UNIVERSIDAD NACIONAL AGRARIA DE LA SELVA

FACULTAD DE RECURSOS NATURALES RENOVABLES

ESCUELA PROFESIONAL DE INGENIERÍA AMBIENTAL

FACTORES FÍSICOS PARA UN MODELO ESTADÍSTICO DE SUSCEPTIBILIDAD A DESLIZAMIENTOS DE MASAS EN EL DISTRITO DE SILLAPATA EL AÑO 2021

Tesis

Para optar el título de:

INGENIERO AMBIENTAL

PRESENTADO POR:

AHNEL KAREN, ZELAYA MOYA

Tingo María – Perú

2022

UNIVERSIDAD NACIONAL AGRARIA DE LA SELVA

Tingo María – Perú

FACULTAD DE RECURSOS NATURALES RENOVABLES

ACTA DE SUSTENTACIÓN DE TESIS

Los que suscriben, miembros del Jurado de Tesis, reunidos con fecha 10 de Diciembre del 2021 a hora 8:20 a. m. a través de la Sala Virtual de Conferencias Microsoft Teams de la Escuela Profesional de Ingeniería Ambiental de la Facultad de Recursos Naturales Renovables para calificar la Tesis titulada:

"FACTORES FÍSICOS PARA UN MODELO ESTADÍSTICO DE SUSCEPTIBILIDAD A DESLIZAMIENTOS DE MASAS EN EL DISTRITO DE SILLAPATA EL AÑO 2021"

Presentado por la Bachiller: **ZELAYA MOYA, Ahnel Karen** después de haber escuchado la sustentación y las respuestas a las interrogantes formuladas por el Jurado, se declara **APROBADO** con el calificativo de **"EXCELENTE"**

En consecuencia, el sustentante queda apto para optar el Título de **INGENIERO AMBIENTAL**, que será aprobado por el Consejo de Facultad, tramitándolo al Consejo Universitario para el otorgamiento del Título correspondiente.

Tingo María, 22 de Marzo de 2022

Dr. LUIS EDUARDO ORÈ CIERTO
PRESIDENTE DEL JURADO

Ing. MSc. VÍCTOR M. BETETA ALVARADO
MIEMBRO DEL JURADO

Ing. MSc. ERLE O. BUSTAMANTE SCAGLIONI
MIEMBRO DEL JURADO

Ing. MSc. FRANKLIN DIONISIO MONTALVO
ASESOR DE LA TESIS

UNIVERSIDAD NACIONAL AGRARIA DE LA SELVA

FACULTAD DE RECURSOS NATURALES RENOVABLES

ESCUELA PROFESIONAL DE INGENIERÍA AMBIENTAL

FACTORES FÍSICOS PARA UN MODELO ESTADÍSTICO DE SUSCEPTIBILIDAD A DESLIZAMIENTOS DE MASAS EN EL DISTRITO DE SILLAPATA EL AÑO 2021

Autora : Ahnel Karen Zelaya Moya

Asesor : Ing. MSc. Franklin Dionisio Montalvo

Programa de investigación : Gestión ambiental

Línea(s) de investigación : Sistemas de gestión

Eje temático de investigación : Gestión de riesgos naturales y ambientales

Lugar de ejecución : Centro poblado Sillapata, región Huánuco

Duración :

Fecha de inicio: Noviembre, 2020.

Término : Julio, 2021

Financiamiento : MONTO S/. 2’860.00

FEDU : ---

Propio : S/. 2’860.00

Otros : ---

Tingo María – Perú

2022

UNIVERSIDAD NACIONAL AGRARIA DE LA SELVA

Tingo María – Perú

FACULTAD DE RECURSOS NATURALES RENOVABLES

ACTA DE SUSTENTACIÓN DE TESIS

Los que suscriben, miembros del Jurado de Tesis, reunidos con fecha 10 de Diciembre del 2021 a hora 8:20 a. m. a través de la Sala Virtual de Conferencias Microsoft Teams de la Escuela Profesional de Ingeniería Ambiental de la Facultad de Recursos Naturales Renovables para calificar la Tesis titulada:

"FACTORES FÍSICOS PARA UN MODELO ESTADÍSTICO DE SUSCEPTIBILIDAD A DESLIZAMIENTOS DE MASAS EN EL DISTRITO DE SILLAPATA EL AÑO 2021"

Presentado por la Bachiller: **ZELAYA MOYA, Ahnel Karen** después de haber escuchado la sustentación y las respuestas a las interrogantes formuladas por el Jurado, se declara **APROBADO** con el calificativo de **"EXCELENTE"**

En consecuencia, el sustentante queda apto para optar el Título de **INGENIERO AMBIENTAL**, que será aprobado por el Consejo de Facultad, tramitándolo al Consejo Universitario para el otorgamiento del Título correspondiente.

Tingo María, 22 de Marzo de 2022

Dr. LUIS EDUARDO ORÈ CIERTO
PRESIDENTE DEL JURADO

Ing. MSc. VÍCTOR M. BETETA ALVARADO
MIEMBRO DEL JURADO

Ing. MSc. ERLE O. BUSTAMANTE SCAGLIONI
MIEMBRO DEL JURADO

Ing. MSc. FRANKLIN DIONISIO MONTALVO
ASESOR DE LA TESIS

DEDICATORIA

El estudio de las amenazas naturales son como crónicas de una muerte anunciada por la naturaleza, que estoy aprendiendo a leer.

Este pequeño esfuerzo está dedicado a mi familia, papás, hermanos y sobrinos, su amor y su ejemplo, me dieron la motivación en este accidentado camino; especialmente a Zadi y Aurelia, mi hogar, mis buenos compañeros, maestros de vida y padres.

También me lo dedico a mí misma, porque vendrán caminos más accidentados, crónicas más difíciles de leer, pero este me ha enseñado que con estrategia y firmeza se puede.

AGRADECIMIENTO

Agradezco a mis padres, quienes forjaron mi vida, su paciencia y compañerismo en los días difíciles siguen siendo el mejor regalo. A mis hermanas y hermano por estar a mi lado a pesar de la distancia.

A mi alma mater, la Universidad Nacional Agraria de la Selva, por sus enseñanzas y cobijo durante los años de mi formación universitaria; al ing. Franklin Dionisio por su asesoría, demás docentes, amigos y amigas, por todo su apoyo.

Agradecimiento especial al ing. Noé Zelaya, mi hermano, por hacer las veces de asesor y coach, su apoyo incondicional durante las partes más difíciles del desarrollo de resultados del presente trabajo fueron claves.

ÍNDICE

Página

I. INTRODUCCIÓN 1

1.1. Objetivo general 2

1.2. Objetivos específicos 2

II. REVISIÓN DE LITERATURA 3

2.1. Estado del arte 3

2.2. Marco teórico 4

2.2.1. Gestión del riesgo de desastres 4

2.2.2. Peligro o amenaza 4

2.2.3. Deslizamientos de masa 5

2.2.4. Susceptibilidad a deslizamiento en masa 8

2.2.5. Modelos lineales generalizados 14

2.2.6. Criterio de información de Akaike 16

2.2.7. Validación del modelo 17

III. MATERIALES Y MÉTODOS 20

3.1. Lugar de ejecución 20

3.1.1. Ubicación de la zona de estudio 20

3.1.1. Accesibilidad 20

3.1.2. Condiciones ambientales 22

3.1.3. Características socioeconómicas 22

3.2. Materiales y equipos 23

3.2.1. Materiales 23

3.2.2. Equipos 23

3.3. Metodología 23

3.3.1. Tipo de investigación 23

3.3.2. Flujograma general del proceso 24

3.3.3. Recopilación de información 24

3.3.4. Inventario de puntos 25

3.3.5. Mapeo de factores físicos del terreno 25

3.3.6. Extracción de información a puntos filtrados 28

3.3.7. Exploración de los factores físicos 29

3.3.8. Aplicación del modelo lineal generalizado 30
3.3.9. Ajuste y simplificación del modelo 30
3.3.10. Análisis de la significancia de las variables finales 30
3.3.11. Predicción del mapa de susceptibilidad 30
3.3.12. Proceso de validación del modelo 31
IV. RESULTADOS Y DISCUSIÓN 32
4.1. Exploración de los factores físicos 32
4.2. Desarrollo del modelo de susceptibilidad a deslizamientos de masa 34
4.2.1. Selección de variables y modelo 34
4.2.2. Validación del modelo 37
4.3. Determinación de la relación entre factores físicos y el modelo 38
V. CONCLUSIONES 42
VI. PROPUESTAS A FUTURO 43
VII. REFERENCIAS BIBLIOGRÁFICAS 44

ÍNDICE DE TABLAS

Tabla Página

1. Ejemplos de las variables más significativas para elaborar modelos de susceptibilidad, según diversos investigadores 9
2. Principales métodos estadísticos para distintas combinaciones de variables respuesta y explicativas usada por los GLM 14
3. Funciones de vínculo posibles usadas para distintas distribuciones en GLM 15
4. Ubicación geográfica según el Sistema UTM Zona 18 S - Datum WGS 84 20
5. Rutas de acceso y movilidad dentro de la zona de estudio 22
6. Fuente y resolución espacial de las variables utilizados 26
7. Coberturas vegetales presentes en el área de estudio 27
8. Formaciones geológicas presentes en el área de estudio 28
9. Características de las variables (dependientes e independientes) empleadas 29
10. Matriz de correlaciones de las variables cuantitativas 32
11. Variables de correlaciones altas 32
12. Factores físicos más influyentes para ser usados en el GLM 34
13. Selección del modelo lineal generalizado 34
14. Resumen del proceso de selección de variables para el modelo final 35
15. Test de razón de verosimilitudes usando todas las variables en GLM 62
16. Resumen del ajuste de los datos al modelo lineal generalizado 62
17. Selección del mejor modelo de susceptibilidad a deslizamientos de masas 63

ÍNDICE DE FIGURAS

Figura Página

1. Desplazamiento de rotación en una ladera (Suárez, 2009). 6
2. Aproximado de muertes humanas por deslizamientos registrados durante 1900-2021 (CRED/ UCLouvain, 2021). 6
3. Total de afectados por deslizamientos a niveles nacionales (periodo 1900-2021) (CRED/ UCLouvain, 2021). 7
4. Total de daños estimados (US$) por desastres de deslizamientos entre los años 1900-2021 (CRED/ UCLouvain, 2021). 7
5. Clasificación del resultado de una prueba diagnóstica (Concejero, 2009). 17
6. Curva ROC de dos pruebas diagnósticas A y B, y línea de no discriminación (ND). 19
7. Mapa de ubicación y localización del área en estudio. 21
8. Flujograma del proceso de tesis. 24
9. Gráfico de mosaicos de las variables de tipo cualitativo. 33
10. Mapa de susceptibilidad a deslizamientos en masa del distrito de Sillapata, usando el modelo lineal generalizado logístico final. 36
11. Porcentaje de los niveles de susceptibilidad a deslizamientos de masas obtenidos para el distrito de Sillapata. 37
12. Curva ROC para validación del modelo de susceptibilidad a deslizamientos de masa. 38
13. Gráfico de regresión de las variables del modelo final 39

