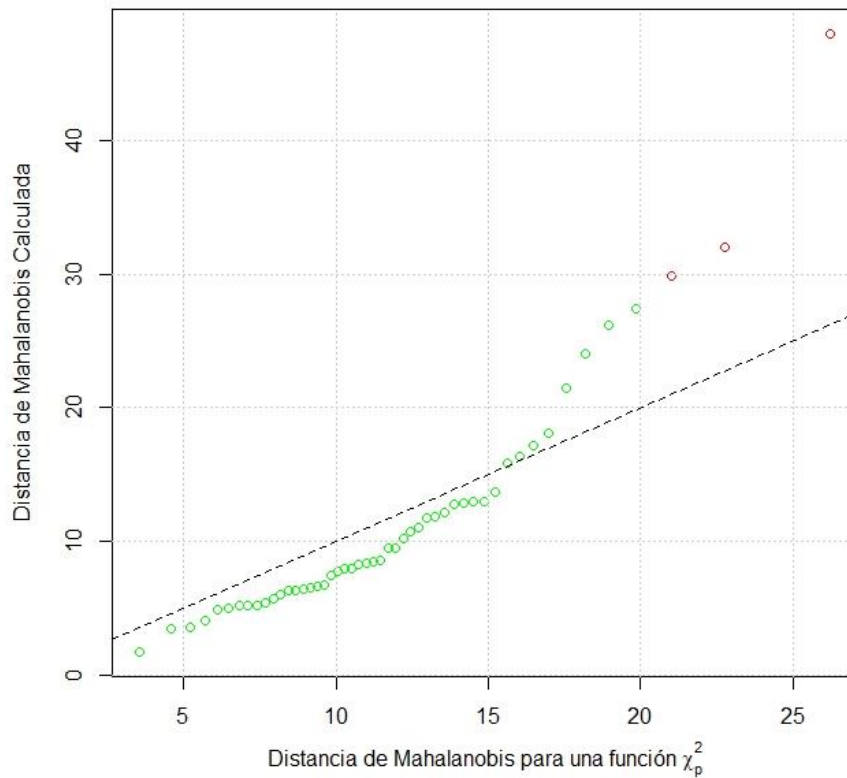
- Si la base de datos cumple o no con el principio de normalidad, es decir, si las distribuciones de los datos se acoplan a una función de distribución normal multivariada, para esto se utiliza la distancia de Mahalanobis (sección 5.2)
- Si existen datos atípicos teniendo en cuenta todas las variables de manera simultánea, esto también lo indica el cálculo de la distancia de Mahalanobis (sección 5.2).
- Si existen variables prescindibles en el cálculo predictivo de posibles modelos, es decir, si existen algunas variables que puedan ser omitidas sin afectar el comportamiento de la base de datos, para esto se utiliza el análisis por componentes principales (sección 5.3).

5.2 Distancia de Mahalanobis

Una propiedad importante del vector d_M (definido en la sección 2.2.6, Ec. 2-5) es que si se ordena de manera ascendente y la muestra sigue una distribución normal multivariada la probabilidad de ocurrencia de cada una de las distancias sigue una distribución de probabilidad Chi-cuadrado, lo que permite encontrar datos atípicos dada una probabilidad de ocurrencia y verificar si la base de datos sigue una distribución normal multivariada.

En la figura 5-5, se realiza una comparación entre los valores ordenados de las distancias de Mahalanobis obtenidas (eje y), y el valor que debería dar si la distribución de los datos corresponde a una función chi-cuadrado (eje x), para esto se realiza el cálculo con la función inversa de chi-cuadrado $qchisq$ en R donde, al ingresar el percentil y los grados de libertad, en este caso 12 (cantidad de variables), se obtiene el valor de la distancia de Mahalanobis para ese percentil.

Figura 5-5. Distribución de la distancia de Mahalanobis comparada con una distribución Chi-cuadrado de la base de datos.



Fuente: Propia.

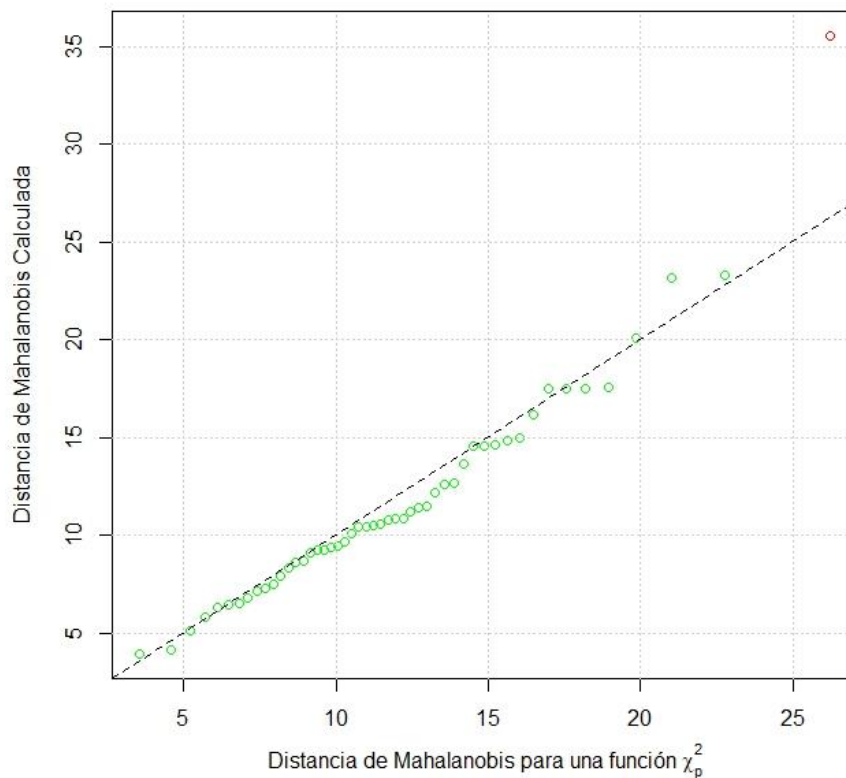
La línea punteada de la Figura 5-5 representa un ajuste perfecto de la distribución de los datos a una función Chi-Cuadrado, es decir que si una base de datos cumple con el principio de normalidad en su totalidad, todas las distancias de Mahalanobis quedarán colocadas sobre esta línea, se puede observar la discordancia de los datos con esta recta lo que propone una ausencia de cercanía con la distribución normal multivariada. Otro punto que resaltar es la existencia de datos atípicos, si se toma como referencia el 99.7% de los datos equivalente en una distribución normal a $\bar{X} \pm 3\sigma$, existen dos datos atípicos correspondientes a los movimientos Z12_14 y Z11_36_1 con una distancia de Mahalanobis de 29.95 y 46.96, bajo el entendido que el límite al 99.7% es de 29.79.

El vector de distancias se obtiene utilizando la Ec 2-5, donde X es la matriz dada por los valores medidos para cada uno de los 50 casos de estudio con sus correspondientes 12

variables numéricas, dando como resultado una matriz de 50x12, \bar{X} es la matriz de promedios de cada una de las doce variables, este valor es igual para cada uno de los 50 casos y se pueden tomar de la Tabla 5-3 obteniendo una matriz de 50x12 donde todas las filas son iguales, S^{-1} es la matriz inversa de covarianzas obtenida en la Tabla 5-2, se aplica el producto que se encuentra dentro del radical de la Ec 2-5 obteniendo una matriz de 50x50 simétrica y se aplica raíz cuadrada. Las distancias de Mahalanobis corresponden a la diagonal principal.

Realizando el mismo proceso para la base de datos transformada con log+1 se obtiene la figura 5-6.

Figura 5-6. Distribución de la distancia de Mahalanobis comparada con una distribución Chi-cuadrado con la base de datos transformada usando la función Log+1.



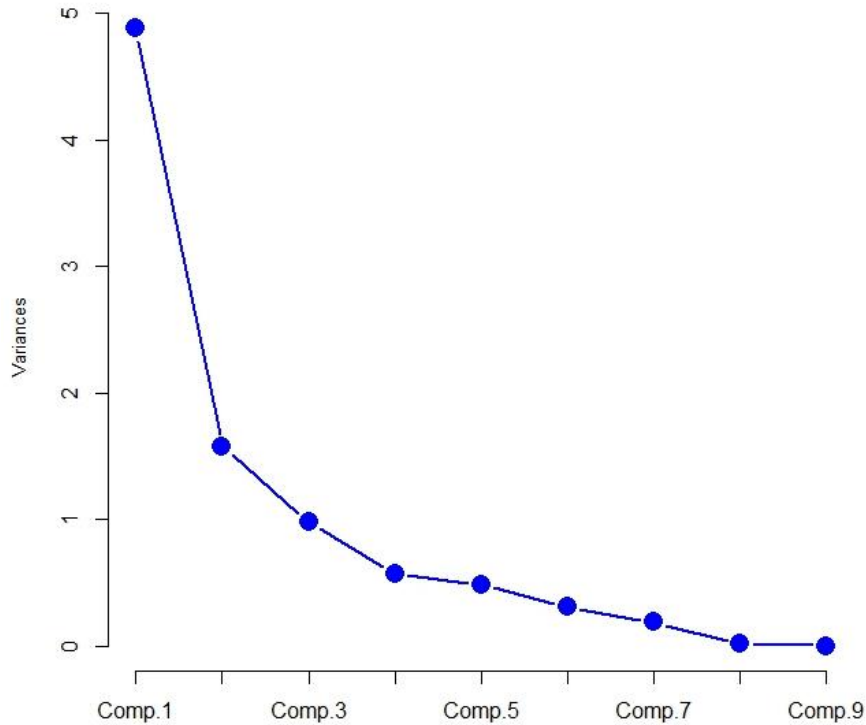
Fuente: Propia

Es notable la mejora en el ajuste de la distribución a la recta 1:1, lo que sugiere un mejor ajuste a una distribución normal multivariada, es decir, esta base de datos transformada permite el uso de técnicas como la regresión lineal multivariada dando más opciones al momento de plantear modelos predictivos.

5.3 Análisis por Componentes Principales

El algoritmo utilizado para realizar este proceso es el proporcionado por la función *princomp* de la librería estándar de R. El análisis por componentes principales busca disminuir el número de variables existentes en una base de datos reemplazándolas por un nuevo conjunto de variables llamadas “componentes”, es decir, se busca pasar de ancho de la masa desplazada (W_r) volumen de la masa desplazada (V), etc a componente 1 (C1) componente 2 (C2), etc, esto con el objetivo de disminuir la dimensionalidad del problema pasando de 9 variables a 2 sin afectar de manera significativa el conjunto de datos.

Para esto la rutina *princomp* realiza un proceso de optimización maximizando la varianza de la componente principal 1 (C1), luego maximiza la varianza de la componente principal 2 (C2) y así sucesivamente, siendo C1 la que posee la mayor variabilidad, En la figura 5-7 se muestran las varianzas de cada componente siendo las tres primeras las más relevantes. Al calcular la proporción de cada varianza dentro de una varianza acumulada de las nueve componentes resulta en 52,63% y 16,63% para las componentes 1 y 2 respectivamente, dando un porcentaje de varianza acumulada de 69,23%, es decir, el 69.23% de la varianza se encuentra en las dos primeras componentes permitiendo realizar un análisis bidimensional utilizando únicamente C1 y C2.

Figura 5-7. Importancia de las componentes principales.

Fuente: Propia.

La figura 5-8 representa un campo bidimensional haciendo uso de las dos primeras componentes como ejes principales, los vectores en verde corresponden a cada una de las nueve variables originales (L_r , W_r , V , etc.). Se puede notar una uniformidad en la magnitud de cada uno de los nueve vectores lo que propone una importancia similar entre las variables originales, concluyendo que es necesario utilizar la totalidad de número de variables dentro de la metodología del cálculo de la distancia de viaje mostrada en el Capítulo 6. En caso de que algún vector de las variables originales exhibiera una magnitud evidentemente menor a las otras, podría descartarse para el cálculo de la distancia de viaje

predicciones, en este caso se decidió utilizar la totalidad de la información realizando una transformación con funciones logarítmicas.

El análisis por componentes logra dar una visualización más profunda del comportamiento de los datos evidenciando si existen variables que pueden omitirse para reducir gasto computacional o descartar hipótesis que el usuario establezca de manera preliminar.

6 Estimación de la distancia de viaje de movimientos en masa con técnicas machine learning

Este capítulo contiene el desarrollo, calibración, y validación del modelo de *Machine Learning* conocido como *Random Forest* aplicado para el cálculo de la distancia de viaje L haciendo uso de la totalidad de la información recabada para los 50 movimientos de la muestra tomada.

Haciendo uso del lenguaje de programación R, se utiliza la librería *ranger* que no solo permite generar un modelo de *Random Forest*, sino que también facilita su calibración y mejora los tiempos de procesamiento comparados con librerías similares como *randomForest*. Para lograr la producción del modelo se tienen en cuenta los siguientes pasos:

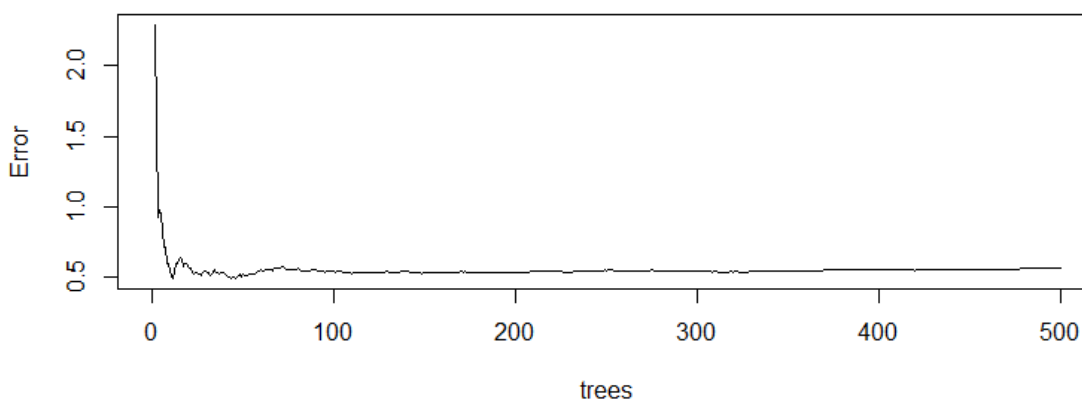
1. Se prepara la base de datos dejando únicamente la información a utilizar, para una versión 1.0 se utilizarán únicamente las variables geométricas con un total de 12 y dos variables categóricas que será el tipo de movimiento y el tipo de obstáculo tal y como se muestra en el Anexo A, fijando la distancia de viaje como variable de respuesta.
2. Se genera un primer modelo de acercamiento para evaluar cuántos árboles se necesitan para estabilizar el error y fijar este valor en futuros cálculos, esto reduce drásticamente el tiempo de procesamiento.
3. Para obtener el modelo con el menor error se genera una matriz con cada combinación posible de hiperparámetros, dando un total de 550 modelos.
4. Se establecen los valores para cada uno de los cuatro hiperparámetros consignados en la Sección 2.3.2 del modelo óptimo los cuales rigen las características generales de cada árbol que compone el bosque y se calcula el error relativo para cada uno de

los movimientos en masa comparando el valor calculado de la distancia de viaje con dicho modelo y el valor real medido en campo.

6.1 Modelo generado

Para el modelo generado se utilizó como primera aproximación un modelo que utiliza el 70% de la información (35 casos aleatorios) para armar los árboles de decisión y el 30% restante como grupo de control para comparar los valores reales con los calculados. Se obtiene que para un valor de 400 árboles el error ya se estabiliza como lo muestra la Figura 6-1, por lo tanto, los modelos generados en el paso 3 se fijará el número de árboles de decisión en 400, ya que, de no hacerse, este valor varía junto con los demás hiperparámetros. Como ejemplo se pueden tomar dos que son el número de árboles (entre 1 y 400) y la cantidad de variables posibles de las cuales se escoge en cada nodo (entre 1 y 12), el número de modelos esta dado por el producto de las posibilidades entre los dos hiperparámetros resultando en 4800 diferentes combinaciones, al tener en cuenta los 3 parámetros adicionales resulta en 220000 la cantidad de modelos a evaluar aumentando considerablemente el tiempo de cálculo, esta es la razón por la cual se fija el tamaño del bosque.

Figura 6-1. Variación del error con respecto al No. de árboles.



Fuente: Propia.

La matriz de hiperparámetros cuenta con 4 columnas y 550 filas, este último valor corresponde a la cantidad de modelos que se generarán y se obtendrá el óptimo. Una vez

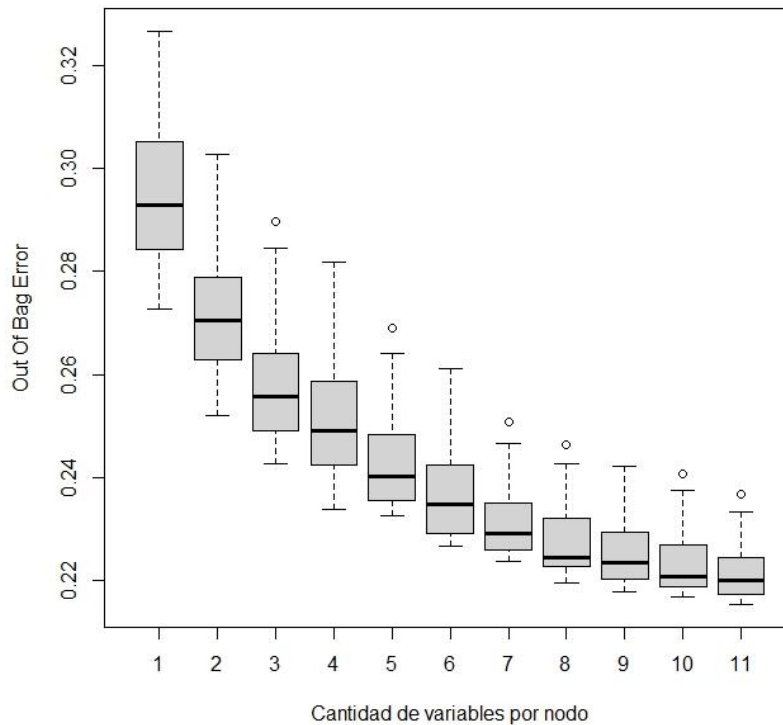
realizado el paso anterior se obtiene que el modelo con el menor error es aquel que trabaja con un número de variables por nodo de 11, una cantidad de nodos de 2 y un tamaño de muestra del 80% como se muestra en la tabla 6-1.

Tabla 6-1. Muestra de los 10 modelos con el menor error.

	Número de variables por Nodo (Mtry)	Cantidad de nodos por árbol (Node_Size)	Porcentaje de casos para entrenamiento (Sample_Size)	Error (OOBE)
1	11	2	80%	2,1547E-01
2	11	4	100%	2,1553E-01
3	11	3	80%	2,1555E-01
4	11	3	100%	2,1558E-01
5	11	4	80%	2,1559E-01
6	11	1	80%	2,1580E-01
7	11	2	100%	2,1587E-01
8	11	5	100%	2,1601E-01
9	11	1	100%	2,1612E-01
10	11	5	80%	2,1621E-01

6.2 Variación del error

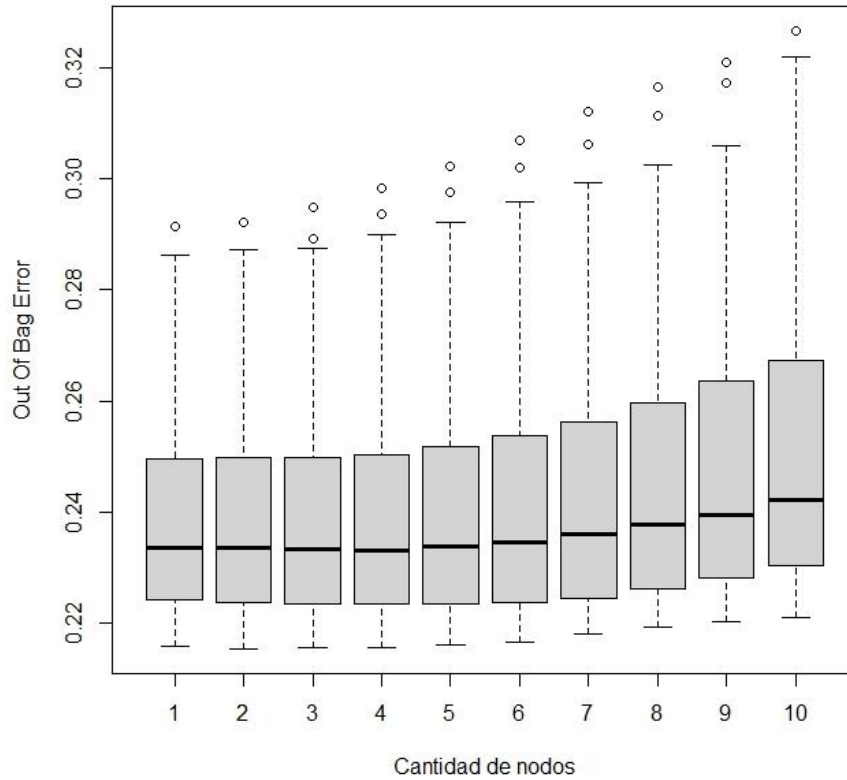
Como se puede ver, la variación del error modificando número de variables aleatorias que toma en cada nodo (Mtry) posee un comportamiento exponencial inverso, siendo 11 el que presenta el menor error promedio correspondiendo al modelo escogido.

Figura 6-2. Variación del error con respecto a las variables tomadas en cada nodo.

Fuente: Propia.

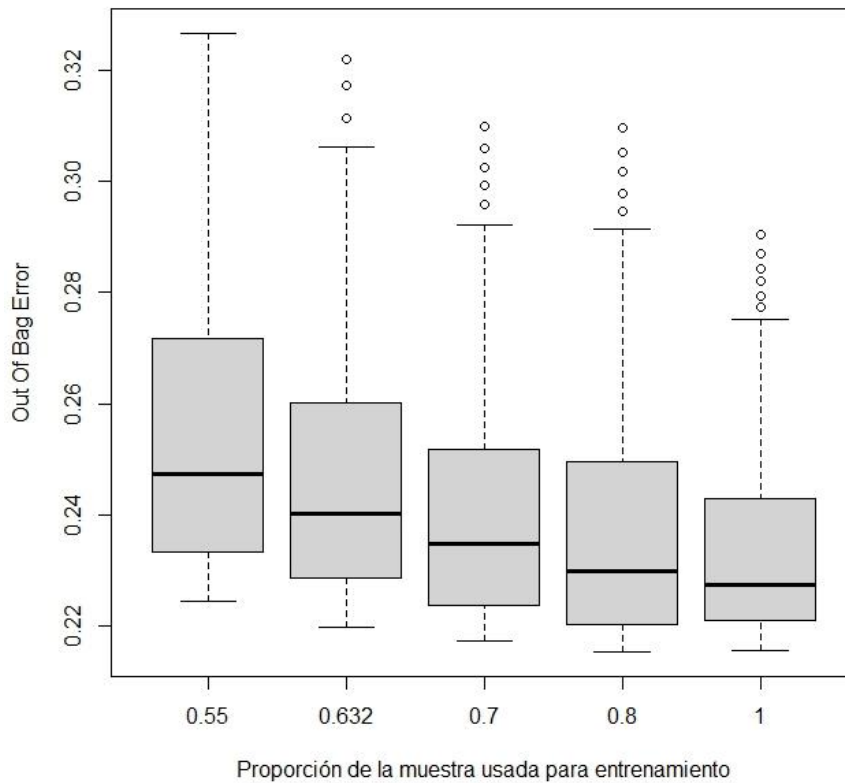
En la Figura 6-3, se muestra un ligero aumento del error conforme aumenta la cantidad de nodos que se generan en cada árbol de decisión siendo 2 el valor con menor error, este hiperparámetro presenta una menor sensibilidad que $Mtry$ para la base de datos recopilada lo cual permite fijar un valor de este hiperparámetro para reducir la cantidad de modelos que se ponen a prueba reduciendo así el costo computacional.

Figura 6-3. Variación del error con respecto al número máximo de nodos.



Fuente: Propia.

La variación del error con respecto al tamaño de la muestra que toma el modelo para entrenarse se mantiene relativamente constante, pero existe una reducción notable en la variabilidad de los resultados cuando la muestra de entrenamiento llega al 80%, y se reduce aún más esta variabilidad si el porcentaje toma un valor del 100% pero el modelo que presenta menor error se encuentra al tomar el 80% de la muestra para el entrenamiento, queda a decisión del usuario optar por la variabilidad del error, o el mínimo error registrado.

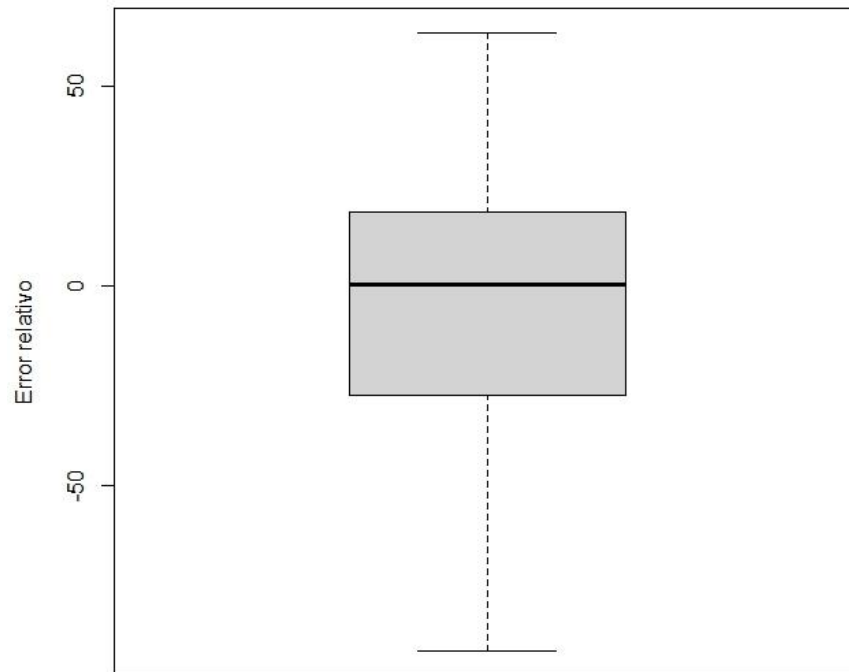
Figura 6-4. Variación del error con respecto al tamaño de la muestra de entrenamiento.

Fuente: Propia.

Una vez calibrado el modelo se procede a validar el modelo, realizando una comparativa entre el error relativo que presenta cada uno de los 50 movimientos estudiados dando como resultado Figura 6-5, si se compara con los modelos estudiados en el Capítulo 4, el rango intercuartílico es de 43.89% con una mediana de 0.23% y un promedio de -26.56% (producto del dato atípico identificado en la sección 5.2) lo que supone una subestimación y sobreestimación similar de las estimaciones con una relativa baja variabilidad en el error.