RESUMEN

Los deslizamientos de masas pueden generar grandes pérdidas y daños, el presente trabajo de investigación estudió su componente de susceptibilidad mediante métodos estadísticos, en el territorio del distrito de Sillapata – Perú, una zona con antecedentes alarmantes. Las variables independientes fueron doce factores físicos, de tipo cualitativo (categóricos) y cuantitativo (continuos) elegidos por su representatividad en campo y factibilidad de obtención; frente a una variable respuesta, de tipo binaria y resultante del inventario de 176 puntos en campo (37,5% puntos de deslizamientos y 62,5% puntos de no deslizamiento); las cuales se relacionaron aplicando el algoritmo de los modelos lineales generalizados con la función de vínculo logit, generando una serie de modelos, posteriormente se usó el Criterio de Información de Akaike para seleccionar el modelo que mejor se ajustó a los datos recogidos de campo. El modelo final, de cinco variables (grados de pendiente, índice de humedad del terreno, índice diferencial de vegetación normalizado, distancia a ríos y curvatura de perfil del terreno), tuvo como mejor umbral de susceptibilidad el valor 0,422 (siendo 0 zona estable y 1 deslizamiento) y una capacidad de discriminación (AUC ROC) de 81,2%. El mapa del modelo final, con valores continuos de susceptibilidad, se categorizó, obteniendo 33,41% con baja susceptibilidad, 33,92% con susceptibilidad media, 23,4% con alta susceptibilidad y 9,27% con muy alta susceptibilidad en todo el territorio. Por último, las variables NDVI y Distancia a ríos mostraron asociación negativa, mientras que Curvatura de perfil, Grados de pendiente y TWI mostraron asociación positiva con la variable respuesta.

Palabras clave: susceptibilidad a deslizamientos de masas, modelo lineal generalizado, mapa.

ABSTRACT

Landslides can generate great losses and damages, this research studied its susceptibility component through statistical methods, in Sillapata district - Peru, an area with alarming antecedents. Independent variables were twelve physical factors, qualitative (categorical) and quantitative (continuous) chosen for their representativeness in the field and feasibility of obtaining them; versus a response variable, of binary type and resulting from the inventory of 176 points in the field (37,5% slip points and 62,5% non-slip points); which were related by applying the algorithm of generalized linear models with a logit link function, generating a series of models, later the Akaike Information Criterion was used to select the model that best fit the data collected in the field. The final model, with five variables (slope degree, topographic wetness index, normalized difference vegetation index, distance from rivers and profile curvature), had as the best threshold susceptibility 0,422 (where 0 was stable zone and 1 landslide) and a discrimination capacity (AUC ROC) of 81,2%. The final model map, with continuous susceptibility values, was categorized, obtaining 33,41% with low susceptibility, 33,92% with medium susceptibility, 23,4% with high susceptibility and 9,27% with very high susceptibility in all the territory. Finally, the variables NDVI and Distance from rivers showed a negative association, while Profile curvature, Slope degree and TWI showed a positive association with the response variable.

Keywords: landslides susceptibility, generalized linear model, map.

I. INTRODUCCIÓN

A nivel mundial, los movimientos de masa son responsables de casi el 17% de las muertes causadas por amenazas naturales (Lacasse et al., 2010), la mayor parte de estas pérdidas se dan en países poco desarrollados debido a la combinación de factores naturales y humanos y que contrastan con la realidad de países desarrollados del grupo recurrente, que reportan más pérdidas económicas que humanas.

El Perú está ubicado en el Cinturón de Fuego del Pacífico, atravesado por la Cordillera de los Andes y el Anticiclón del Pacífico sur y, presenta diversidad y complejidad de geografía, hidrometeorología, geología, entre otras características físicas, indispensables para su riqueza de vida y paisajes, que a la vez lo exponen a diversos fenómenos de origen natural, que nos vienen afectando desde tiempos históricos. De acuerdo con la base de datos internacional de desastres EM-DAT, entre 1900 y 2021, Perú es el sexto país donde ocurren más deslizamientos de masa, sin embargo, tiene el primer lugar en pérdidas fatales (9977 víctimas) y presenta la mayor estimación de daños con $ 4’340’902’000 (precio ajustado al año 2019), por lo mismo a nivel mundial.

Según los registros del Instituto de Defensa Civil, durante los primeros seis meses de cada año, los deslizamientos de masa son la segunda emergencia más frecuente, y son recurrentes en zonas de selva y sierra. Uno de los eventos más alarmantes de la sierra huanuqueña fue el deslizamiento de nivel 4 ocurrido en el distrito de Sillapata, que afectó múltiples infraestructuras, por lo que las autoridades nacionales como el INGEMMET recomendaron la reubicación de la población damnificada.

Pese a los avances en el sistema de gestión de riesgos de desastres, aún es poco común en el Perú evaluar el peligro, vulnerabilidad y riesgo de forma sistemática, mostrando inoperancia en etapas post desastre como la respuesta, rehabilitación y reconstrucción, que en casos menos fatales deja poblaciones con servicios básicos obstruidos de forma prolongada. A la fecha, las bases de datos a nivel nacional no cuentan con información aprovechable a escala local, resultando difícil proteger a la población y las inversiones frente a futuros impactos de desastres por fenómenos naturales.

Los deslizamientos de masa son originados por fenómenos de geodinámica interna y se caracterizan por tener mayor impacto en entornos andinos como la sierra peruana, carácter

violento y recurrente en el tiempo, por ello, es importante contar con estudios que puedan hacer una identificación preliminar de los lugares con mayor probabilidad de ocurrencia (mayor susceptibilidad). Desde un punto de vista objetivo, con enfoques estadísticos es posible obtener zonificaciones de peligro e incluso riesgo, útiles para conocer efectos probables asociados a los deslizamientos y para la planificación territorial.

A continuación, se estudiará el distrito de Sillapata, buscando conocer los factores que podrían intervenir en mayor grado en el modelado de su susceptibilidad por deslizamientos de masa, como resultado se obtendrán zonificaciones con métricas de aproximación a la realidad. Para el estudio se planteó la interrogante: *¿Qué factores físicos intervienen en el modelo de susceptibilidad a deslizamiento de masas en el distrito de Sillapata el año 2021?* y, de acuerdo con las visitas preliminares, se formuló la siguiente hipótesis de investigación: *Los factores físicos que intervienen en el modelo de susceptibilidad a deslizamiento de masas en el distrito de Sillapata el año 2021, son los factores distancia a vías y elevaciones.*

1.1. Objetivo general

Evaluar los factores físicos que intervienen en el modelo de susceptibilidad a deslizamientos de masa en el distrito de Sillapata el año 2021.

1.2. Objetivos específicos

Identificar los factores físicos que intervienen en el modelado de la susceptibilidad a deslizamientos de masa del área de estudio.

Desarrollar el modelo de susceptibilidad a deslizamientos de masa más adecuado.

Determinar la relación entre los factores físicos y el modelo de susceptibilidad a deslizamientos con mejor rendimiento.

II. REVISIÓN DE LITERATURA

2.1. Estado del arte

Recientemente muchos investigadores aplican diferentes métodos para evaluar el fenómeno mundial de deslizamientos de tierras, en este apartado se mencionarán algunas de las investigaciones que han servido de guía para el desarrollo del presente trabajo.

Pourghasemi et al. (2021) evaluaron y mapearon la susceptibilidad a deslizamiento de tierras en una provincia de Irán, empleando y comparando cuatro algoritmos de aprendizaje automático. Previamente, generaron diez factores condicionantes del terreno, de los cuales los más significativos, según el algoritmo de mínimos cuadrados parciales fueron la altitud, la distancia a vías, las unidades litológicas y los grados de pendiente. A través de múltiples validaciones se determinó que el Árbol de Regresión potenciado y Modelo Linear Generalizado tuvieron mejor rendimiento.

En una cuenca de los andes colombianos Ospina-Gutiérrez y Aristizábal (2021) evaluaron el desempeño y capacidad de predicción de 11 modelos de susceptibilidad por movimientos en masa. Se usaron 14 variables predictoras, y a través del ACP se determinó que las variables: curvatura estándar, geología y flujo acumulado, aportaron más a los modelos. La evaluación concluyó que modelos como el Gradiente estocástico, el XGboost y el Bosque Aleatorio, es decir modelos no paramétricos de algoritmos no ensamblados, mostraron mejores resultados.

En otra cuenca, en medio oriente, Mohamed y Pourghasemi (2020), mapearon la susceptibilidad a deslizamiento de tierras, empleando y comparando siete métodos de aprendizaje automático. Para ello se generaron doce variables de deslizamientos, cuya contribución fue evaluada a través de técnicas de bosque aleatorio, los resultados muestran una mayor contribución en: Grados, Aspecto y Longitud de pendiente, seguido por Distancia a vías. En la comparación, el modelo de Bosque Aleatorio (RF) y Análisis Discriminante Lineal (LDA), tuvieron mejor rendimiento.

Por su parte Chen et al. (2019) evaluaron y compararon el rendimiento de dos modelos: Análisis de datos funcionales (ADF) y Modelos Lineales Generalizados, para mapear la susceptibilidad a deslizamientos de tierras en Nanzheng – China. Para su investigación trabajaron con 12 factores condicionantes, cuya importancia se evaluó a través de una

correlación por atributos (CAE), resultando de mayor importancia la altitud, seguido del NDVI y distancia a vías. En el trabajo, el modelo ADF funcionó mejor.

Londoño - Linares (2017), aplicó un método multivariado como el Análisis Discriminante para calcular la susceptibilidad a deslizamientos en la cuenca del río Chinchina – Colombia, y obtener un mapa de zonificación de la susceptibilidad a deslizamientos de tierras. Para ello, generó 17 factores, siete de los cuales fueron descartados; la matriz de estructura y estandarización muestran que Distancia a vías, Elevaciones, Distancia a fallas y a población, fueron más influyentes. El resultado: zonificaciones a escalas regionales, con precisiones y sensibilidades altas.

2.2. Marco teórico

2.2.1. Gestión del riesgo de desastres

La gestión del riesgo es el enfoque y la práctica sistemática de gestionar la incertidumbre para minimizar los daños y las pérdidas potenciales, mientras que la gestión del riesgo de desastres se describe como el proceso sistemático de utilizar directrices administrativas, organizaciones, destrezas y capacidades operativas para ejecutar políticas y fortalecer las capacidades de afrontamiento, con el fin de reducir el impacto adverso y la posibilidad de que ocurra un desastre (UNISDR, 2009).

Durante la conferencia mundial sobre la Reducción de los Desastres y la adopción del “Marco de Acción de Hyogo para 2005 – 2015” – MAH, los países participantes asumieron compromisos como el de “velar para que la Reducción del Riesgo de Desastre constituya una prioridad nacional y local dotada de una sólida base institucional”, generando una corriente de organización basada en la Gestión del Riesgo de Desastres, con un enfoque integral, en el que todos los actores del desarrollo asuman funciones específicas desde la base de sus competencias y atribuciones, priorizando como tema central la GRD y sus componentes (gestión Prospectiva, Correctiva y Reactiva) y no solo la administración del desastre (PCM, 2014).

2.2.2. Peligro o amenaza

La oficina de Estrategia Internacional para la Reducción de Desastres de las Naciones Unidad - UNISDR (2009), define la amenaza como un fenómeno, sustancia, actividad humana o condición peligrosa que pueden ocasionar la muerte, lesiones u otros impactos a la salud, al igual que daños a la propiedad, la pérdida de medios de sustento y de servicios, trastornos sociales y económico, o daños ambientales. El Marco de Acción de Hyogo menciona

que las amenazas relevantes en el campo de la reducción del riesgo de desastres son amenazas de origen natural y desastres y riesgos ambientales y tecnológicos conexos, siendo un ejemplo claro de peligro los deslizamientos de masa.

Por su parte CENEPRED (2014), describe al peligro originado por fenómenos de origen natural como la probabilidad de que un fenómeno, potencialmente dañino, de origen natural, se presente en un lugar específico, con una cierta intensidad y en un período de tiempo y frecuencia definidos; mencionando que, fuera de Perú, puede ser denominado amenaza.

2.2.3. Deslizamientos de masa

Según la Asociación Internacional de Geología, la UNESCO y la United Nations Relief Organization (UNDRO) el término deslizamiento de tierra o *landslide* en inglés, corresponde a una categoría de fenómenos caracterizados por movimientos de masa donde la gravedad es la principal fuerza en juego y la velocidad del desplazamiento es una de sus principales características; además, tienen límites nítidos (Montiel et al., 2007). La definición de Suarez (2009) precisa que son movimientos de masa de roca, residuos o tierra hacia abajo de un talud, cuya naturaleza incluye procesos originados por fuerzas gravitacionales, hidráulicas, entre otras. El INDECI (s/a) clasifica a los deslizamientos de tierra como un peligro natural y lo define como la ruptura y desplazamiento de pequeñas o grandes masas de suelos, rocas, rellenos artificiales o combinaciones de estos, pudiendo ser procesos lentos o violentos.

Los desplazamientos en masa se subdividen en subtipos denominados deslizamientos rotacionales, deslizamientos traslacionales o planares y, deslizamientos compuestos (de rotación y traslación). La superficie de rotura define el tipo de deslizamiento, y esta diferenciación es importante porque puede definir el sistema de análisis y el tipo de estabilización a emplear (Suárez, 2009).

Suarez (2009) además menciona que las superficies curvas, cóncavas o en forma de cuchara se asocian a deslizamientos rotacionales, que generalmente tienen una relación Dr/Lr entre 0,15 y 0,33 (ver **figura 1**); mientras las superficies semiplanas o ligeramente onduladas se asocian a los movimientos traslacionales y generalmente tienen una relación Dr/Lr de menos de 0.1; por último, es frecuente que el patrón de desplazamiento general incluya movimientos de traslación y rotación, presentando hundimientos o extensiones laterales en forma conjunta, a este se le denomina deslizamientos compuestos. La mayoría de movimientos incluyen varios tipos de desplazamiento aunque suele predominar uno.

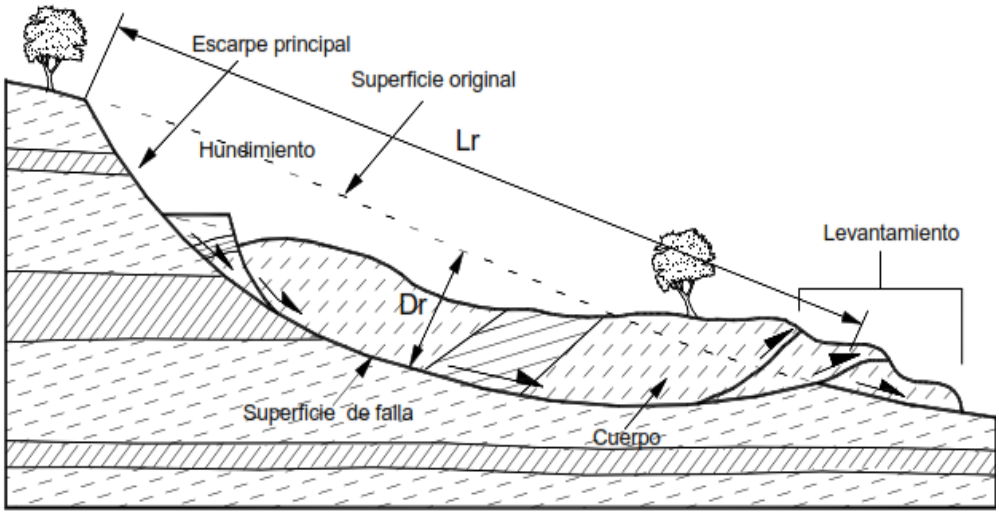


Figura 1. Desplazamiento de rotación en una ladera (Suárez, 2009).

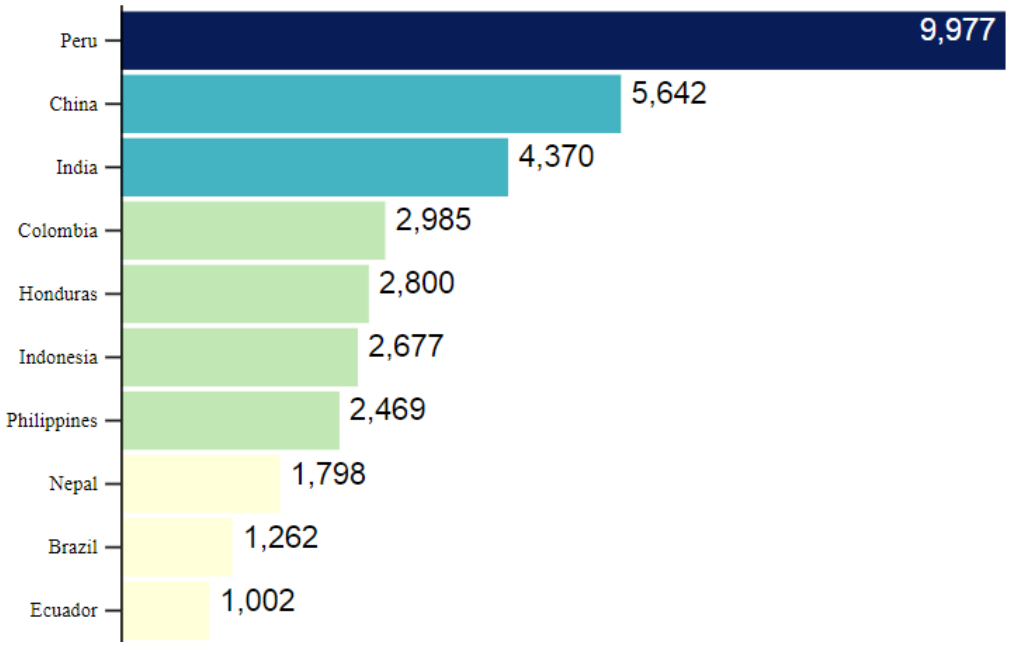


Figura 2. Aproximado de muertes humanas por deslizamientos registrados durante 1900-2021 (CRED/ UCLouvain, 2021).

Los deslizamientos producen cambios en la geomorfología, daños en el ambiente, la infraestructura y la vida humana (ver **figura 2**), siendo el total de daños superior al de otros eventos naturales como terremotos e inundaciones, especialmente en países con incipiente planificación como el Perú (ver **figura 3** y **4**). La ubicación más frecuente coincide ser en zonas montañosas, ejemplo de ello son los países cercanos a la cordillera de los andes, que según la base de datos internacional de desastres EM-DAT (CRED/ UCLouvain, 2021) Perú es el sexto país con mayor ocurrencia de deslizamientos importantes a nivel mundial (25 eventos entre 1900-2021) y a nivel del continente americano solo es superado por Colombia.

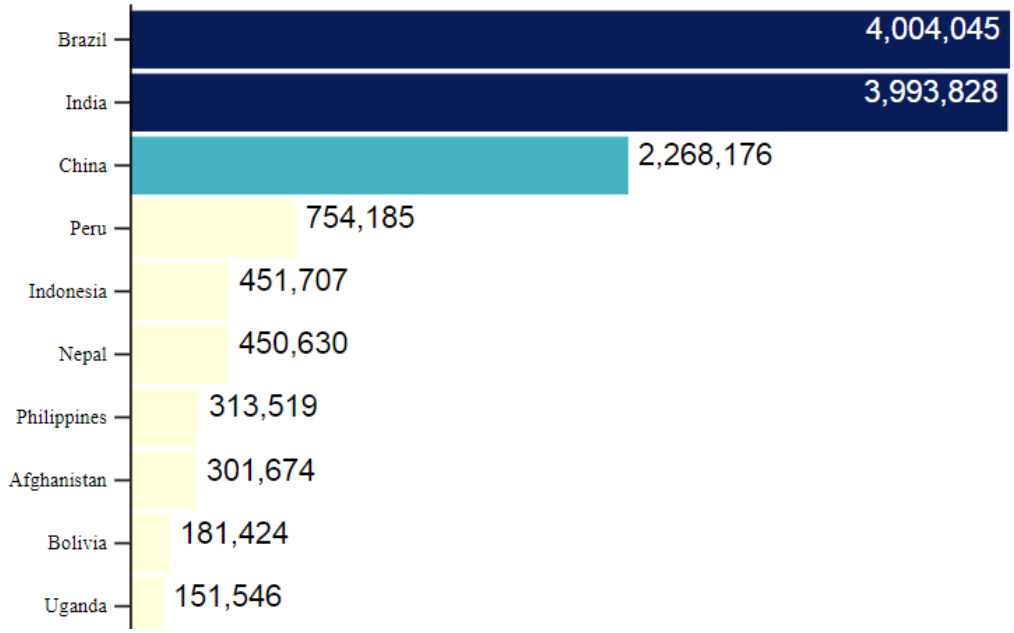


Figura 3. Total de afectados por deslizamientos a niveles nacionales (periodo 1900-2021) (CRED/ UCLouvain, 2021).