La validación con un caso de estudio externo a la base de datos tomada para entrenar y calibrar el modelo se desarrolla en el capítulo 7 junto con su comparación con respecto a los otros modelos predictivos establecidos en el capítulo 4.

Figura 6-5. Error relativo del calculado con el Random Forest. haciendo uso de los 50 movimientos muestrados.



6.3 Comentarios finales

Para el desarrollo de este capítulo se buscó una alternativa a los modelos de regresión lineal que se encuentran comúnmente en la literatura, logrando implementar variables de distintos tipos, dimensiones y ordenes de magnitud, siempre y cuando se realicen los procesos de análisis estadísticos previos para garantizar la estabilidad en los resultados, mostrando al usuario cómo calibrar, validar y desarrollar un modelo de *Machine Learning* teniendo en cuenta los parámetros que influyen en el resultado, tomando el *Out of Bag Error* como indicativo de calidad.

7 Probabilidad de excedencia de la distancia de viaje

En este capítulo se presenta un modelo propuesto para el cálculo de probabilidades de excedencia con base en un algoritmo de Random Forest. En la sección 7.1 se presenta el procedimiento para el cálculo de las probabilidades de excedencia y su aplicación en un movimiento de la zona de estudio. En la sección 7.2 se presentan resultados del cálculo de las distancias de viaje utilizando los diferentes modelos que se ilustraron en el capítulo 4.

7.1 Modelo propuesto para el cálculo de excedencia de la distancia de viaje con base en un algoritmo de *Random Forest*

Como se menciona en la sección 1.1, uno de los problemas actuales relacionados con los modelos basados en correlaciones lineales es el resultado de un único valor con una gran incertidumbre. Similar a lo aplicado en el campo de la hidrología (Monsalve, 1995), es posible calcular un rango de alcance dando como resultado una curva de probabilidad de excedencia, estos métodos dependen directamente de la cantidad de datos disponibles, y para solucionar esto se pueden realizar simulaciones de Montecarlo. Al usar el modelo generado con *Random Forest* como “ecuación de trabajo” es posible simular y obtener n distancias de viaje tomando en consideración la incertidumbre que presentan las variables independientes obteniendo una metodología que da como resultado un valor medio calculado acompañado de un intervalo de confianza.

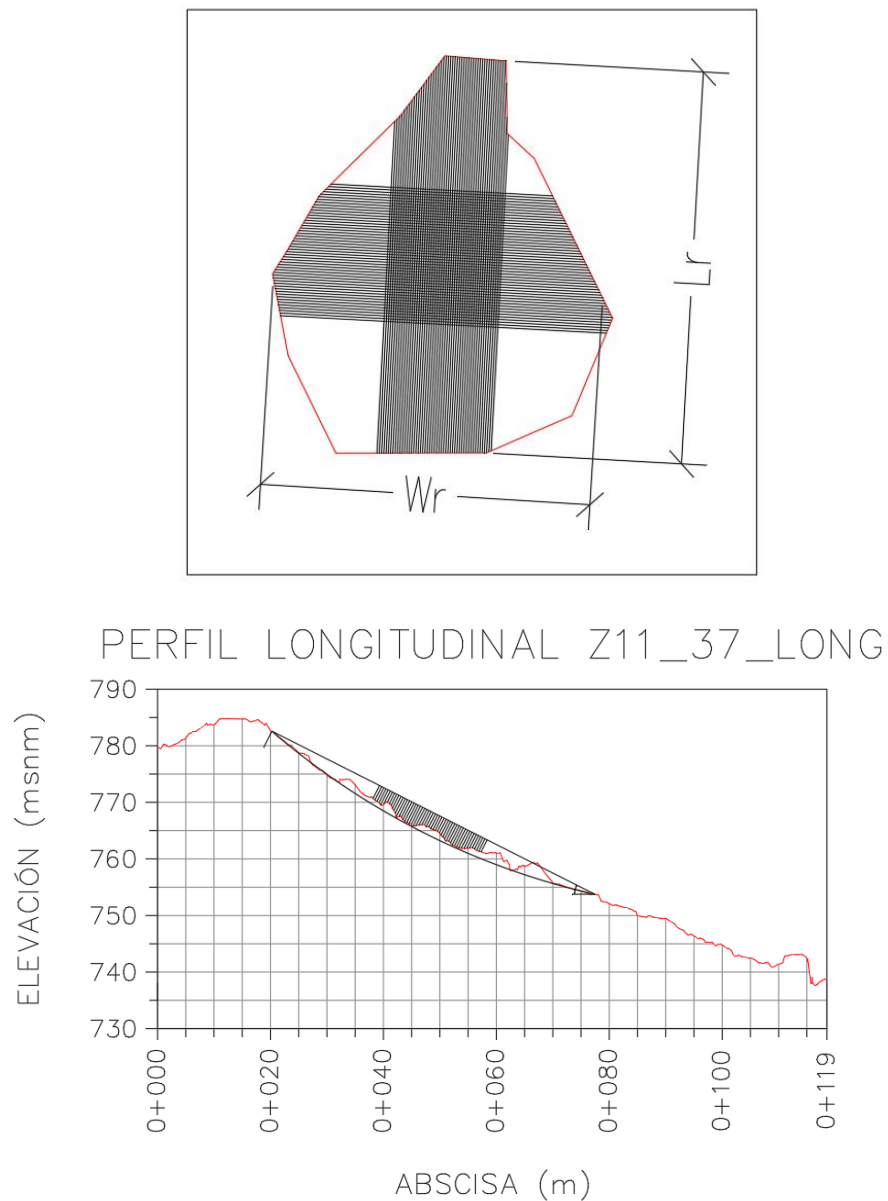
Dentro de las doce variables (descritas en la sección 5.1) integradas al modelo de RF, las relacionadas directamente con la geometría de la superficie de falla son las que presentan

una incertidumbre al momento de ser medidas, para solucionar esto, se obtiene un valor promedio y una desviación estándar para L_r , W_r y D_r midiendo en el tercio medio de la superficie de falla posibles 50 valores como se muestra en la figura 7-1, este procedimiento se realiza con cada uno de los movimientos registrados.

Para obtener un valor proporcional de desviación estándar (dada en metros) a un caso de estudio se debe calcular el coeficiente de variación promedio de todos los casos, para lo cual se obtuvo 0,075; 0,109 y 0,198 para L_r , W_r , y D_r respectivamente. La función de probabilidad usada para obtener los valores aleatorios iniciales y calcular las simulaciones es de tipo normal haciendo uso de la función *rnorm* incluida en el paquete estándar de R.

En el Caso de L_r , W_r , y D_r , se tienen como valor de entrada el valor obtenido en campo en caso de que el movimiento ya hubiera ocurrido buscando una validación del modelo, o puede aplicarse a un valor teórico calculado a través de un estudio geotécnico de estabilidad que permita encontrar la superficie de falla con el menor factor de seguridad. Los otros valores de entrada son la desviación estándar como producto del coeficiente de variación y el valor medido o calculado según sea el caso. Es importante tener en cuenta que es necesario transformar estos valores aleatorios con la función $\text{Log}+1$ ya que el modelo de RF está construido bajo esta condición. En la Figura 7-1 se muestra un ejemplo mediante el cual se tomaron las medidas en el tercio medio de la superficie de falla, obteniendo diferentes valores de L_r , D_r y W_r y así poder calcular la desviación estándar de cada una de estas tres variables para este movimiento, este proceso se realizó con todos los 50 movimientos de la base de datos.

Figura 7-1. Medición de L_r , D_r y W_r en el tercio medio de la superficie de falla. Movimiento Z11_37 (Girón, Santander).



Fuente: Propia.

Los valores de A_p se obtienen calculando el área de una elipse usando $L_r/2$ y $W_r/2$ como radios, V se calcula haciendo uso de la ecuación propuesta en la Guía Metodológica (Ávila, y otros, 2016), el componente aleatorio se encuentra en el coeficiente de expansión de material tomando 1,48 como valor medio (promedio entre los valores mínimos y máximos

propuestos) y una desviación estándar de 0,15 asumiendo que la diferencia entre el valor máximo y mínimo corresponde a un rango que contiene el 99,8% de los datos.

Estos valores aleatorios entran al modelo de RF obteniendo una distancia de viaje, este proceso corresponde a una única simulación, el proceso de ingreso de las variables se realiza usando el código del Anexo C, utilizando la función $predict(X, Y)$ donde X es el modelo generado por la función $ranger$ y Y es el vector de valores de cada una de las 12 variables [Lr, Wr, Dr, Ap, V, α , δ , msnm, DO, L, H, H/L] que requiere el modelo como parámetros de entrada para calcular la distancia de viaje donde L inicialmente es igual a 0 ya que es el valor objetivo a calcular. La cantidad de simulaciones está dada por el intervalo de confianza que se quiera obtener, en este caso es del 99,8% o tres desviaciones estándar por encima y por debajo del valor promedio, es decir, un mínimo de 1000 simulaciones, este proceso se realiza produciendo un valor aleatorio dentro del rango que permita el coeficiente de variación obtenido con la información de la Figura 7-1 para cada una de las variables que compone el vector Y, obteniendo diferentes valores de L.

Para buscar una precisión al segundo decimal se opta por realizar 10000 simulaciones ya que el aumento en el desgaste computacional es mínimo (pasa de 7 a 8 segundos), esto puede variar dependiendo de la cantidad de parámetros de entrada. Para construir la curva de probabilidad de excedencia (Figura 7-2), se construye un vector de probabilidad donde el valor de cada componente está dado por la Ec.7-1, donde N_s es el número de simulaciones e i es la posición dentro del vector variando entre 1 y N_s

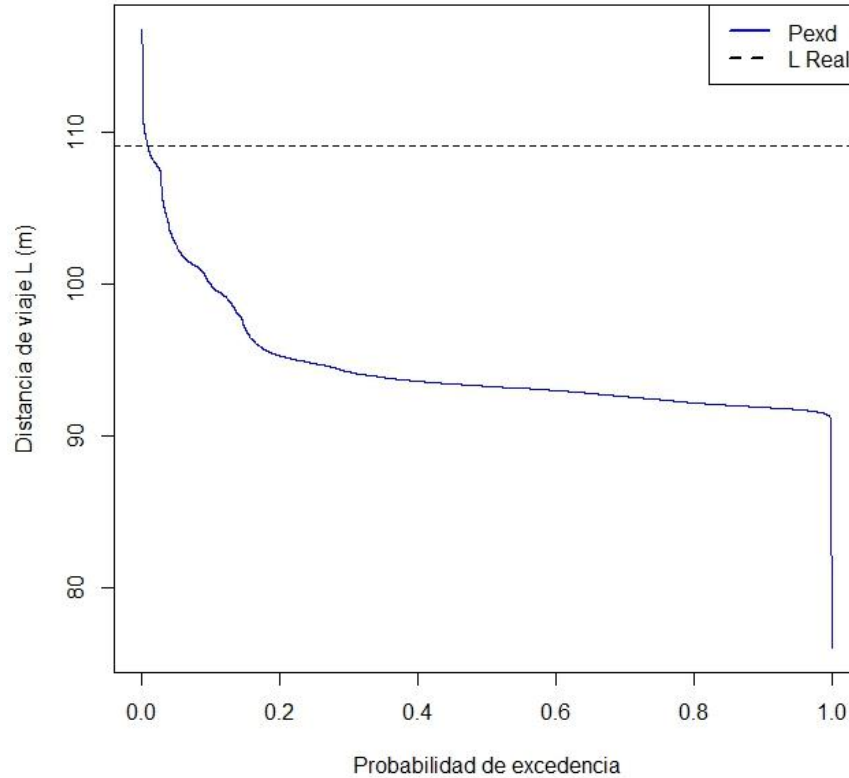
$$P[i] = \frac{i}{N_s} \quad \text{Ec. 7-1}$$

Luego se ordena el vector de distancias de viaje calculado L de mayor a menor y se traza la curva con P en las ordenadas y L en las abscisas.

Para probar el funcionamiento del algoritmo se reservó un movimiento ubicado en el municipio de Tona, el levantamiento está identificado como Tona_A2 y el movimiento con el nombre de Z11_CE1. Este caso tiene las siguientes características: Lr=92.52m, Wr=40.73, Dr=4.27, $\alpha=40^\circ$, $\beta=31^\circ$, msnm=2090, TO=2 (Obra Vial), DO=31.01m y TM=3 (Flujo). Nótese la ausencia de V y Ap, estas dimensiones son calculadas con Lr, Wr y Dr, completando así los 12 parámetros necesarios para el funcionamiento del modelo. En la

Figura 7-2 se encuentra la relación entre la probabilidad de excedencia contra la distancia de viaje correspondiente, el valor real medido de L es de 109.08m lo cual está dentro del rango de variación que proporciona la curva.

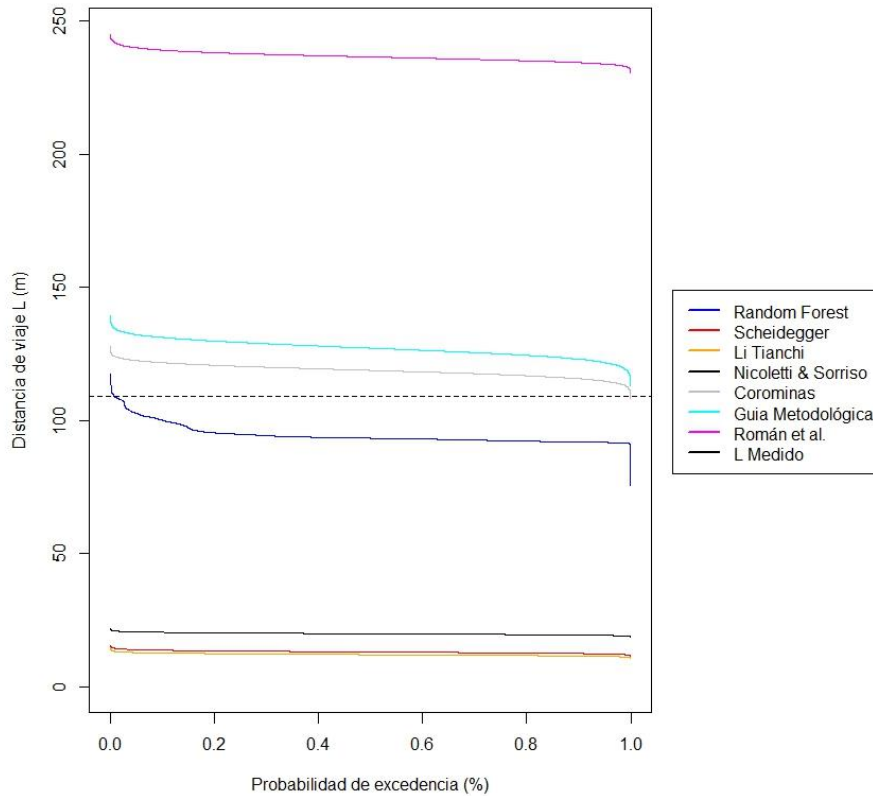
Figura 7-2. Probabilidad de Excedencia de distancia de Viaje. Movimiento Z11_CE1. (Tona, Santander)



Fuente: Propia.

7.2 Comparación de metodologías para el cálculo de probabilidades de excedencia de distancias de viaje

Como fase final de la metodología de cálculo que se expondrá en el Capítulo 8, se realiza una comparativa entre los diferentes modelos realizando de manera similar las simulaciones de Montecarlo, variando el parámetro de entrada que cada uno exige siendo en la gran mayoría el volumen dando como resultado la figura 7-3.

Figura 7-3. Comparativa de la probabilidad de excedencia para los diferentes modelos.

Fuente: Propia.

Reafirmando los comportamientos en la distribución del error presentados en el Capítulo 4, los modelos de Scheidegger, Li Tianchi y Nicoletti & Sorriso presentan valores un 80% menores subestimando el valor real, los mejores ajustes resultan de las ecuaciones propuestas por Corominas, la Guía Metodológica, y el modelo presentado en este capítulo (sección 7.1) siendo este último el único que no utiliza H como parámetro de entrada pero con un gasto computacional mayor. El modelo generado por Román et al. (2024) proporcionan valores conservadores duplicando el valor real, lo cual no es necesariamente un resultado indeseado ya que se puede interpretar con un “factor de seguridad” de 2, valor común dentro de la práctica.

8 Metodología para la estimación de distancias de viaje de movimientos en masa

En el presente capítulo se desarrollan una propuesta fundamentada sobre buenas prácticas ingenieriles para la aplicación de modelos empíricos y el propuesto en el capítulo 7 para la estimación de distancias de viaje de movimientos en masa teniendo en cuenta la información anteriormente desarrollada en los capítulos 3 a 7. Con base en las lecciones aprendidas en dichos capítulos, se propone una metodología a seguir teniendo en cuenta las ventajas y limitaciones de los modelos anteriormente presentados.

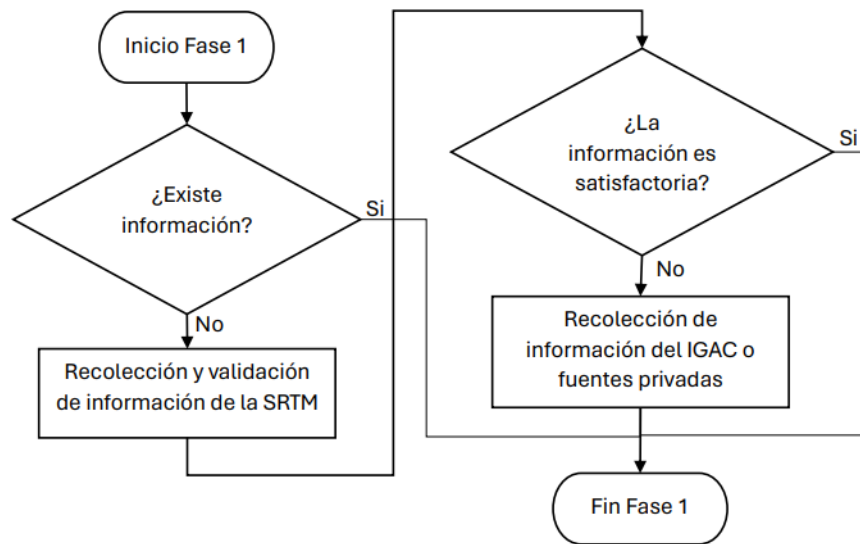
La metodología propuesta se divide en cuatro fases: La fase 1 corresponde a los lineamientos para la ejecución de actividades de recolección, levantamiento y ensamblaje de información morfométrica. La fase 2 corresponde a los lineamientos para la verificación de la base de datos empleada. La fase 3 corresponde a la validación de los modelos para la estimación de las distancias de viaje. La fase 4 corresponde al cálculo de la probabilidad de excedencia de la distancia de viaje.

8.1 Fase 1: Recolección, levantamiento y ensamblaje de información

En esta fase es importante definir el área de estudio, la cual probablemente corresponderá a una zona impactada por los procesos de movimiento en masa. Dados los métodos propuestos para la estimación de distancias de viaje, se recomienda definir una escala de trabajo de entre 1:2000 y 1:10000. Es recomendable que la escala de trabajo permita una observación razonablemente detallada de las dimensiones de los movimientos en masa contenidos en su interior, pero que al mismo tiempo brinde una ventana espacial que abarque varios deslizamientos en la zona para su consecuente validación.

La información fundamental para la aplicación de la presente metodología corresponde sustancialmente a información morfométrica. En la figura 8-1 se presenta el diagrama en el que se sugiere el flujo de trabajo para la elección de la información morfométrica necesaria para la elaboración de la primera fase del estudio. Dependiendo del contexto, existen diversas fuentes de información, desde libre y abierta hasta información de pago. Por ejemplo, existen datos provenientes de misiones topográficas internacionales, como los ofrecidos por la Shuttle Radar Topography Mission (SRTM). Los datos proporcionados por la SRTM son libres y abiertos, pero tienen una resolución de 30 x 30 m. Si los datos de la SRTM no fuesen satisfactorios por cuestiones de resolución, en el caso colombiano el IGAC centraliza información morfométrica y topográfica a diferentes escalas. La información puede estar restringida a ciertas zonas del territorio colombiano y en otros casos se requiere la compra directa de la información. Es importante encontrar un balance entre la calidad de la información proporcionada y la resolución obtenida para el estudio y se privilegia la información obtenida de levantamientos topográficos de la zona de estudio ya que presentan la información más reciente y se puede controlar la calidad y detalle de esta.

Se recomienda, en cada caso realizar el análisis del ajuste de los Modelos de Elevación digital del terreno, a las geformas presentes en la zona de análisis. Por ejemplo, evaluar con detalle el ajuste de las curvas de nivel a ríos, vías o escarpes de laderas o incluso de deslizamientos históricos.

Figura 8-1. Fase 1. Diagrama de flujo. Recolección de información.

Fuente: Propia.

De cualquier modo, y como punto de buena praxis se recomienda la recolección y gestión de información de amplia índole, aun cuando no haga parte del estudio primario y no afecte la consecución de resultados. En este caso se recomienda la recolección de orden geológico, geomorfológico, de cobertura/uso del suelo, hidrológico, todas al menos en la escala de trabajo proporcionada.

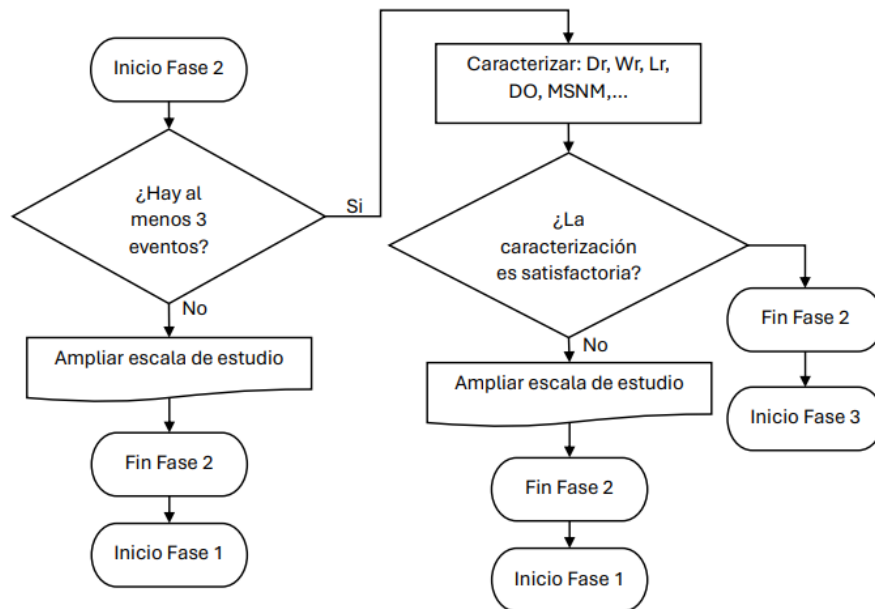
8.2 Fase 2: Verificación de información en la zona de estudio

La fase 2 se desarrolla con base en la información recolectada en la fase 1. Para la aplicación de los modelos empíricos presentados, es recomendable valorar la situación de los movimientos en masa en el área de interés para valorar a priori la cantidad de movimientos que se encuentran en la zona. Como se mostró anteriormente en el capítulo 3.4.3, en contextos como el colombiano, los movimientos en masa que fluyen decenas o centenas de metros (incluso kilómetros) desde la fuente o superficie de falla son aproximadamente el 75% del total de los movimientos encontrados en los registros de inventario. En ese sentido, con una escala espacial justa, es posible abarcar un área suficiente que permita la delimitación de diferentes deslizamientos históricos que den pautas del comportamiento típico de los deslizamientos en la zona de estudio.

En ese sentido se recomienda la reconstrucción de al menos tres deslizamientos (cuantos más sea posible, mejor) que se hayan desarrollado en el área. Una herramienta importante para tener en cuenta en el contexto colombiano es el Servicio de Información de Movimientos en (Servicio Geológico Colombiano, 2024).

Se recomienda delimitar cada movimiento en masa y definir sus dimensiones características, así como su distancia de viaje. Esta fase es de vital importancia porque, como se ha visto en el capítulo 4, el uso de los modelos empíricos está sujeto a un cierto nivel de incertidumbre, por lo tanto, es importante realizar a priori un diagnóstico de la situación de los movimientos en masa en la región de estudio. Como resultado de esta fase se requiere la caracterización de las distancias fundamentales como la pendiente, el volumen de la masa desplazada, y el área paramétrica entre otras.

Si el análisis de la información en este nivel de detalle no arroja resultados satisfactorios por diversos motivos; como la falta de deslizamientos históricos en la zona de estudio o la imposibilidad de adquirir las variables anteriormente mencionadas dada la falta de resolución del modelo digital de elevación, se recomienda terminar esta fase y comenzar el proceso nuevamente desde la fase 1. La figura 8-2 resume en un diagrama de flujo los pasos necesarios para la consecución de la Fase 2.

Figura 8-2. Fase 2. Validación de la información de la zona de estudio.

Fuente: Propia.

8.3 Fase 3: Validación de los modelos a utilizar en el cálculo de la distancia de viaje

La información recabada en la Fase 2 permite tener el *input* necesario para los modelos a utilizar, para el caso presentado en el Capítulo 7 ubicado en el municipio de Tona – Santander, vemos que los modelos que presentan mejor ajuste son los de Corominas (1996), la Guía Metodológica y el Random Forest, esta conclusión se logró luego de aplicar los modelos presentados en el Capítulo 4, por lo tanto, es necesario encontrar aquellos que permitan la estimación más acertada para el área de estudio. En la Tabla 8-1 se presentan los valores del error relativo con respecto al valor real de cada uno de los 50 movimientos de la base de datos del Anexo A resumiendo los resultados del capítulo 4.

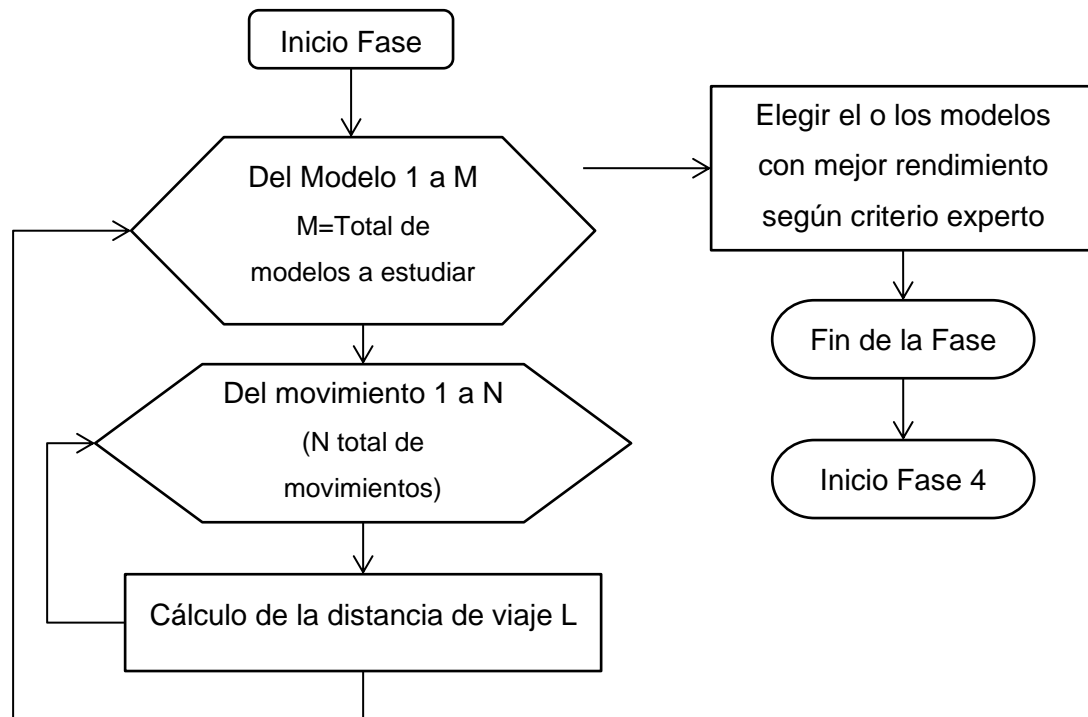
Tabla 8-1. Errores relativos del caso de estudio para cada modelo estudiado.