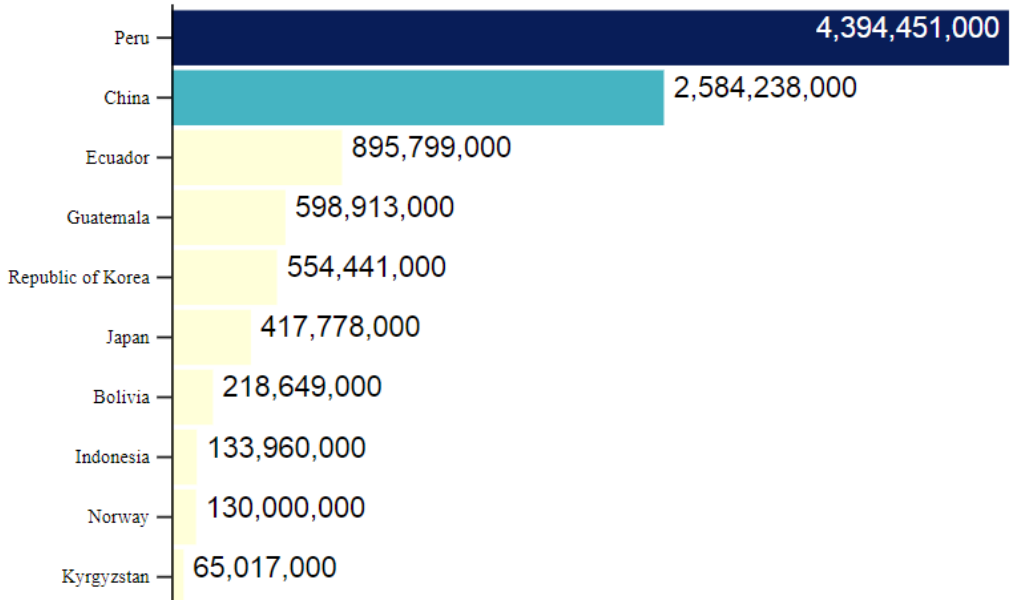


Figura 4. Total de daños estimados (US$) por desastres de deslizamientos entre los años 1900-2021 (CRED/ UCLouvain, 2021).

Problemas de esta naturaleza aumentarán en el futuro debido a la expansión poblacional; sin embargo, muchas de estas pérdidas son evitables si el problema se identifica previamente y se implementan medidas de prevención o control (Mohamed y Pourghasemi, 2020; Suárez, 2009).

2.2.4. Susceptibilidad a deslizamiento en masa

Es la distribución espacial de la probabilidad de ocurrencia de deslizamientos de masa como resultado de complejas asociaciones entre varios factores ambientales (Pourghasemi et al., 2021; Mohamed et al., 2020; Londoño-Linares, 2017; Hess et al., 2017), factores ambientales que pueden ser clasificados en condicionantes y desencadenantes del fenómeno y su respectivo ámbito geográfico (CENEPRED, 2014) o, que pueden ser diferenciados en factores naturales y causados por actividades humanas (Pourghasemi et al., 2021; Londoño-Linares, 2017; Suárez, 2009; Moreno et al., 2006; Cubito et al., 2005) (ver **tabla 1**).

2.2.4.1. Mapeo de susceptibilidad a deslizamientos

La predicción de deslizamientos y la zonificación de susceptibilidad a los mismos, puede ser un insumo esencial para que los formuladores de políticas desarrollen estrategias de mitigación del peligro y/o planificación del territorio, por ello muchas investigaciones científicas relativas a gestión de desastres por fenómenos naturales, se han enfocado en la evaluación de métodos y generación de mapas de susceptibilidad a deslizamientos (Pourghasemi et al., 2021; Ospina-Gutiérrez y Aristizábal, 2021; Sun et al., 2020) los que posteriormente pueden complementarse con frecuencia de ocurrencia (análisis de peligros) y con análisis de exposición y vulnerabilidad, para obtener cuantificaciones de riesgo, en términos de pérdidas y daños, o zonificaciones (Londoño-Linares, 2017).

Existen múltiples métodos para evaluar y mapear la susceptibilidad a movimientos en masa, cuyo rendimiento dependerá de la calidad y cantidad de datos disponibles, de la escala de trabajo y de la selección de una metodología adecuada, existiendo dos enfoques: cuantitativos y cualitativos (Ospina-Gutiérrez y Aristizábal, 2021; Wang et al., 2020; Chen et al., 2019; Hess et al., 2017) que en general se distinguen en cuatro principales métodos: a) geomorfológicos, b) heurísticos, que trabajan cualitativamente a través del juicio de expertos, siendo subjetivos; los cuantitativos presentan objetividad y son c) determinísticos, que establecen relaciones matemáticas basados en análisis de estabilidad, su aplicación es efectiva a escala de un solo talud y d) estadísticos o probabilísticos, que establecen relaciones estadísticas multivariadas o binarias aplicables en áreas heterogéneas (Ospina-Gutiérrez y Aristizábal, 2021; Mohamed y Pourghasemi, 2020; Wang et al., 2020; Londoño-Linares, 2017).

Tabla 1. Ejemplos de las variables más significativas para elaborar modelos de susceptibilidad, según diversos investigadores

Varnes (1984)	**Binaghi, et al (1998)**	**Guzzerti, et al (1999)**	**Baeza & Corominas (2001)**	**Cleriel, at al (2002)**	**Dai & lee (2002)**	**Zezere (2002)**	**Remondo, et al (2003)**	**Lee & Talib (2005)**
Litología	Litología	Litología	Litología	Litología	Litología	Litología	Litología	Litología
Vegetación	Uso del suelo	Vegetación	Uso del suelo	Uso del suelo	Uso del suelo		Uso del suelo	Índice de vegetación
Geomorfología			Espesor de depósito superficial			Espesor de depósito superficial	Espesor de depósito superficial	
	Elevaciones Orientaciones Pendientes	Elevaciones Orientaciones Pendientes Curvatura	Densidad de árboles Elevaciones Orientaciones Pendientes Concavidad-Convexidad Pendiente de la cuenca Área de la cuenca Longitud de la cuenca	Pendientes	Elevaciones Orientaciones Pendientes Concavidad-Convexidad	Pendientes	Elevaciones Orientaciones Pendientes Concavidad-Convexidad Pendiente de la cuenca Área de la cuenca Longitud de la cuenca Rugosidad	Orientaciones Pendientes Concavidad-Convexidad
Estructuras	Distancia a fracturas			Relación pendiente-estratificación	Estructura	Relación pendiente-estratificación		Distancia a fracturas
Condiciones Hidrológicas	Distancia a red de drenaje				Distancia a red de drenaje	Presencia de canales fluviales		Distancia a red de drenaje
Condiciones Climáticas		Condiciones Climáticas y Sísmicas		Precipitación			Insolación	

Fuente: Bonachea (2006).

Los métodos estadísticos manejan la hipótesis de que los deslizamientos futuros tendrán patrones de distribución similares a los ya registrados (Mohamed y Pourghasemi, 2020), tratando de explicar así la relación entre un conjunto de factores de inestabilidad conocidos (factores del terreno o ambiente) y el registro de deslizamientos pasados y presentes, estableciendo relaciones entre variables explicativas y una variable dependiente (Ospina-Gutiérrez y Aristizábal, 2021; Londoño-Linares, 2017). Los SIG y las técnicas de sensoramiento remoto facilitan el manejo de información provenientes de una extensión de área, por ello son herramientas de uso extendido en estas aplicaciones (por ejemplo, Ospina-Gutiérrez y Aristizábal, 2021; Hassangavyar et al., 2020; Sun et al., 2020; Mohamed y Pourghasemi, 2020; Chen et al., 2019; Londoño-Linares, 2017; Schachter, 2008; Baeza, 1994).

A pesar de los avances, la posibilidad de un modelo 100% correcto con los métodos estadísticos es altamente complejo, pues dependerá de condiciones como la cantidad de eventos positivos, tamaño de los eventos, selección de factores adecuados (ver **tabla 1** sobre variables recomendadas), la optimización de la información (la pérdida de resolución implica pérdida de confiabilidad en el resultado) y el error de los datos (Sun et al, 2020; Schachter, 2008). Incluso los métodos más recientes, como la aplicación de inteligencia artificial aún no han consensuado el modelo o modelos más adecuados para estudiar la susceptibilidad a deslizamientos de masa (Pourghasemi et al., 2021; Chen et al., 2019) y se vienen probando aplicaciones en los procesos de cálculo y comparándolos (por ejemplo, Hassangavyar et al., 2020; Wang et al., 2020; Sun et al, 2020; Chen et al., 2019) en busca de mejorar en la precisión de predicción de estos eventos que podrían causar desastres.

2.2.4.2. Variables empleadas en el presente trabajo

De acuerdo a las variables recomendadas y los datos disponibles, se escogieron un grupo de variables para estudiar en la zona de estudio, estas fueron:

1. Elevación

La elevación o altitud del terreno, es controlada por factores como geología, geomorfología y meteorología, esto incluye las unidades litológicas, meteorización, acción del viento y patrones de precipitación (Mohamed y Pourghasemi, 2020) mismas que pueden afectar en la ocurrencia de deslizamientos, siendo considerado un predictor significante de susceptibilidad a deslizamientos en bastantes estudios (Vasu y Lee, 2016; Kavzoglu et al., 2019; Chen et al., 2019; Pourghasemi et al., 2021).

2. Grados de Pendiente

Uno los factores que más influye en la inestabilidad de suelos son los grados de pendiente, pues la inclinación de una pendiente está relacionada con los esfuerzos cortantes que actúan sobre la pendiente y con el desplazamiento de la masa del deslizamiento (Tien Bui et al., 2017; Mohamed y Pourghasemi, 2020), por lo tanto, la fuerza impulsora de los movimientos de masa aumenta al aumentar el ángulo de pendiente (Guillard y Zezere, 2012). A medida que aumentan los grados de pendiente, la frecuencia de deslizamientos se intensifica significativamente (Pandey, 2015; Pourghasemi et al., 2021).

3. Longitud de flujo

Es la longitud máxima que podría recorrer una gota de agua que drenaría hacia una celda (Londoño-Linares, 2017). Esta variable también indica el tamaño acumulado de la cuenca, la capacidad de esta para concentrar agua subterránea y la posible acumulación de sedimentos; en estudios se ha observado que las celdas cercanas a una cresta de cuenca muestran menos roturas que las celdas alejadas (Landinez y Beltrán, 2019). Al tratarse de la longitud medida sobre el terreno, este parámetro se pondera con la pendiente y sus unidades son metros, en este estudio.

4. Índice de Humedad Topográfica

Refleja el contenido de humedad controlado por la topografía y determina la dirección del flujo de agua subterránea (Hassangavyar et al., 2020). El índice consiste en la descarga que contribuye el área de pendiente ascendente de cada celda de la cuadrícula (α) y el gradiente de pendiente (β) (Beven y Kirkby, 1979).

$$TWI = \ln \frac{\alpha}{\tan \beta}$$

Según Meinhardt et al. (2015) hay un incremento en la densidad de deslizamientos acompañado de decrecientes valores de TWI, lo cual puede explicarse por la tendencia del gradiente de pendiente a ser mayor sobre valores más bajos de TWI, o también, que las áreas con valores TWI altos son muy planas para deslizamientos de tierra, o que estos suelos ya están erosionados. Para Sun et al. (2020) el contenido y distribución de agua en el suelo, afectaría la condición de roca, suelo y vegetación sobre la superficie de la pendiente, impactando sobre los deslizamientos.