	Valor Calculado de la distancia de viaje L (m)	Error Relativo (%)
Random Forest	94,53	-13,91%
Sheidegger	13,06	-88,11%
Li Tianchi	12,04	-89,03%
Nicoletti & Sorriso	19,89	-81,89%
Corominas	118,7	8,11%
Guía Metodológica	127	15,66%
Roman et al.	236,6	115,48%

Es tarea del usuario de la metodología decidir qué solución utilizar dependiendo del contexto en el que se realice el estudio, a continuación, se listan algunas posibilidades para la toma de decisiones.

- Se toma el modelo que presente el menor error, en este caso sería el de Corominas con 118.7m de distancia de viaje.
- Se utiliza un promedio entre los tres o cuatro modelos que presentan menor error (Corominas, RF, y Guía Metodológica) dando como resultado 113.41m y un error relativo de 3.29%.
- Utilizar únicamente aquellos que presenten una estimación con un error relativo positivo (Corominas, Guía Metodológica y Román et al.) resultado en L=160.76m y un error relativo del 46%, esto con el objetivo de tener un factor de seguridad, en este caso sería de 1.46.

Estas son solo tres de muchas posibilidades y está en el criterio de la persona encargada el cómo analiza y usa los resultados de la Fase 3. En caso tal que ninguno de los resultados de los modelos utilizados sea satisfactorios, se pueden optar por otros adicionales que no se hayan nombrado en el presente documento o se puede cambiar la metodología utilizada. La Figura 8-3 resume el procedimiento a seguir de la Fase 3.

Figura 8-3. Fase 3. Validación del modelo.

Fuente: Propia.

8.4 Fase 4. Evaluación de la probabilidad de excedencia de la distancia de viaje

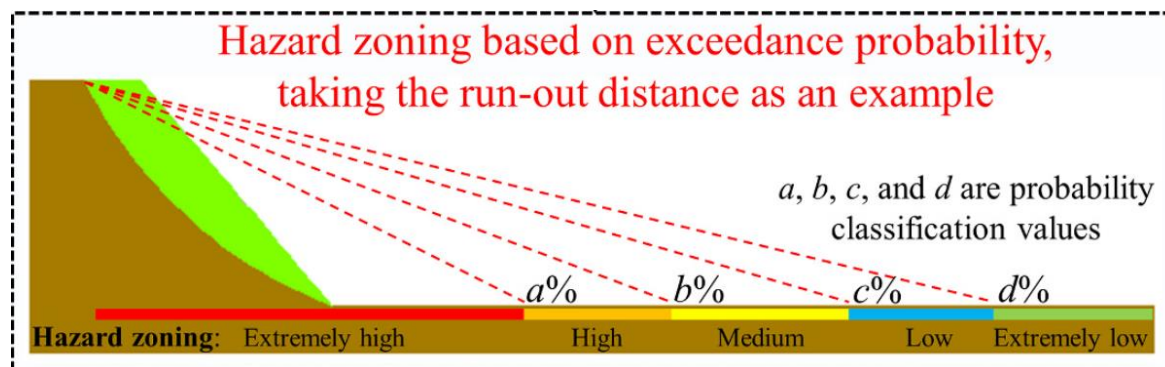
Una vez establecidos los modelos que se utilizarán como resultado de la Fase 3, se aplica el método de cálculo expuesto en el Capítulo 7, en caso de no poseer valores estimados de coeficientes de variación se pueden utilizar los expuestos en el presente documento. También se debe tener en cuenta cuáles son los modelos de estimación escogidos ya que pueden variar en los parámetros de entrada, y dependiendo de la variabilidad de dichos nuevos parámetros o de la confianza en la calidad de estos será necesario obtener sus coeficientes de variación.

El uso de la figura 7-2 puede darse en dos vías según sea el caso, el primero está asociado a la vulnerabilidad de un elemento expuesto, es decir, una obra vial, un barrio residencial o cualquier elemento que pueda ser alcanzado por el volumen desplazado, por lo tanto, el parámetro de entrada será la distancia de viaje que se medirá desde el posible origen del

material hasta el elemento expuesto, este valor entra a la gráfica en el eje de las ordenadas y se observa su probabilidad de excedencia correspondiente, esto permite establecer si el riesgo es alto, medio o bajo según criterio experto.

El segundo escenario, nuevamente haciendo uso de una curva similar a la figura 7-2, plantea la necesidad de la zonificación de la amenaza donde, estableciendo unos valores de excedencia asociados a unos niveles de riesgo o amenaza, se establece para cada uno de estos niveles una distancia de viaje permitiendo la toma de decisiones en cuanto trazados, replanteos u obras de protección para mitigar dicha amenaza similar a lo mostrado en la figura 8-4, los valores a , b , c y d dependerán de la escala de amenaza que el usuario utilice.

Figura 8-4. Zonas de amenaza de movimientos en masa basados en la probabilidad de excedencia de la distancia de viaje.



Fuente: (Sun, y otros, 2021)

En la Tabla 8-2 se presenta la categoría de la amenaza vs la probabilidad de no excedencia entendiendo este parámetro como el complemento de la probabilidad de excedencia, para el caso presentado en la Figura 7-3 en caso de que un elemento expuesto se encuentre a 100m de la fuente y utilizando el modelo de RF la Pexd es de 10%, o de no excedencia de 90% dando como resultado una Amenaza muy baja.

Tabla 8-2. Categorías de amenaza establecidas para el ejemplo de acuerdo con las probabilidades de no excedencia de la longitud de viaje.

Categoría de amenaza	Probabilidad de no excedencia
Amenaza muy alta	$\leq 2,5\%$
Amenaza alta	(2,5% - 25,0%]
Amenaza media	(25% - 50,0%]
Amenaza baja	(50% - 75,0%]
Amenaza muy baja	(75,0% - 97,5%]

Fuente: (Moncayo Legarda, 2021).

9 Conclusiones y recomendaciones

9.1 Conclusiones

Los resultados obtenidos del trabajo funcionan principalmente como un punto de partida para diversas aplicaciones en el campo de la gestión del riesgo, permitiendo la correcta toma de decisiones evaluando los riesgos por procesos de remoción en masa a los que se expone la sociedad.

Los modelos estudiados y la metodología propuesta se pueden aplicar a sistemas de alerta temprana que permitan la planificación correcta de los sistemas de prevención, evaluando bajo una escala de riesgo las zonas priorizadas. También se puede aplicar en la planificación de obras civiles haciendo uso de mapas de zonificación y estimando costos de mitigación de la amenaza, donde algunas veces será más económico replantear la ubicación en la obra en lugar de realizar una obra de protección.

La metodología también permite ir un paso más allá de los resultados que se obtienen de manera cotidiana, transformando un único valor de distancia de viaje a un rango de afectación, también permite tener una guía de como producir y analizar información obtenida para otras zonas dentro del territorio adaptándose a la información disponible y evidenciando necesidades en la toma de información dependiendo del nivel de detalle deseado.

La zona de estudio escogida facilitó el desarrollo del levantamiento de la información ya que la densidad de movimientos en masa permite con un solo levantamiento identificar varios procesos lo que reduce considerablemente los costos asociados a transporte de equipos, también otorga una gran variabilidad entre los tipos de movimientos presentados, con diferentes volúmenes de material movilizado, alturas sobre el nivel del mar, etc, que evitan el sesgo muestral.

El modelo propuesto de *Machine Learning* presenta fuertes ventajas con respecto a los utilizados de manera convencional para el cálculo de la distancia de viaje, ya que separa por completo la variable objetivo L permitiendo realizar labores de estimación sobre movimientos que aún no se han dado pero por la estabilidad del talud estudiado pueden darse. También permite ingresar variables cualitativas y cuantitativas aprovechando la totalidad de la información disponible pero, como en cualquier modelo, presenta inconvenientes a la hora de ser aplicado, como conocimientos previos de programación, una gran cantidad de datos requerida para obtener resultados satisfactorios, y un análisis previo estadístico de dicha información.

Las curvas de probabilidad de excedencia permiten establecer un rango de valores facilitando la toma de decisiones, el siguiente paso a estas curvas es realizar un análisis comparativo con distintas funciones de distribución de probabilidad y ajustar la curva tabulada a una aproximación dada por una expresión matemática, eliminando el carácter discreto de la solución presentada en el capítulo 7.

9.2 Recomendaciones

Es fundamental promover las campañas de toma de información de detalle para estudios de gestión del riesgo preventivos, una de las grandes limitantes evidenciadas es la falta de información o su calidad a escala de detalle ya que, en caso de utilizar metodologías más robustas como las planteadas en este documento como las técnicas de *Machine Learning*, o porque no, desarrollos con inteligencia artificial, requerirán bases de datos igualmente robustas que, a diferencia de otros campos del conocimiento como la hidrología o el estudio de materiales de construcción, no se encuentra fácilmente.

Para fortalecer el acceso a la información con el detalle requerido para aplicar la metodología presentada en este trabajo, se debe fomentar estudios con el mismo detalle en otras zonas de Colombia que también presenten recurrentes casos de movimientos en masa, e incluso alimentar con otro tipo de variables como geológicas o mecánicas del material movilizado y estudiar su comportamiento e influencia siguiendo el proceso realizado en el capítulo 5.

Bibliografía

- Agudelo, C. (2020). La métrica de Mahalanobis [Trabajo de grado]. Universidad Distrital Fransisco José de Caldas.
- Anish, S. (24 de 07 de 2018). *Random Forests in R*. Obtenido de <https://datascienceplus.com/random-forests-in-r/>
- Ávila, G., Cubillos, C., Granados, A., Medina, E., Rodríguez, É., Rodríguez, C., & Ruiz, G. (2016). *Guía metodológica para estudios de amenaza, vulnerabilidad y riesgo por movimientos en masa*. Bogotá D.C.: Servicio Geológico Colombiano.
- Caselles, E. (2021). *Informe técnico final, producto entregado como parte del proyecto de investigación "Método para la estimación de distancias de viaje de movimientos en masa con base en aproximaciones geométricas"*. Bogotá D.C.
- Colombiano, S. G. (s.f.). *Estado de Cartografía Geológica*. Obtenido de https://srvags.sgc.gov.co/Flexviewer/Estado_Cartografia_Geologica/?web=1&wdLOR=c59CC4008-D001-43A0-9DB8-2818A8C3D35D
- Corominas, J. (1996). The angle of reach as a mobility index for small and large landslides. *Can. Geotech. J.*, 982-994.
- Cruden, D., & Varnes, D. (1996). Landslide types and processes. *Landslides*, 36-75.
- Devore, J. (2008). *Probabilidad y Estadística para Ingeniería y Ciencias*. Mexico D.F: Cengage Learning.
- Finlay, P., & Mostyn, G. F. (1999). Landslide risk assessment: prediction of travel. *Can. Geotech. J.*, 556-562.
- Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The Elements of Statistical Learning Data Mining, Inference and Prediction*. Stanford, California: Springer.
- Hungr, O., Evans, S., Bovis, M., & Hutchinson, J. (2001). A review of the classification of landslides of the flow type. *Environmental & Engineering Geoscience*, 221-238.
- Hungr, O., Leroueil, S., & Picarelli, L. (2014). The Varnes classification of landslide types, an update. *Landslides*, 167-194.

- Hutchinson, J. (1988). General report : morphological and geotechnical parameters of landslides in relation to geology and hydrogeology. *Landslides*, (págs. 3-35). Londres.
- Instituto de Hidrología, M. y. (s.f.). *Catalogo Estaciones IDEAM*. Obtenido de <https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Catalogo-Estaciones-IDEAM/n6vw-vkfe>
- Instituto de Hidrología, M. y. (s.f.). *Indicadores Ambientales - IDEAM*. Obtenido de <https://experience.arcgis.com/experience/568ddab184334f6b81a04d2fe9aac262/page/Indicadores-Ambientales/>
- Kaya, I. (12 de 2 de 2020). *Anomaly Detection and Mahalanobis Distance for multivariate problems*. Obtenido de <https://medium.com/@ikaya754/anomaly-detection-and-mahalanobis-distance-25b21b7cfe5b>
- Marvin, N., Wagner, S., & Probst, P. (12 de 11 de 2023). *Package 'ranger'*. Obtenido de <https://cran.r-project.org/web/packages/ranger/ranger.pdf>
- Moncayo Legarda, C. (2021). Trabajo Final de Maestría: Evaluación de la distancia de viaje de movimientos en masa en Colombia a partir de registros históricos. Universidad Nacional de Colombia.
- Moncayo, S., & Ávila, G. (2022). Landslide Travel Distances in Colombia from National Landslide Database Analysis. *Progress in Landslide Research and Technology, Volume 1 Issue 1*, 315-325.
- Monsalve, G. (1995). *Hidrología en la ingeniería*. Bogotá D.C.: Departamento de Publicaciones Escuela Colombiana de Ingeniería.
- Muhammad, Q. (2015). Landslide runout distance prediction based on mechanism and cause of soil or rock mass movement. *Journal of the Civil Engineering Forum*, 29-35.
- Nicolett, P., & Sorriso, V. (1991). Geomorphic Controls of the Shape and Mobility of Rock Avalanches. *Geol. Soc. Am. Bull.*, 1365-1373.
- Raychaudhuri, S. (2008). Introduction to Monte Carlo simulation. *Winter Simulation Conference* (págs. 91-100). Miami: IEEE.
- Roman, D., Ortiz, J., Tapias, M., & Oviedo, E. (2024). Empirical Estimation of Landslide Runout Distance Using Geometrical Approximations in the Colombian North–East Andean Region. *Sustainability*.

- Scheidegger, A. (1973). On the Prediction of the Reach and Velocity of Catastrophic. *Rock Mechanics*, 231-236.
- Servicio Geológico Colombiano. (2024). *SIMMA*. Obtenido de <https://simma.sgc.gov.co/#/public/basic/>
- Sun, X., Peng, Z., Tianbin, L., Sheng, W., Rafael, J., & Xianda, F. (2021). From probabilistic back analyses to probabilistic run-out predictions of. *Engineering Geology*, 16.
- Tianchi, L. (1983). A Mathematical Model for Predicting the Extent of a major Rockfall. *Geomorphol*, 473-482.
- UNGRD. (19 de 08 de 2020). *Unidad Nacional para la Gestión del Riesgo de Desastres*. Obtenido de <https://portal.gestiondelriesgo.gov.co/Paginas/Noticias/2020/Riesgo-por-movimientos-en-masa-en-Colombia.aspx>
- Varnes, D. (1978). Slope movement types and processes. *Landslides: analysis and control*, 11-33.
- Weiwei, Z., Xuanmei, F., Runqiu, H., Xiangjun, P., Qiang, X., & Weile, L. (2017). Empirical prediction for travel distance of channelized rock. *Nat. Hazards Earth Syst. Sci.*, 833-844.
- Zongxing, Z., Chengren, X., Huiming, T., Robert, C., Aijun, S., & Xiao, L. (2017). Prediction of landslide runout based on influencing factor analysis. *Environ Earth Sci*, 17.

Anexos

A. Resumen información primaria

En este anexo se encuentra en formato de tabla la base de datos de información primaria utilizada durante el desarrollo de este trabajo final de maestría donde se registran 50 movimientos y 17 variables.

ID Movimiento	Latitud	Longitud	Longitud de la superficie de falla	Ancho de la masa desplazada	Profundidad de la superficie de falla	Área paramétrica	Volumen masa desplazada	Ángulo del talud prefalla	Dirección del movimiento
ID	Lat	Long	Lr	Wr	Dr	Ap	V	α	δ
-	-	-	m	m	m	m ²	m ³	°	°
Z11_12	7,080741	-73,185670	26,33	12,18	1,51	319,53	3,75E+02	29	0

ID Movimiento	Latitud	Longitud	Longitud de la superficie de falla	Ancho de la masa desplazada	Profundidad de la superficie de falla	Área paramétrica	Volumen masa desplazada	Ángulo del talud prefalla	Dirección del movimiento
ID	Lat	Long	Lr	Wr	Dr	Ap	V	α	δ
-	-	-	m	m	m	m ²	m ³	°	°
Z11_37	7,074047	-73,161358	64,02	23,14	3,93	456,86	4,51E+03	27	25
Z11_41_1	7,023865	-73,178198	13,38	14,50	1,00	170,00	1,50E+02	26	36
Z11_41_2	7,023829	-73,178007	28,93	23,89	1,00	249,00	5,36E+02	46	49
Z11_41_3	7,024153	-73,177752	22,83	27,48	5,17	360,78	2,51E+03	36	43
Z11_41_4	7,024092	-73,177490	20,43	17,84	4,37	289,23	1,23E+03	29	45
Z11_41_5	7,023928	-73,177131	23,76	34,52	1,85	594,88	1,18E+03	42	44
Z11_41_6	7,023537	73,177096	30,55	14,62	5,38	332,67	1,86E+03	29	0
Z11_2_1	7,129321	-73,051648	110,68	11,23	4,50	1173,73	4,33E+03	35	13
Z11_2_2	7,129480	-73,051614	117,47	32,92	1,00	3588,94	3,00E+03	37	0
Z11_2_3	7,130295	-73,050319	21,98	19,08	1,00	336,24	3,25E+02	48	0
Z11_6_1	7,145811	-73,050290	85,94	120,00	1,00	4401,15	7,99E+03	43	0
Z11_6_2	7,146008	-73,051290	49,68	45,32	1,00	1079,26	1,74E+03	44	0
Z11_44_1	7,148478	-73,050558	78,77	61,16	1,73	8012,00	6,46E+03	34	13
Z11_44_2	7,151734	-73,048086	103,57	70,19	9,37	5987,90	5,28E+04	30	5
Z11_45	7,152327	-73,044061	61,98	77,08	5,66	3150,00	2,10E+04	26	21
Z11_46_1	7,155594	-73,040591	44,00	10,36	3,12	93,00	1,10E+03	50	30
Z11_46_2	7,154804	-73,039105	69,46	42,72	3,10	1862,55	7,13E+03	45	14
Z12_1_1	7,399924	-73,194747	318,00	124,05	14,76	21696,00	4,51E+05	34	6
Z12_1_2	7,401635	-73,198890	227,47	138,32	20,44	26682,00	4,98E+05	28	25

ID Movimiento	Latitud	Longitud	Longitud de la superficie de falla	Ancho de la masa desplazada	Profundidad de la superficie de falla	Área paramétrica	Volumen masa desplazada	Ángulo del talud prefalla	Dirección del movimiento
ID	Lat	Long	Lr	Wr	Dr	Ap	V	α	δ
-	-	-	m	m	m	m ²	m ³	°	°
Z12_1_3	7,403399	-73,194076	96,50	38,91	8,87	2637,86	2,58E+04	37	91
Z12_35	7,409831	-73,201629	173,53	23,56	21,42	1487,00	6,79E+04	32	20
Z12_36	7,417656	-73,197481	72,27	15,20	2,99	891,00	2,54E+03	39	28
Z12_33	7,407272	-73,180073	36,99	32,00	3,95	1557,72	3,62E+03	37	0
Z12_14	7,411274	-73,180199	183,00	52,91	8,07	6226,67	6,05E+04	33	11
Z7_21	6,795833	-72,832500	51,26	125,80	4,39	5182,00	2,19E+04	10	29
Z7_19_1	6,797573	-72,847852	25,65	17,10	2,69	209,65	9,14E+02	39	42
Z7_19_2	6,797249	-72,847132	47,46	25,01	1,86	945,87	1,71E+03	23	8
Z7_30	6,783104	-72,836843	45,59	23,43	0,89	532,46	7,34E+02	51	0
Z6_20_1	6,620350	-72,638713	261,64	274,60	23,00	45848,27	1,28E+06	40	0
Z6_20_2	6,618799	-72,638380	115,87	32,54	2,98	3079,18	8,70E+03	44	29
Z6_20_3	6,618404	-72,638663	186,66	65,60	27,14	6284,71	2,58E+05	31	0
Z6_20_4	6,618228	-72,637618	67,97	80,72	4,07	3886,04	1,73E+04	44	17
Z6_20_5	6,619202	-72,636390	195,09	216,87	22,04	25114,23	7,23E+05	38	17
Z6_20_6	6,619271	-72,633225	261,68	181,19	11,20	17755,70	4,12E+05	33	39
Z6_20_7	6,616055	-72,634888	190,67	221,60	6,17	16826,00	2,02E+05	32	0
Z6_20_8	6,615552	-72,636901	66,39	132,20	9,06	5018,32	6,16E+04	45	0
Z5_45_1	6,625037	-72,618562	51,32	39,57	1,22	1195,75	1,92E+03	44	45
Z5_45_2	6,625475	-72,617184	58,10	28,68	1,00	924,47	1,29E+03	46	30

ID Movimiento	Latitud	Longitud	Longitud de la superficie de falla	Ancho de la masa desplazada	Profundidad de la superficie de falla	Área paramétrica	Volumen masa desplazada	Ángulo del talud prefalla	Dirección del movimiento
ID	Lat	Long	Lr	Wr	Dr	Ap	V	α	δ
-	-	-	m	m	m	m ²	m ³	°	°
Z5_45_3	6,626013	-72,623069	137,09	59,54	13,04	4585,80	8,25E+04	44	20
Z5_45_4	6,621997	-72,623014	98,08	27,03	7,68	760,68	1,58E+04	38	0
Z5_45_5	6,622421	-72,622569	54,00	42,04	3,71	1479,56	6,53E+03	44	0
Z5_45_6	6,621023	-72,625673	39,77	19,17	3,65	929,44	2,16E+03	39	0
Z5_45_7	6,624934	-72,618195	38,21	59,27	1,00	891,14	1,75E+03	54	23
Z11_36_1	6,856502	-72,883469	814,94	434,34	58,15	280409,99	1,60E+07	28	45
Z11_36_2	6,851442	-72,891627	315,83	113,08	7,23	24760,29	2,00E+05	41	0
Z11_36_3	6,856467	-72,890039	253,08	136,19	13,36	18241,61	3,57E+05	35	7
Z11_36_4	6,855959	-72,897057	266,29	190,14	15,10	34806,64	5,92E+05	39	0
Z11_36_5	6,850776	-72,891219	139,91	77,56	1,00	6806,98	8,41E+03	51	0
Z11_35	6,850776	-72,891219	27,11	15,55	4,32	289,15	1,41E+03	31	0

ID Movimiento	Elevación	Tipo de Obstáculo	Distancia al Obstáculo	Distancia de viaje medida	Altura máxima de deslizamiento	Coefficiente de fricción equivalente	Tipo de movimiento
ID	msnm	TO	DO	L	H	H/L	TM
-	m	-	m	m	m	-	-
Z11_12	923	2	66,02	66,86	34,34	0,514	2

ID Movimiento	Elevación	Tipo de Obstáculo	Distancia al Obstáculo	Distancia de viaje medida	Altura máxima de deslizamiento	Coefficiente de fricción equivalente	Tipo de movimiento
ID	msnm	TO	DO	L	H	H/L	TM
-	m	-	m	m	m	-	-
Z11_37	780	0	0,00	95,69	44,39	0,464	3
Z11_41_1	822	1	53,22	49,58	29,23	0,590	3
Z11_41_2	832	1	51,26	47,59	28,00	0,588	3
Z11_41_3	840	1	122,20	122,20	41,00	0,336	3
Z11_41_4	834	1	86,80	86,80	34,00	0,392	3
Z11_41_5	839	1	93,52	93,52	39,00	0,417	3
Z11_41_6	823	1	51,50	51,50	25,00	0,485	3
Z11_2_1	1995	2	120,18	203,20	141,92	0,698	2
Z11_2_2	1991	2	119,00	158,59	101,00	0,637	2
Z11_2_3	1914	2	21,97	21,98	25,00	1,137	3
Z11_6_1	1651	2	68,21	102,76	79,00	0,769	2
Z11_6_2	1624	2	37,47	113,33	79,50	0,701	2
Z11_44_1	1690	2	155,87	162,91	116,10	0,713	3
Z11_44_2	1641	2	167,00	168,35	89,00	0,529	3
Z11_45	1847	2	68,58	208,75	105,22	0,504	2
Z11_46_1	1869	0	0,00	126,82	97,52	0,769	3
Z11_46_2	1930	2	47,29	94,39	133,05	1,410	2
Z12_1_1	938	1	1279,95	1279,95	345,00	0,270	3
Z12_1_2	1010	1	1597,63	1597,63	416,00	0,260	3

ID Movimiento	Elevación	Tipo de Obstáculo	Distancia al Obstáculo	Distancia de viaje medida	Altura máxima de deslizamiento	Coefficiente de fricción equivalente	Tipo de movimiento
ID	msnm	TO	DO	L	H	H/L	TM
-	m	-	m	m	m	-	-
Z12_1_3	840	1	1107,74	1107,74	241,00	0,218	3
Z12_35	1100	0	0,00	168,87	92,36	0,547	3
Z12_36	934	0	0,00	233,60	138,57	0,593	3
Z12_33	770	0	0,00	47,76	40,18	0,841	3
Z12_14	782	0	0,00	576,00	173,00	0,300	3
Z7_21	2173	2	50,47	310,00	115,00	0,371	1
Z7_19_1	1557	1	118,00	118,00	53,00	0,449	3
Z7_19_2	1567	1	119,00	119,00	67,00	0,563	3
Z7_30	1555	1	120,00	120,00	53,68	0,447	2
Z6_20_1	1876	1	301,70	301,70	194,63	0,645	3
Z6_20_2	1780	1	115,30	115,30	95,94	0,832	3
Z6_20_3	1775	1	160,76	160,76	94,85	0,590	3
Z6_20_4	1741	1	48,88	48,88	47,24	0,966	2
Z6_20_5	1881	1	223,85	223,85	174,24	0,778	1
Z6_20_6	1876	1	219,31	219,31	143,21	0,653	1
Z6_20_7	1820	1	161,49	161,49	101,34	0,628	2
Z6_20_8	1785	1	47,03	47,03	46,85	0,996	2
Z5_45_1	2027	1	162,15	162,15	148,95	0,919	2
Z5_45_2	2016	1	151,82	151,82	129,69	0,854	2