5. Índice diferencial de agua normalizado NDWI

Sirve para la teledetección del agua líquida generalmente en la vegetación, útil para diferenciar suelos desnudos (húmedos o secos) y también áreas con mezclas de suelo y vegetación (Gao, 1996). Normalmente utiliza dos canales angostos del espectro electromagnético centrados entre 0,86um y 1,24um, infrarrojo cercano (NIR) e infrarrojo corto (SWIR), a través de la ecuación:

$$NDWI = \frac{(NIR - SWIR)}{(NIR + SWIR)}$$

Dados los recursos satelitales de PlanetScope, se usa la adaptación hecha por McFeeters (1996), que se basa en la sustitución de la banda SWIR por la banda visible verde (G), que permite maximizar la reflectancia de la vegetación, quedando resaltadas las masas humedecidas, con la ecuación:

$$NDWI = \frac{(G - NIR)}{(G + NIR)}$$

Los valores de NDWI fluctúan entre -1 a 1. Para suelo y vegetación son valores negativos o cercanos a cero, y para cuerpos de agua resulta positivo y variable dependiendo de su turbidez (sin diferenciar si fuese contenido de sedimentos o clorofila a) (McFeeters, 1996).

6. Índice diferencial de vegetación normalizado NDVI

El NDVI es considerado una variable que influye en los modelos de susceptibilidad de deslizamientos en varios estudios (por ejemplo, Mohamed y Pourghasemi, 2020; Sun et al., 2020; Chen et al., 2019), debido a que refleja la densidad de vegetación, que cumple un importante rol inmovilizando gran cantidad de agua y aumentando la resistencia al corte y la cohesión del suelo de la masa litológica (Mohamed y Pourghasemi, 2020). Su valor varía de -1 a 1, y se halla con la ecuación:

$$NDVI = \frac{(NIR - R)}{(NIR + R)}$$

Donde los espectros electromagnéticos satelitales NIR y R son, respectivamente Infrarrojo cercano y Rojo.

7. Cobertura vegetal

Las actividades humanas como la expansión de áreas urbanas, la deforestación o conversión de paisajes naturales a pastos, juegan significantes roles en la

ocurrencia de deslizamientos, pues alteran las condiciones geológicas originales, y la estabilidad de las pendientes se puede llegar a modificar (Mohamed y Pourghasemi, 2020; Hassangavyar et al., 2020).

8. Geología

La geología se considera un factor de significancia, sus características tienen diferentes valores de resistencia frente a deslizamientos (Pourghasemi et al., 2021), pues la litología y formaciones geológicas en una región determinan los tipos, extensión, profundidad y velocidad de deslizamientos (Hassangavyar et al., 2020). Por su parte, las variaciones litológicas, impactan significativamente sobre diversos tipos de peligros, al variar sus características físicas y mecánicas, como el tipo, la resistencia, grado de intemperismo, durabilidad, densidad y permeabilidad de las unidades litológicas (Mohamed y Pourghasemi, 2020).

9. Distancia a flujos de agua

La escorrentía a lo largo de estos flujos de agua sirve para reducir los fenómenos (encausando los flujos de agua), a la vez que aumenta la presión de agua en las zonas más adyacentes a los flujos (o cauces), iniciando deslizamientos (Mohamed y Pourghasemi, 2020), por ello el impacto más importante de los ríos y quebradas sobre los deslizamientos resulta en la erosión fluvial (Pourghasemi et al., 2021).

10. Distancia a vías

El establecimiento de vías sobre territorios montañosos requiere actividades como corte o excavación de pendientes, llevando a cambiar las condiciones originales de la geología local y debilitando el soporte natural de las pendientes rocosas (Mohamed y Pourghasemi, 2020), estas actividades son un factor antropogénico de impacto significativo sobre la incidencia de deslizamientos (Pourghasemi et al., 2021; Hassangavyar et al., 2020).

11. Curvatura plana

Es una medida perpendicular a la dirección de la pendiente máxima, es decir, brinda información acerca del flujo del agua, si este coincide o se bifurca en una ladera montañosa (Pourghasemi et al., 2021; Meinhardt et al., 2015) y, tiene un impacto directo en la convergencia y dispersión de las escorrentías superficiales (Mohamed y Pourghasemi, 2020).

12. Curvatura de perfil

Es una medida en la dirección de la pendiente máxima (Meinhardt et al., 2015) y afecta la deposición del material gestionando la aceleración o desaceleración de estos materiales en pendiente (Mohamed y Pourghasemi, 2020), además mide la variación de la frecuencia de la pendiente, controla la velocidad del flujo de agua que se drena a la superficie, lo que influye en la erosión y el movimiento de la pendiente inferior del depósito (Pourghasemi et al., 2021; Meinhardt et al., 2015).

2.2.5. Modelos lineales generalizados

Los GLM (del inglés *Generalized Linear Model*) son una extensión de los modelos lineales que son capaces de tratar variables dependientes que tengan distribuciones no normales y varianzas no constantes (heterocedasticidad). Podemos considerar emplear GLM en caso de que la variable respuesta sea: conteo de casos (por ejemplo, la abundancia de una especie), conteo de casos expresados como proporciones (porcentaje de plántulas muertas en un experimento de vivero) o, respuesta binaria (por ejemplo, vivo o muerto, infectado o no infectado) (Cayuela, 2011). El algoritmo fue establecido por Nelder y Wedderburn (1972) y fue enmarcado por McCullagh y Nelder (1989).

Cayuela (2011) detalla dos propiedades importantes de los GLM: 1) La estructura de sus errores, debido a la ausencia de normalidad en la variable dependiente y por lo tanto también en los valores estimados por el modelo, permiten especificar otros tipos de distribución de errores como Poisson, Binomial, Gamma y Exponencial. 2) La función de vínculo, es la encargada de linealizar la relación existente entre la variable respuesta y la(s) variable(s) explicativa(s), transforma la variable respuesta y acota las predicciones del modelo de acuerdo con la naturaleza de la variable respuesta.

Tabla 2. Principales métodos estadísticos para distintas combinaciones de variables respuesta y explicativas usada por los GLM

Variable respuesta	Variables explicativas	Métodos y/o análisis
Continua	Binaria	t-test
	Nominal, >2 categorías	Análisis de varianza
	Ordinal	Análisis de varianza
	Cualitativas y continuas	Regresión múltiple
Nominal / Binaria	Cualitativas	Tablas de conting./ Logística/ Otros glm
	Continuas	Logística / Otros glm
	Categóricas y continuas	Logística / Otros glm

Variable respuesta	Variables explicativas	Métodos y/o análisis
Nominal > 2 categorías	Nominal Categóricas y continuas	Tablas de conting. Logística multinomial
Ordinal	Categóricas y continuas	Logística ordinal
Discretas	Categóricas Categóricas y continuas	Modelos log-lineales Regresión de Poisson

Fuente: Durbán (2014) y Cayuela (2011).

Cabe mencionar que al emplear una función de vínculo los valores estimados por el modelo no son valores de "y", sino que, para obtener los valores estimados de la variable respuesta se debe aplicar la función inversa a la función de vínculo usada (Cayuela, 2011).

A pesar de la existencia de funciones de vínculo canónicas que se aplican por defecto a cada tipo de distribución, no siempre se debe usar una única función de vínculo para una distribución determinada, sino que se recomienda comparar diversas funciones de vínculo para un mismo modelo y determinar cuál presenta mejor ajuste del modelo a los datos (Cayuela, 2011). A continuación, se muestran funciones vínculo posibles para cada distribución de errores.

Tabla 3. Funciones de vínculo posibles usadas para distintas distribuciones en GLM

Distribución de errores	Función de vínculo	Fórmula	Otras funciones de vínculo posibles
Normal	Identidad	μ	Logarítmica
Poisson	Logarítmica	$\log(\mu)$	Identidad, Raíz cuadrada
Binomial	Logit	$\log\left(\frac{\mu}{n-\mu}\right)$	Logarítmica
Gamma	Recíproca	$\frac{1}{\mu}$	Identidad, Logarítmica

Fuente: Cayuela, 2011.

De acuerdo con el principio fundamental de GLM (Chen et al., 2019), la esperanza y variaciones de las variables de respuesta se pueden calcular respectivamente:

$$\mu_i = E[Y_i] = g^{-1}\left(\sum_j X_{ij}\beta_j + \xi_i\right)$$

$$var[Y_i] = \frac{\phi V(\mu_i)}{\omega_i}$$

Donde: Y_i es el vector de variables de respuesta; X_{ij} es la matriz de variables explicativas; β_j es el vector de parámetros pendientes; ξ_i son los términos de interferencia; g(x) es la función de vínculo correspondiente; V(x) es la función de varianza; ϕ es el parámetro de dispersión de V(x), y ω_i es el peso del valor observado de orden i (Chen et al., 2019).

De ese modo los GLM usan la regresión multivariante para determinar la asociación entre las variables explicativas de deslizamientos y la variable dependiente (registro de presencia y ausencia de deslizamientos) (Pourghasemi et al., 2021). La ecuación de la probabilidad de ocurrencia del evento se puede expresar:

$$P = \frac{exp(\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_i x_i)}{1 + exp(\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_i x_i)}$$

Donde: P es la probabilidad de ocurrencia del evento Y (Y, a su vez, representa la variable dependiente o respuesta) y, α_i son los coeficientes de regresión logística de las variables explicativas empleadas.

Por último, se debe transformar la función logística, y la función de vínculo quedaría:

$$g(y_i) = \alpha_0 + \sum \alpha_i x_i + \varepsilon_i$$

Donde: ε_i es el error residual.

2.2.6. Criterio de información de Akaike

En el proceso de construcción de un modelo, cuando no se conoce exactamente el mecanismo generador de los datos, la forma funcional con que intervienen las variables, ni tampoco los desfases temporales que pueden existir, los contrastes estadísticos formales pierden importancia. Akaike (1974), desde la teoría de la decisión estadística, traslada el problema del ajuste de un modelo a la elección de la función de pérdida más adecuada, proponiendo elegir como función de pérdida (o criterio de especificación) el mínimo criterio AIC (Márquez, 2002).

El criterio de Akaike AIC se basa en la medida de información de Kullback-Leiber (1951), la cual permite interpretar la distancia entre dos distribuciones (la observada y la teórica) a partir de la log-verosimilitud de un modelo (Márquez, 2002), esta distancia es relativa y depende del conjunto de datos, por tanto un valor individual de AIC no es

interpretable por sí solo, pues tienen sentido siempre que se realizan comparaciones empleando los mismos datos (Martínez et al., 2009). El AIC se calcula a partir de:

$$AIC = -2\log \mathcal{L}(\hat{\theta}) + 2\mathrm{K}$$

Donde: log $\mathcal{L}(\ddot{\theta})$ es el logaritmo de la máxima verosimilitud, que permite determinar los valores de los parámetros libre de un modelo estadístico, y K es el número de parámetros del modelo.

En resumen, el AIC halla una función de pérdida que al ser minimizada permite obtener el modelo que mejor se ajusta a los datos (Márquez, 2002), sirve para evaluar la complejidad de dicho modelo e incluso para comparar modelos similares con distintos grados de complejidad o modelos con las mismas variables pero funciones de vínculo distintas (Cayuela, 2011).

2.2.7. Validación del modelo

Una de las condiciones para conocer el desempeño en una prueba o modelo, es la validez, es el grado en que un modelo o prueba mide lo que se desee medir, y se determina con la sensibilidad y la especificidad (Pita y Pértegas, 2003). El esquema de partida para análisis de validez es la tabla de doble entrada que muestra la **figura 5** en que se representa, por un lado, el resultado de una decisión (en las filas), y por otro, el estado real conocido o no (en las columnas) (Concejero, 2009; Pita y Pértegas, 2003). Así, se reproducen, cuatro posibles resultados, que se representan en dicha tabla y, conociendo el número de valores de cada una es posible calcular la sensibilidad y especificidad para el punto de corte que los origina (Cerda y Cifuentes, 2011).

		Estado real		
		Positivo	Negativo	Suma
Diagnóst.	Positivo	Verdaderos Positivos (VP) a	Falsos Positivos (FP) (Falsas alarmas) b	a+b
	Negativo	Falsos Negativos (FN) (omisiones) c	Verdaderos Negativos (VN) d	c+d
	Suma	(a+c)	(b+d)	a+b+c+d

Figura 5. Clasificación del resultado de una prueba diagnóstica (Concejero, 2009).

La **figura 5** describe completamente el comportamiento de un sistema diagnóstico o de detección con dos únicos posibles resultados (positivo o negativo). Además, es necesario especificar dos cantidades para obtener una representación completa del comportamiento diagnóstico. Estas dos medidas son la sensibilidad y especificidad (Concejero, 2009).

Sensibilidad: es la proporción de casos diagnosticados como afirmativos, a partir del criterio o regla de decisión establecido, en los que se comprueba que efectivamente sucede el estado que se pretende detectar o diagnosticar (Concejero, 2009). También es conocida como fracción de verdaderos positivos FVP (Pita y Pértegas, 2003).

Especificidad: es la proporción de casos diagnosticados como negativos, a partir de la regla de decisión establecida, en los que se comprueba que efectivamente no sucede el estado que se pretende detectar o diagnosticar (Concejero, 2009). Conocida como fracción de verdaderos negativos FVN (Pita y Pértegas, 2003).

$$Sensibilidad = \frac{VP}{VP + FN} \qquad Especificidad = \frac{VN}{VN + FP}$$

El análisis estadístico ROC (*Receiver operating characteristic*) grafica la tasa de éxito (sensibilidad) frente a la tasa de falsa alarma (especificidad) de cada posible punto de corte de un diagnóstico cuya escala de medición es continua, para tareas de detección con sólo dos resultados posibles (sí/ no, presente/ ausente) (Concejero, 2009; Cerda y Cifuentes, 2011). La curva ROC se construye en base a la unión de distintos puntos de corte, correspondiendo el eje Y a la sensibilidad y el eje X a (1 – especificidad), esto permite determinar el punto de corte en el que se alcanza la sensibilidad y especificidad más altos y, evaluar la capacidad de diferenciar presencias de ausencias, a través del mayor Índice de Youden, que es capaz de determinar el punto de corte con la sensibilidad y especificidad más altas para un mismo punto (Cerda y Cifuentes, 2011).

En la **figura 6**, para cada curva ROC las flechas indican el punto de corte que determina la sensibilidad y especificidad conjuntas más alta (Cerda y Cifuentes, 2011).

El área bajo la curva ROC (AUC o *Area under curve*) es una herramienta útil para evaluar la capacidad discriminativa del modelo a lo largo de todo el rango de puntos de corte posibles. La línea ND de la **figura 6** describe lo que sería una curva ROC incapaz de discriminar presencias de usencias y una prueba diagnóstica tendrá mayor capacidad

discriminativa en la medida que sus puntos de corte tracen una curva ROC lo más cercana posible a los lados izquierdos y superior del gráfico. Si una curva ROC coincide con la línea ND, poseerá un AUC = 0,5, y a medida que el AUC de una prueba se acerque a 1,0 (prueba o modelo perfecto), mayor será su capacidad discriminativa (Cerda y Cifuentes, 2011).

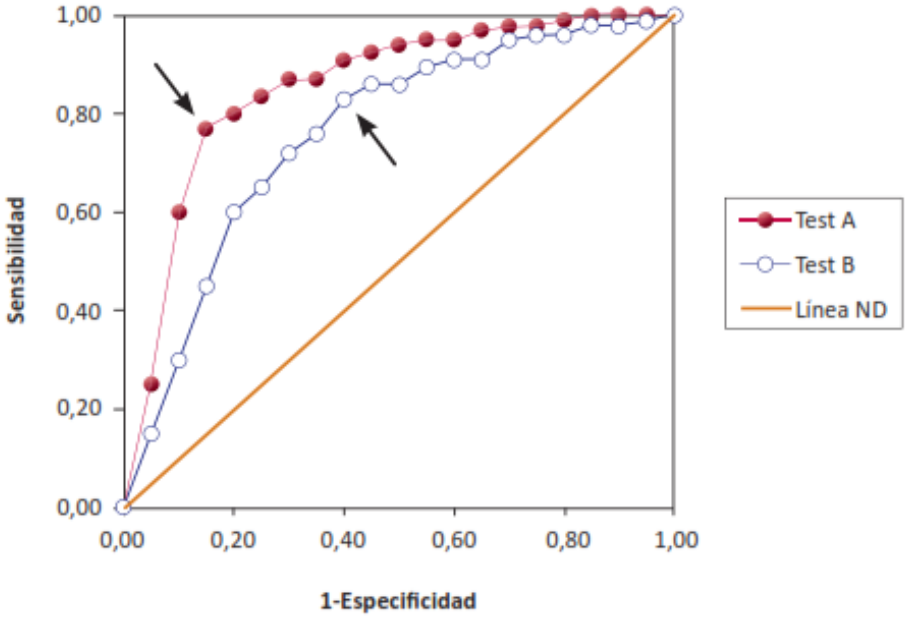


Figura 6. Curva ROC de dos pruebas diagnósticas A y B, y línea de no discriminación (ND).

III. MATERIALES Y MÉTODOS

3.1. Lugar de ejecución

3.1.1. Ubicación de la zona de estudio

El distrito de Sillapata, ubicado en la zona andina del país, políticamente se ubica en la provincia de Dos de Mayo, región Huánuco, y se estudia en el presente trabajo debido a la geodinámica externa activa, tal es así que el año 2019, ocurrió un desastre de deslizamiento de masas que dejó cuantiosos daños, por lo que INGEMMET (2019) recomendó el traslado del centro poblado a una zona de acogida (ver el mapa de ubicación en **figura 7**).

A partir de la división política de límites distritales del INEI y del modelo digital de elevación, se tiene las siguientes coordenadas de ubicación geográficas:

Tabla 4. Ubicación geográfica según el Sistema UTM Zona 18 S - Datum WGS 84

Punto	Este	Sur	Altitud
Punto límite Norte	306600.1	8923808.5	3068
Punto límite Oeste	301960.1	8916457.9	3176
Punto límite Sur	306549.6	8908671.4	3572
Punto límite Este	309771.6	8921176.1	3416
Punto más bajo	305910	8923109	3076
Punto más alto	307031.8	8916903.8	4073
Capital	305478	8920898	3423

Fuente: Elaboración propia.

3.1.1. Accesibilidad

El acceso desde la ciudad de Huánuco se realiza a través de la ruta nacional PE-3N, hasta la ciudad de La Unión, a partir de la cual existen 3 rutas vecinales de acceso y otras rutas de movilidad hacia distintos centros poblados de la zona de estudio (ver **tabla 5**).

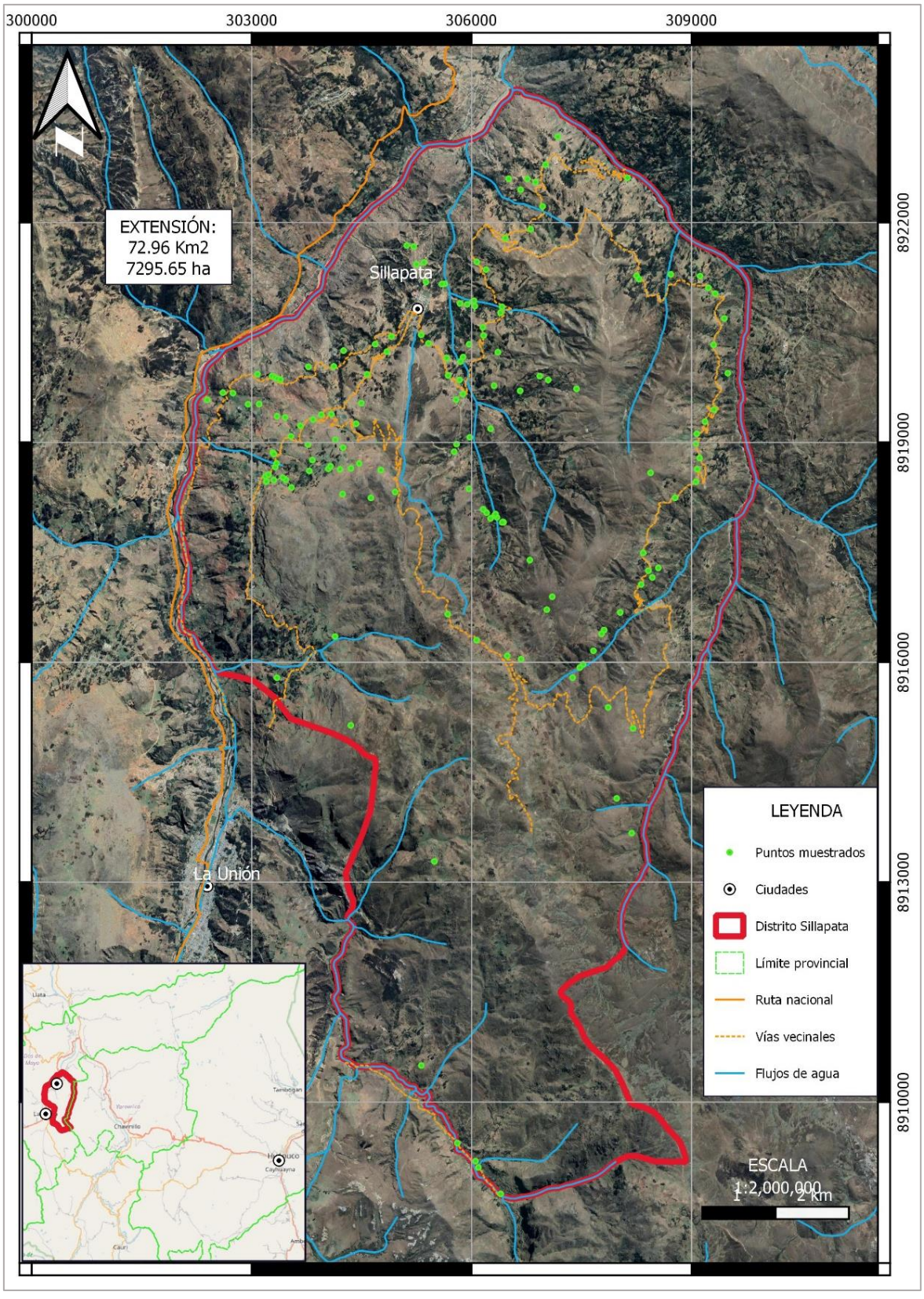


Figura 7. Mapa de ubicación y localización del área en estudio.