ID Movimiento	Elevación	Tipo de Obstáculo	Distancia al Obstáculo	Distancia de viaje medida	Altura máxima de deslizamiento	Coefficiente de fricción equivalente	Tipo de movimiento
ID	msnm	TO	DO	L	H	H/L	TM
-	m	-	m	m	m	-	-
Z5_45_3	2043	1	356,93	356,93	180,24	0,505	3
Z5_45_4	1904	1	98,08	76,99	61,06	0,793	2
Z5_45_5	1887	1	54,00	38,73	37,67	0,973	1
Z5_45_6	1855	1	44,45	44,45	31,68	0,713	3
Z5_45_7	2017	1	161,63	161,63	133,35	0,825	2
Z11_36_1	2781	1	814,94	7670,00	1403,00	0,183	3
Z11_36_2	2574	1	315,83	267,26	241,04	0,902	3
Z11_36_3	2497	1	253,08	267,21	179,67	0,672	3
Z11_36_4	2350	1	266,29	246,60	214,06	0,868	3
Z11_36_5	2562	1	139,91	110,60	112,47	1,017	3
Z11_35	886	0	0	75,80	39,27	0,518	1

B. Código en R utilizado para realizar el análisis descriptivo

En este Anexo se encuentra el código utilizado para la exploración de la base de datos establecida en el Anexo A, el lenguaje de programación utilizado es R y las librerías utilizadas pertenecen al paquete estándar del software Rstudio.

```
#Librerías utilizadas.
library(readxl)
library(writexl)
library(dplyr)

#Ingreso y arreglo de la base de datos para su uso.
data <- read_excel("C:/Users/USUARIO/Desktop/Maestría/Trabajo de
Grado/Muestreos/Compilación Movimientos/00_DatBase29092023.xlsx",
                  sheet = "Base de datos depurada", range
= "B5:R55")
data <- data[-c(1), ]

data$Lat <- as.numeric(data$Lat)
data$Long <- as.numeric(data$Long)
data$Lr <- as.numeric(data$Lr)
data$Wr <- as.numeric(data$Wr)
data$Dr <- as.numeric(data$Dr)
data$Ap <- as.numeric(data$Ap)
data$V <- as.numeric(data$V)
data$α <- as.numeric(data$α)
data$δ <- as.numeric(data$δ)
data$msnm <- as.numeric(data$msnm)
data$DO <- as.numeric(data$DO)
data$L <- as.numeric(data$L)
data$H <- as.numeric(data$H)
data$HL <- as.numeric(data$HL)

#Resumen estadístico de la base de datos.
data1 <- data[,c("Lr", "Wr", "Dr", "Ap", "V", "α", "δ", "msnm", "DO", "L", "H", "HL")]
resumen <- summary(data1)

#Resumen estadístico de la base de datos con transformación Log+1.
```

```

log.data1 <- log10(data[,c("Lr", "Wr", "Dr", "Ap", "V", "α", "δ", "msnm", "DO", "L", "H", "HL")]
+1)
summary(log.data1)

#Matriz de Covarianzas.
cov <- round(cov(data1), 3)

#Matriz de correlaciones.
corr <- round(100*cor(data1), 3)

#Cálculo de los coeficientes de correlación
vcvlog.data <- 1:ncol(log.data1)
vsdlog.data <- 1:ncol(log.data1)
i <- 1
for( i in 1:ncol(data1))
{
  vsdlog.data[i] <- sd(log.data1[,c(i)])
  vcvlog.data[i] <- sd(log.data1[,c(i)])/mean(log.data1[,c(i)])
}
vcvlog.data
vsdlog.data
colMeans(log.data1)
z <- rbind.data.frame(vcvlog.data, vsdlog.data, colMeans(log.data1))

#Estandarización de las variables igualando sus desviaciones estandar a 0
#para mejorar la visualización en un grafico de cajas y bigotes

log.data1sc <- log.data1 %>%
mutate_at(c('Lr', 'Wr', 'Dr', 'Ap', 'V', 'α', 'δ', 'msnm', 'DO', 'L', 'H', 'HL'),
~(scale(.) %>% as.vector))
windows()

windows()
par(mar = c(5,7,4,2) + 0.1)
boxplot(log.data1sc, horizontal = TRUE, border = "blue", col = "lightblue",
las = 1, cex.axis = 1, boxwex = 0.5, xlab = "puntaje", ylab =
"variable", outline = FALSE)

#Visualización de las correlaciones entre variables y su distribución a
traves de un histograma

windows()
pairs(log.data1)

windows()
pairs(data1, lwd = 1, pch = 16, cex = 0.5, gap = 0, xaxt = "n", yaxt =
"n",
panel = function(x,y, ...){
  points(x, y, col = "red", ...)
  abline(lm(y ~ x), col = "blue", lwd = 2)
},

```

```

diag.panel = function(x, ...) {
  usr <- par("usr"); on.exit(par(usr))
  par(usr = c(usr[1:2], 0, 1.5))
  h <- hist(x, plot = FALSE)
  breaks <- h$breaks; nB <- length(breaks)
  y <- h$counts; y <- y/max(y)
  rect(breaks [-nB], 0, breaks[-1], y, col = "gray", ...)
}
)

#Análisis con la distancia de Mahalanobis

#USO DE LA RUTINA mahalanobis USANDO LA PARTE CUANTITATIVA DE LA BASE DE
DATOS (12 VARIABLES 49 INDIVIDUOS)
n <- nrow(data1)
p <- ncol(data1)
d <- apply(data1, MARGIN = 1,
           FUN = function(x) mahalanobis(x , center = colMeans(data1), cov
= cov(data1)))
#ORDEN DEL VECTOR d DE MANERA ASCENDENTE
sd <- sort(d)
#FUNCIÓN CUANTIL PARA UNA DISTRIBUCIÓN CHI-CUADRADO
qc <- qchisq(p = (1:n - 0.5)/n, df = p)

outs.idx <- which(sd > qchisq(p = 0.997, df = p))

col.out <- rep("green",n)
col.out[outs.idx] <- "red"

windows()
plot(x = qc, y=sd, col = col.out,
     xlab = expression(paste("Distancia de Mahalanobis para una función ",
chi[p]^2)) , ylab = "Distancia de Mahalanobis Calculada")
grid()
abline(a = 0, b = 1, col = "black", lty = 2)

#DISTANCIA DE MAHALANOBIS log+1
#USO DE LA RUTINA mahalanobis USANDO LA PARTE CUANTITATIVA DE LA BASE DE
DATOS (12 VARIABLES 49 INDIVIDUOS)
n <- nrow(log.data1)
p <- ncol(log.data1)
d <- apply(log.data1, MARGIN = 1,
           FUN = function(x) mahalanobis(x , center = colMeans(log.data1),
cov = cov(log.data1)))
#ORDEN DEL VECTOR d DE MANERA ASCENDENTE
sd <- sort(d)
#FUNCIÓN CUANTIL PARA UNA DISTRIBUCIÓN CHI-CUADRADO DE
qc <- qchisq(p = (1:n - 0.5)/n, df = p)

outs.idx <- which(sd > qchisq(p = 0.997, df = p))

col.out <- rep("green",n)

```

```
col.out[outs.idx] <- "red"

windows()
plot(x = qc, y=sd, col = col.out,
      xlab = expression(paste("Distancia de Mahalanobis para una función ",
chi[p]^2)) , ylab = "Distancia de Mahalanobis Calculada")
grid()
abline(a = 0, b = 1, col = "black", lty = 2)

#Análisis de componenetes principales con la base de datos transformada
(Log+1)

data2 <- data[,c("Lr", "Wr", "Dr", "Ap", "V", "α", "δ", "msnm", "DO")]
data2 <- log10(data2+1)
pc <- princomp(data2, cor = TRUE, scores = TRUE)
summary(pc)
pc$loadings

windows()
screepplot(pc, col="blue", pch = 16, type = "lines", cex = 2, lwd = 2,
           cex.axis = 0.8, cex.lab = 0.8, main = " ")

head(pc$scores)

#Grafico para encontrar las variables principales.
windows()
biplot(pc, col = c(2,3),
       xlab = "primera componente",
       ylab = "segunda componente",
       main = "Biplot de variables geométricas")
grid()
```

C. Código en R para modelo de Random Forest y simulaciones

En este Anexo se encuentra el código utilizado para la generación del modelo Random Forest utilizando la información del Anexo A, el lenguaje de programación utilizado es R y las librerías utilizadas pertenecen al paquete estándar del software Rstudio, exceptuando el paquete ranger, el cual permite generar una rutina para establecer modelos de Machine Learning y las simulaciones de Montecarlo para el cálculo de la probabilidad de excedencia.

```
library(readxl)
library(writexl)
library(ranger)
library(ggplot2)

#Ingreso base de datos
data <- read_excel("C:/Users/USUARIO/Desktop/Maestría/Trabajo de
Grado/Muestreos/Compilación Movimientos/00_DatBase07042024.xlsx",
                 sheet = "Base de datos depurada", range = "B5:R56")

data <- data[-c(1), ]

data$Lat <- as.numeric(data$Lat)
data$Long <- as.numeric(data$Long)
data$Lr <- as.numeric(data$Lr)
data$Wr <- as.numeric(data$Wr)
data$Dr <- as.numeric(data$Dr)
data$Ap <- as.numeric(data$Ap)
data$V <- as.numeric(data$V)
data$α <- as.numeric(data$α)
data$δ <- as.numeric(data$δ)
data$msnm <- as.numeric(data$msnm)
data$DO <- as.numeric(data$DO)
data$L <- as.numeric(data$L)
data$H <- as.numeric(data$H)
data$HL <- as.numeric(data$HL)

data <- data[ , -c(1,2,3,15,16)]
data3 <- data
data1 <- log10(data[ , -c(9,12)]+1)
df <- data.frame(data1, data[ , c(9,12)])
df <- df[ , c("Lr", "Wr", "Dr", "Ap", "V", "α", "δ", "msnm", "TO", "DO", "L", "TM")]
data <- df

#Grilla de hiperparámetros.
```

```

hyper_grid <- expand.grid(
  n           = seq(1, nrow(data)),
  mtry        = seq(1, 11, by= 1),
  node_size   = seq(1, 10, by = 1),
  sample_size = c(.55, .632, .70, .80, 1),
  OOB_RMSE    = 0,
  Test_error  = 0
)

#Encontrar el modelo óptimo.

for(i in 1:nrow(hyper_grid))
{
  n = hyper_grid[i,1]
  data1 <- data[-n,]
  model <- ranger(
    formula    = L ~. ,
    data       = data1,
    num.trees  = 500,
    mtry       = hyper_grid$mtry[i],
    min.node.size = hyper_grid$node_size[i],
    sample.fraction = hyper_grid$sample_size[i],
    seed       = 100 # Notese el seteo de la semilla
  )

  hyper_grid$OOB_RMSE[i] <- sqrt(model$prediction.error)
  pred=predict(model,data=data[n,])
  hyper_grid$Test_error[i] <- (pred$predictions-data[n,11])/data[n,11]
}

#función de filtrado

hyper_grid2 <- expand.grid(
  mtry        = seq(1, 11, by= 1),
  node_size   = seq(1, 10, by = 1),
  sample_size = c(.55, .632, .70, .80, 1),
  OOB_RMSE    = 0,
  Test_error  = 0
)

hyper_grid$Test_error <- as.numeric(hyper_grid$Test_error)
for (i in 1:nrow(hyper_grid2)) {
  x <- subset(hyper_grid, mtry==hyper_grid2[i,1] &
node_size==hyper_grid2[i,2] & sample_size==hyper_grid2[i,3])
  hyper_grid2[i,4]<-mean(x$OOB_RMSE)
  hyper_grid2[i,5]<-mean(x$Test_error)
}

#Variación del error con respecto a mtry
windows()

```

```
boxplot(formula = OOB_RMSE ~ mtry, data = hyper_grid2, xlab="Cantidad
de variables por nodo", ylab="Out Of Bag Error")

#Variación del error con respecto a cantidad de nodos
windows()
boxplot(formula = OOB_RMSE ~ node_size, data = hyper_grid2,
xlab="Cantidad de nodos", ylab="Out Of Bag Error")

#Variación del error con respecto a tamaño de la muestra de
entrenamiento
windows()
boxplot(formula = OOB_RMSE ~ sample_size, data = hyper_grid2,
xlab="Proporción de la muestra usada para entrenamiento", ylab="Out Of
Bag Error")

#Variación del error con respecto a mtry

boxplot(formula = Test_error ~ mtry, data = hyper_grid2)

#Variación del error con respecto a cantidad de nodos

boxplot(formula = Test_error ~ node_size, data = hyper_grid2)

#Variación del error con respecto a tamaño de la muestra de
entrenamiento

boxplot(formula = Test_error ~ sample_size, data = hyper_grid2)

#validación del modelo con mtry=11, node_size=1, sample_size=1.

a <- numeric(nrow(data))

for(i in 1:nrow(data))
{
  data1 <- data[-i,]
  model <- ranger(
    formula      = L ~. ,
    data         = data1,
    num.trees    = 500,
    mtry         = 11,
    min.node.size = 2,
    sample.fraction = 0.8,
    seed         = 100,
    write.forest = TRUE
  )
  pred=predict(model,data=data[i,])
  a[i] <- pred$predictions
}

a <- 10^a
V.calculado = a
V.real = 10^data1$L
```

```

Error.Relativo = (V.calculado-V.real)/V.calculado*100
windows()
boxplot(Error.Relativo, outline = FALSE, ylab="Error relativo")
boxplot(Error.Relativo)

#Simulaciones de montecarlo
seed=100
hyper_grid3 <- expand.grid(
  Lr= seq(1, 10000, by=1),
  Wr=0,
  Dr=0,
  Ap=0,
  V=0,
  alpha=0,
  delta=0,
  msnm=0,
  TO=0,
  DO=0,
  L=0,
  TM=0
)

for(i in 1:nrow(hyper_grid3))
{
  seed=100
  hyper_grid3[i,1]=log10(rnorm(1, mean=92.52, sd=92.52*0.0753)+1)
  hyper_grid3[i,2]=log10(rnorm(1, mean=40.73, sd=40.73*0.109)+1)
  hyper_grid3[i,3]=log10(rnorm(1, mean=4.27, sd=4.27*0.198)+1)

  hyper_grid3[i,4]=log10((pi*(10^hyper_grid3[i,1])/2*(10^hyper_grid3[i,2])
/2))

  hyper_grid3[i,5]=log10(1/6*pi*10^hyper_grid3[i,1]*10^hyper_grid3[i,2]*10
^hyper_grid3[i,3]*rnorm(1, mean=1.48, sd=0.15))
  hyper_grid3[i,6]=log10(40)
  hyper_grid3[i,7]=log10(31)
  hyper_grid3[i,8]=log10(2090)
  hyper_grid3[i,9]=2
  hyper_grid3[i,10]=log10(31.01)
  hyper_grid3[i,11]=0
  hyper_grid3[i,12]=3
}

for(i in 1:nrow(hyper_grid3))
{
  seed=100
  pred=predict(model,hyper_grid3[i,])
  hyper_grid3[i,11]=pred$predictions
}