Tabla 5. Rutas de acceso y movilidad dentro de la zona de estudio

Ruta vial	Trayectoria referencial	Distancia	Jerarquía
HU-736	La Unión – Sillapata	7.55 Km	Trocha
HU-719	La Unión - Sillapata - Yanas	18.37 Km	Trocha
HU-738	HU-719 – Buena vista	13.03 Km	Trocha
HU-737	HU-719 – Santa Rufina	17.54 Km	Trocha
HU-735	La Unión (Chumipata) - Progreso	2.97 Km	Trocha

Fuente: Provías Nacional, Google satélite. Elaboración propia.

3.1.2. Condiciones ambientales

El distrito se extiende al margen derecho del río Vizcarra, efluyendo todas sus aguas al mismo; de acuerdo con el Observatorio del agua – ANA, pertenece al curso alto de la Intercuenca Alto Marañón V.

Según el Mapa ecológico del Perú (MINAGRI, 1995), el distrito está influenciado por las zonas de vida Bosque húmedo montano tropical y el Bosque seco montano bajo tropical. La clasificación climática por el método de Thornthwaite (SENAMHI, 2020) indica que el distrito cuenta con los climas: lluvioso con otoño e invierno seco/ frío, en menor parte también existe el lluvioso con abundante humedad en todas las estaciones/ templado y, el semiseco con otoño e invierno seco/ templado. El MINAM (2015), indica como tipo de cobertura 'Agricultura costera y andina', de productos nativos como papa, habas, olluco, quinua, cebada, trigo, etc., actualmente existen terrenos en zonas de mayor altitud en desuso, carencia de infraestructuras de riego funcionales (predominando la agricultura de secano), además plantaciones de árboles nativos y exóticos que generalmente sirven solo para limitar propiedades. La cobertura de 'Pajonales andinos' se encuentra en la parte de mayor altitud del distrito.

3.1.3. Características socioeconómicas

El distrito de Sillapata fue declarado en emergencia por el DS N° 040-2019-PCM debido a los eventos de deslizamientos de masa ocurridos el mes de junio. El distrito cuenta con una población total de 1'814 personas, en su mayoría adultos y adultos mayores (INEI 2017), a la fecha no cuenta con Plan de Desarrollo Urbano ni con zonificación de uso de suelos, la provincia, también presenta desactualización respecto a sus instrumentos de planificación territorial.

Según el PDC Huánuco (GRPPAT, 2016), hay mayor presencia de las actividades económicas: agrícola, industria y comercio. La actividad agrícola es diversificada, en su mayoría destinada al autoconsumo y demanda local y, la topografía accidentada hace difícil la aplicación de tecnologías nuevas para mejorar la productividad. El producto con mayor producción en las campañas de cultivo es la papa.

La ganadería característica es de baja calidad genética, pues se cría a campo abierto, sin adecuada orientación técnica, salvo excepciones. La minería es desarrollada artesanalmente en varias zonas de Dos de Mayo y el distrito de Sillapata es uno de los lugares destacables en la explotación de mineral aurífero (GRPPAT, 2016). Se ha observado actividades de explotación de mineral no metálico como la caliza.

El área, es vulnerable a cambios en los patrones climáticos y presenta alta fragilidad de ecosistemas es sus praderas altoandinas, como desaparición de flora y fauna silvestre, disminución de recursos hídricos, pérdida de rendimientos de la producción y aparición de plagas y enfermedades. Es susceptible a riesgos de desastres, como inundaciones, derrumbes, huaicos y hundimientos [de tierras] (GRPPAT, 2016).

3.2. Materiales y equipos

3.2.1. Materiales

Ficha de inventario de campo, útiles de escritorio, modelo digital de elevación MDE Alos Palsar (escena: Path 110- Frame 6980, re: 12.5 m, fecha: 02/18/11), mosaico satelital mensual ortorectificado de 4 bandas (RGB y Nir) PlanetScope (Quads: 587-968 y 587-967, re: 4.68 m, fecha: 02/21) de Planet Labs Inc., mosaico satelital Google-Satellite, software Microsoft Excel, software QGis 3.12.2, software SASPlanet, software R y Rstudio versión 4.1.0.

3.2.2. Equipos

Laptop Lenovo ideapad z500 core I5, GPS Etrex 30x, Cámara Fotográfica Panasonic DMC-FH8.

3.3. Metodología

3.3.1. Tipo de investigación

Tipo de investigación es: Aplicada – No experimental.

Alcance de la investigación es: Descriptivo – correlacional.

Unidad de análisis es: el distrito de Sillapata.

3.3.2. Flujograma general del proceso

El orden de pasos seguido durante todo el proceso investigativo se muestra en el siguiente flujograma:

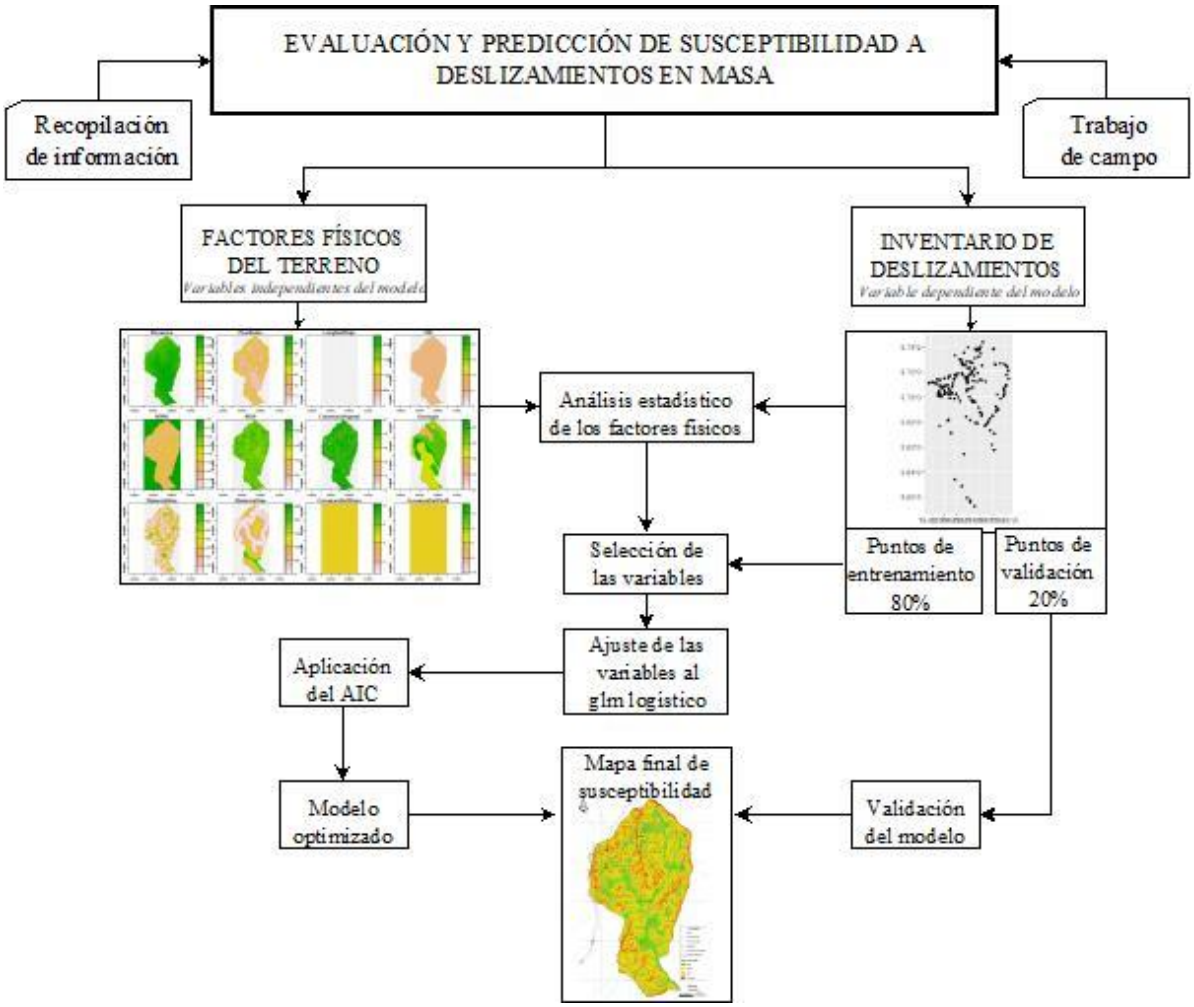


Figura 8. Flujograma del proceso de tesis.

3.3.3. Recopilación de información

Las entidades del estado, como parte de su gestión local, almacenan información sobre alguno de los deslizamientos ocurrido en el distrito que hayan causado afectaciones, por ello se solicitó dicha información a la Secretaría de Defensa Civil de la Municipalidad Distrital de Sillapata, Instituto Vial Provincial de Dos de Mayo. Así mismo, se revisó la base de datos del SINPAD de Defensa Civil Nacional y Gobierno Regional Huánuco.

También se recopiló información geográfica y satelital que cubriera todo el área en estudio: una escena del modelo digital de elevación Alos Palsar (desde el portal ASF Data Search de la NASA: https://search.asf.alaska.edu/#/) cuyos datos fueron rellenados

posteriormente (función *fill*); dos quads del mosaico satelital PlanetScope (compuesto de los mejores píxeles diarios durante el periodo mensual de febrero de 2021) seleccionado y descargado desde el complemento Planet_Explorer de QGIS; Mosaico satelital Google-Satellite en alta definición; Servicio WMS del Mapa de Geología 1:100'000 – INGEMMET; base de datos vectoriales de las Rutas nacionales y vecinales descargados desde la plataforma de PROVIAS NACIONAL.

3.3.4. Inventario de puntos

Debido a la escasa información geográfica recopilada como información secundaria de deslizamientos, se realizó trabajo de campo y fotointerpretación satelital para identificar puntos deslizamiento y no deslizamiento. Los recorridos de campo se hicieron desde las tres cumbres más altas del distrito siguiendo rutas hacia las zonas más bajas. Se tomaron las coordenadas de deslizamientos activos y pasados en las coronas del deslizamiento, pero en caso estos fueran inaccesibles, se optó por tomar el pie del deslizamiento. Alguno de los puntos de "no deslizamiento" se tomaron en campo y otros con apoyo del mosaico Google-Satellite de alta definición. Los puntos de campo se plasmaron en fichas de inventario de deslizamientos.