```

```
#Simulaciones para generar la probabilidad de excedencia

#pasar de log10 a normal los valores calculados
L<-10^hyper_grid3$L

#organizar de mayor a menor
L<-sort(L,decreasing = TRUE)

#columna del orden del valor calculado
Orden=seq(1, 10000, by=1)

#Vector de prob de excedencia en ceros
Pexd=numeric(nrow(hyper_grid3))

#Ensamblaje de la matriz
MC=data.frame(Orden,L,Pexd)

#Vector de prob de excedencia dependiendo del orden
for(i in 1:nrow(MC))
{
  MC[i,3]=MC[i,1]/(MC[(nrow(MC)),1]+1)
}

#Gráfica probabilidad de excedencia solo random forest
windows()
plot(x=MC$Pexd, y=MC$L, type="l", lwd=1, col = "blue", lty=1,
xlab="Probabilidad de excedencia", ylab="Distancia de viaje L (m)")
  legend (x="topright",
         c("Pexd", "L Real"), # Textos de la leyenda
         lty = c(1, 2),      # Tipos de línea
         col = c("blue","black"), # Colores
         lwd = 2)
abline(h=109.08, lty=2)

#Gráfica probabilidad de excedencia comparativa
#sheidegger
seed=100
hyper_grid4 <- expand.grid(
  Lr= seq(1, 10000, by=1),
  Wr=0,
  Dr=0,
  V=0,
  H=0,
  L=0
)
for(i in 1:nrow(hyper_grid4))
{
  seed=100
  hyper_grid4[i,1]=rnorm(1, mean=92.52, sd=92.52*0.0753)+1
  hyper_grid4[i,2]=rnorm(1, mean=40.73, sd=40.73*0.109)+1
  hyper_grid4[i,3]=rnorm(1, mean=4.27, sd=4.27*0.198)+1
}
```

```

hyper_grid4[i,4]=1/6*pi*hyper_grid4[i,1]*hyper_grid4[i,2]*hyper_grid4[i,
3]*rnorm(1, mean=1.48, sd=0.15)
  hyper_grid4[i,5]=84.39
  L=0
}

for(i in 1:nrow(hyper_grid4))
{
  seed=100

pred=hyper_grid4[i,5]/(10^(0.624+0.15666*log10(hyper_grid4[i,4]/1000)))
  hyper_grid4[i,6]=pred
}

SH<-sort(hyper_grid4$L,decreasing = TRUE)
MC=data.frame(MC,SH)
#Li Tianchi
seed=100
hyper_grid4 <- expand.grid(
  Lr= seq(1, 10000, by=1),
  Wr=0,
  Dr=0,
  V=0,
  H=0,
  L=0
)
for(i in 1:nrow(hyper_grid4))
{
  seed=100
  hyper_grid4[i,1]=rnorm(1, mean=92.52, sd=92.52*0.0753)+1
  hyper_grid4[i,2]=rnorm(1, mean=40.73, sd=40.73*0.109)+1
  hyper_grid4[i,3]=rnorm(1, mean=4.27, sd=4.27*0.198)+1

hyper_grid4[i,4]=1/6*pi*hyper_grid4[i,1]*hyper_grid4[i,2]*hyper_grid4[i,
3]*rnorm(1, mean=1.48, sd=0.15)
  hyper_grid4[i,5]=84.39
  L=0
}

for(i in 1:nrow(hyper_grid4))
{
  seed=100
  pred=hyper_grid4[i,5]/(10^(0.664+0.1529*log10(hyper_grid4[i,4]/1000)))
  hyper_grid4[i,6]=pred
}
LT<-sort(hyper_grid4$L,decreasing = TRUE)
MC=data.frame(MC,LT)
#Nicoletti & Sorriso
seed=100
hyper_grid4 <- expand.grid(
  Lr= seq(1, 10000, by=1),

```

```

    Wr=0,
    Dr=0,
    V=0,
    H=0,
    L=0
  )
for(i in 1:nrow(hyper_grid4))
{
  seed=100
  hyper_grid4[i,1]=rnorm(1, mean=92.52, sd=92.52*0.0753)+1
  hyper_grid4[i,2]=rnorm(1, mean=40.73, sd=40.73*0.109)+1
  hyper_grid4[i,3]=rnorm(1, mean=4.27, sd=4.27*0.198)+1

hyper_grid4[i,4]=1/6*pi*hyper_grid4[i,1]*hyper_grid4[i,2]*hyper_grid4[i,
3]*rnorm(1, mean=1.48, sd=0.15)
  hyper_grid4[i,5]=84.39
  L=0
}

for(i in 1:nrow(hyper_grid4))
{
  seed=100
  pred=hyper_grid4[i,5]/(10^(0.527+0.0847*log10(hyper_grid4[i,4]/1000)))
  hyper_grid4[i,6]=pred
}
NS<-sort(hyper_grid4$L,decreasing = TRUE)
MC=data.frame(MC,NS)
#Corominas
seed=100
hyper_grid4 <- expand.grid(
  Lr= seq(1, 10000, by=1),
  Wr=0,
  Dr=0,
  V=0,
  H=0,
  L=0
)
for(i in 1:nrow(hyper_grid4))
{
  seed=100
  hyper_grid4[i,1]=rnorm(1, mean=92.52, sd=92.52*0.0753)+1
  hyper_grid4[i,2]=rnorm(1, mean=40.73, sd=40.73*0.109)+1
  hyper_grid4[i,3]=rnorm(1, mean=4.27, sd=4.27*0.198)+1

hyper_grid4[i,4]=1/6*pi*hyper_grid4[i,1]*hyper_grid4[i,2]*hyper_grid4[i,
3]*rnorm(1, mean=1.48, sd=0.15)
  hyper_grid4[i,5]=84.39
  L=0
}

for(i in 1:nrow(hyper_grid4))
{

```

```

    seed=100
    pred=hyper_grid4[i,5]/(10^(-0.047-0.085*log10(hyper_grid4[i,4]/1000)))
    hyper_grid4[i,6]=pred
  }

C<-sort(hyper_grid4$L,decreasing = TRUE)
MC=data.frame(MC,C)
#GUIA
seed=100
hyper_grid4 <- expand.grid(
  Lr= seq(1, 10000, by=1),
  Wr=0,
  Dr=0,
  V=0,
  H=0,
  L=0
)
for(i in 1:nrow(hyper_grid4))
{
  seed=100
  hyper_grid4[i,1]=rnorm(1, mean=92.52, sd=92.52*0.0753)+1
  hyper_grid4[i,2]=rnorm(1, mean=40.73, sd=40.73*0.109)+1
  hyper_grid4[i,3]=rnorm(1, mean=4.27, sd=4.27*0.198)+1

hyper_grid4[i,4]=1/6*pi*hyper_grid4[i,1]*hyper_grid4[i,2]*hyper_grid4[i,
3]*rnorm(1, mean=1.48, sd=0.15)
  hyper_grid4[i,5]=84.39
  L=0
}

for(i in 1:nrow(hyper_grid4))
{
  seed=100
  pred=hyper_grid4[i,5]/(10^(-0.049-0.108*log10(hyper_grid4[i,4]/1000)))
  hyper_grid4[i,6]=pred
}

G<-sort(hyper_grid4$L,decreasing = TRUE)
MC=data.frame(MC,G)
#ART1
seed=100
hyper_grid4 <- expand.grid(
  Lr= seq(1, 10000, by=1),
  Wr=0,
  Dr=0,
  V=0,
  L=0
)
for(i in 1:nrow(hyper_grid4))
{
  seed=100
  hyper_grid4[i,1]=rnorm(1, mean=92.52, sd=92.52*0.0753)+1

```

```

hyper_grid4[i,2]=rnorm(1, mean=40.73, sd=40.73*0.109)+1
hyper_grid4[i,3]=rnorm(1, mean=4.27, sd=4.27*0.198)+1

hyper_grid4[i,4]=1/6*pi*hyper_grid4[i,1]*hyper_grid4[i,2]*hyper_grid4[i,
3]*rnorm(1, mean=1.48, sd=0.15)
}

for(i in 1:nrow(hyper_grid4))
{
  seed=100
  pred=228.68+0.0005*hyper_grid4[i,4]
  hyper_grid4[i,5]=pred
}

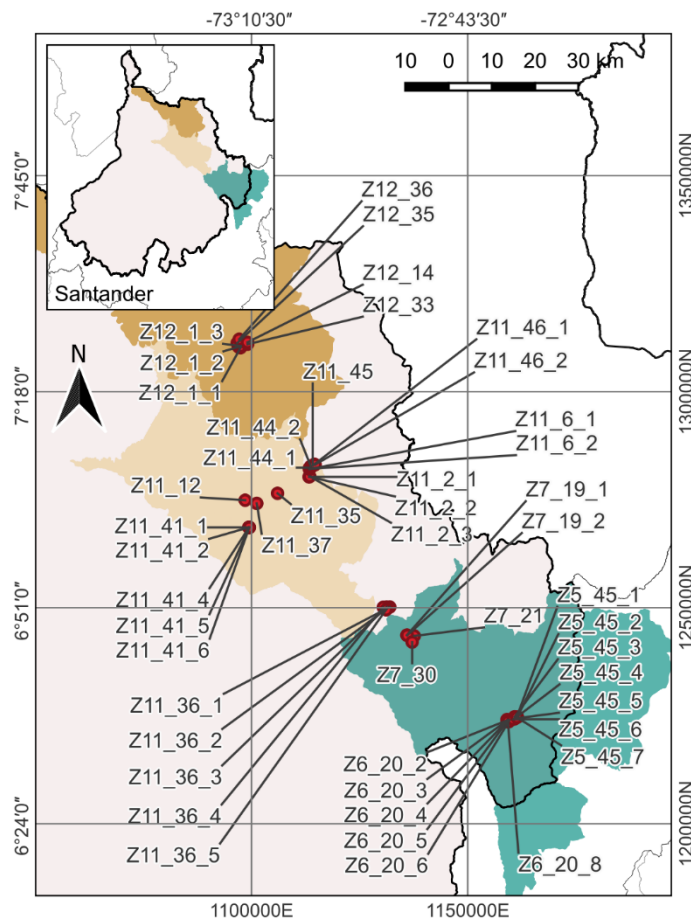
A1<-sort(hyper_grid4$L,decreasing = TRUE)
MC=data.frame(MC,A1)

windows(width = 9, height = 9)
par(mar = c(6, 5, 4, 12))
plot(MC$Pexd, MC$L, type="l", lwd=1, col = "blue", lty=1, ylim =
c(min(L, SH, LT, NS, C, G, A1), max(L, SH, LT, NS, C, G, A1)),
      xlab = "Probabilidad de excedencia (%)", ylab = "Distancia de viaje
L (m)")
lines(MC$Pexd, MC$SH, type="l", lwd=1, col = "red", lty=1, ylab="",
      xlab="")
lines(MC$Pexd, MC$LT, type="l", lwd=1, col = "orange", lty=1, ylab="",
      xlab="")
lines(MC$Pexd, MC$NS, type="l", lwd=1, col = "black", lty=1, ylab="",
      xlab="")
lines(MC$Pexd, MC$C, type="l", lwd=1, col = "gray", lty=1, ylab="",
      xlab="")
lines(MC$Pexd, MC$G, type="l", lwd=1, col = "cyan", lty=1, ylab="",
      xlab="")
lines(MC$Pexd, MC$A1, type="l", lwd=1, col = "magenta", lty=1, ylab="",
      xlab="")
abline(h=109.08, lty=2)
legend (x="topright",
        inset = c(-0.41,0.4),
        c("Random Forest", "Scheidegger", "Li Tianchi", "Nicoletti &
Sorriso", "Corominas", "Guia Metodológica", "Daniel et al.", "L
Medido"), # Textos de la leyenda
        lty = c(1, 1, 1, 1, 1, 1, 1, 1, 2), # Tipos de línea
        col =
c("blue","red","orange","black","gray","cyan","magenta","black"),
# Colores
        lwd = 2,
        xpd=TRUE)

```


D. Recopilación de información primaria – Fichas técnicas

Este anexo busca resumir todo el trabajo realizado en el procesamiento de los datos topográficos y fotogramétricos realizados en campo en la zona de estudio y que permiten tener la información base presentada en el Anexo A.



El complemento de este anexo se encuentra en el documento con nombre Anexo D: Fichas Técnicas.