El inventario final con todas las coordenadas (66 puntos de deslizamiento y 110 puntos de no deslizamiento) se almacenó en formato .csv. Posteriormente se proyectaron al sistema UTM 18S datum WGS 84 y codificaron en sistema binario (1 = Deslizamiento y 0 = No deslizamiento), usando el software R con el script 1 del **anexo F**.

3.3.5. Mapeo de factores físicos del terreno

La mayor parte de los factores físicos a utilizar en el modelo estadístico de deslizamientos de masa, son un sustituto de los parámetros físicos que influyen en la materialización de estos fenómenos cuya estimación a gran escala no es factible (Pourghasemi et al., 2021). Debido a que no hay directrices estandarizadas para elegir los factores que controlan la ocurrencia de deslizamientos, en este trabajo la selección se basó en la revisión de literatura, la visita de campo y los datos disponibles.

Usando sistemas de información geográfica se procesaron y georeferenciaron (en proyección UTM 18S y datum WGS84) la información geográfica recopilada, generando los 12 factores físicos escogidos (**tabla 6**), los cuales se mapearon en formato ráster, alinearon a tamaño de píxel 4,68m y recortaron de acuerdo al polígono del distrito Sillapata (extraído del IGN), resultando en mapas individuales que se muestran en el **anexo G**.

Tabla 6. Fuente y resolución espacial de las variables utilizados

Variable	Origen	Resolución / escala
Elevación	MDE ALOS PALSAR	12,5m x 12,5m
Pendiente (grados)	MDE ALOS PALSAR	12,5m x 12,5m
Longitud de flujo	MDE ALOS PALSAR	12,5m x 12,5m
TWI	MDE ALOS PALSAR	12,5m x 12,5m
NDWI	Mosaico PlanetScope	4,68m x 4,68m
NDVI	Mosaico PlanetScope	4,68m x 4,68m
Cobertura vegetal	Mosaico PlanetScope	4,68m x 4,68m
Geología	Mapa geológico - INGEMMET	1: 100 000
Distancia a flujos de agua	MDE ALOS PALSAR	12,5m x 12,5m
Distancia a vías	Red de vías - PROVIAS	No tiene
Curvatura plana	MDE ALOS PALSAR	12,5m x 12,5m
Curvatura de perfil	MDE ALOS PALSAR	12,5m x 12,5m

Ver distribución y frecuencia de cada factor o variable en el **anexo G**.

Las variables Elevación, Pendiente, Longitud de Flujo, TWI, Distancia a flujos de agua, Curvaturas, NDWI y NDVI se generaron haciendo uso de técnicas de teledetección de modelos digitales de elevación e imágenes satelitales, con el apoyo de herramientas SIG.

La variable Cobertura vegetal, a través del mapa de cobertura vegetal sirve para conocer el tipo de vegetación que crece sobre terreno a partir de las actividades económicas que se realizan en dicho lugar. Debido a la complejidad de su elaboración, se detallará a continuación.

Se recolectaron 149 puntos de las coberturas (en campo y mediante foto-interpretación de imágenes de alta resolución de Google-Satellite). Se realizó una clasificación píxel por píxel del mosaico satelital PlanetScope (del mes de febrero de 2021) en Python, mediante las librerías Gdal y Scikit Learn (Pedregosa et al., 2011), utilizando como datos de entrenamiento los puntos colectados, luego se aplicó el algoritmo de Máquina de Vectores de Soporte, cuyos hiperparámetros fueron optimizados mediante una grilla de búsqueda, y con el mejor modelo se realizó la predicción para toda la imagen, obteniendo una exactitud de 99,8%. En el área de estudio se identificaron y agruparon las coberturas vegetales en cinco categorías principales: Cultivos, Forestal, Humedales, Pastizales y Suelo desnudo (ver **tabla 7**).

Tabla 7. Coberturas vegetales presentes en el área de estudio

Nombre	Descripción*	Muestra fotográfica
Cultivos	Terrenos dedicados a producir alimentos, fibras y otras materias primas, relacionado a cultivos en rotación, descanso o barbecho. Los cultivos observados son: tubérculos (Papa, Olluco y Mashua), cereales (Maíz, Trigo, Cebada, Centeno y Avena), leguminosas (Habas, Alfalfa y Chocho), no se consideró áreas con pastos cultivados.	
Forestal	Referido a coberturas de tipo boscoso dispuesta de forma fragmentada, densa o abierta, incluyen plantaciones forestales y matorrales con vegetación secundaria o en transición. Las especies observadas en extensión son: Quinual, Quishuar, Sauco, Aliso, Eucalipto, Ciprés y Pino.	
Humedal	Lo constituyen superficies de agua (poca extensión) y coberturas húmedas, como terrenos anegadizos temporal o permanentemente inundados. En la zona se ha encontrado las categorías: Cuerpos de agua, Zonas pantanosas, Turberas y vegetación acuática sobre cuerpos de agua.	
Pastizal	Comprende cubiertas sembradas o naturales de pastizales, herbazales o pajonales. Se ha observado extensiones de pasto denso (limpios y enmalezados), pajonales andinos (Ichu y Cashaucsha) y otros herbazales andinos (Kunkuna, Champa, Ucush casha, Orejuelo, Valeriana, Condor cebolla, Crespillo, entre otros).	
Suelo desnudo	Corresponde a zonas urbanizadas, redes viales de comunicación, escombreras y zonas de disposición de residuos sólidos y también se ha relacionado a zonas erosionadas.	

* Descripción establecida con base en la clasificación hecha por IDEAM, 2010. Fuente: Trabajo de campo.

Las unidades geologías y características litológicas útiles para describir la variable Geología del modelo, se extrajeron del Mapa Geológico Nacional del INGEMMET, de esta manera se identificaron cinco unidades geológicas (ver **tabla 8**).

Tabla 8. Formaciones geológicas presentes en el área de estudio

Nombre	Formación	Litología	Edad
PE-cm	Complejo Marañón	Esquistos y gneis. Afloramiento de pizarras, esquistos y filitas de colores marrones a grises	Neoproterozoico
JTr-p	Grupo Pucara	Calizas gris azulinas en bancos medios a gruesos con nódulos de chert	Triásico - Jurásico
Ps-m	Grupo Mitu	Andesita púrpura y flujos de lava, andesita piroclástica, conglomerado de guijarros de andesita y arenisca roja en proporciones variables	Paleógeno temprano
Q-al	Depósitos Aluviales	Acumulación de grava, arena, limo y arcillas, con clastos subangulosos a angulosos de diferente composición	
Qp-lu	Formación La Unión	Conglomerados polimícticos gruesos y areniscas semiconsolidados en matriz arenosa	Pleistoceno

Fuente: INGEMMET, 1996.

Por último, la variable Distancia a vías se generó a partir de la base de datos espacial generada por PROVIAS, esta capa de vías fue actualizada con imagen de alta resolución de Google satellite. Con el uso de herramientas de SIG, se creó una capa de distancias euclidianas de vías a todo el distrito.

3.3.6. Extracción de información a puntos filtrados

Con el uso de herramientas SIG, las doce capas ráster de los factores físicos se combinaron en un solo archivo ráster llamado “12VARIABLES.tif”. Dicha información fue extraída en la nube de puntos del inventario, este procedimiento es posible hacerlo empleando el script 2 del **anexo F**, obteniendo una base de datos general que se usará para entrenamiento y validación del modelo final.

En la **tabla 9**, se muestra el resumen de las variables consideradas en la nueva base datos general: las independientes (variables físicas), la dependiente (susceptibilidad a deslizamiento) y la ubicación de cada punto.

Tabla 9. Características de las variables (dependientes e independientes) empleadas

Tipo variable		Variable	Unidad de medida	Área de estudio	Tipo	Rango de ocurrencia
VI	1-	Elevación	Metros	Geomorfología	Cuantitativa continua	3085msnm a 4140msnm
VI	2-	Grados de pendiente	Grados sexagesimales	Geomorfología	Cuantitativa continua	0° a 70,91°
VI	3-	Longitud de flujo	Metros	Hidrogeología	Cuantitativa continua	0 m a 23027,13 m
VI	4-	TWI	Porcentaje	Geomorfología	Cuantitativa continua	1,618% a 21,503%
VI	5-	NDWI	Adimensional	Uso del suelo	Cuantitativa continua	-0,828 a -0,096
VI	6-	NDVI	Adimensional	Uso del suelo	Cuantitativa continua	-0,302 a 0,991
VI	7-	Cobertura vegetal	Adimensional	Uso del suelo	Cualitativa categórica	1(Cultivos) 2(Forestal) 3(Humedal) 4(Pastizal) 5(Suelo desnudo)
VI	8-	Geología	Adimensional	Geología	Cualitativa categórica	2(Q-al) 3(Qp-lu) 5(JTr-p) 6(PE-cm1) 7(PE-cm2) 9(Ps-m)
VI	9-	Distancia flujo de agua	Metros	Hidrografía	Cuantitativa continua	0 m a 806,42 m
VI	10-	Distancia vías	Metros	Uso del suelo	Cuantitativa continua	0 m a 2610,82 m
VI	11-	Curvatura plana	Adimensional	Geomorfología	Cuantitativa continua	-12,504 a 9,695
VI	12-	Curvatura de perfil	Adimensional	Geomorfología	Cuantitativa continua	-13,073 a 5,805
VD		Susceptibilidad a deslizamientos	Adimensional	Gestión de riesgos	Cualitativa binomial	0 a 1,000

VI: Variable independiente. VD: Variable dependiente.

3.3.7. Exploración de los factores físicos

Para asegurar resultados óptimos, es necesario revisar, depurar y seleccionar las variables, por ello se realizaron análisis estadísticos del comportamiento de los 176 puntos y sus datos, esto es posible usando el script 3 del **anexo F**.

Para el caso de las 10 variables independientes cuantitativas se realizó un análisis de correlación, mientras las 2 variables cualitativas se analizaron usando gráfico de mosaicos. Este proceso descartó variables altamente correlacionadas a otras, por considerarse duplicadas.

3.3.8. Aplicación del modelo lineal generalizado

Como punto de partida, se dividió de forma aleatoria la base de datos general (variables independientes y dependiente) según el principio de Pareto, es decir el 80% de los puntos para entrenamiento y 20% para validación del modelo, esto es posible usando el script 4 del **anexo F**. Vale mencionar que, en este punto del proceso, se identificó un problema de desbalance debido a la minoritaria cantidad de clases que se quiere predecir (puntos de deslizamiento "1"), causando que el modelo no represente adecuadamente y las predicciones sean solo de la clase mayoritaria, resultando en altos valores de las métricas de evaluación (Ospina-Gutiérrez y Aristizábal, 2021).

El modelo lineal generalizado a aplicar se eligió a partir del comportamiento de las variables independientes (variables continuas y categóricas) y la variable dependiente (variable binomial), de esta manera, se usó la función glm() logística y la base de datos de entrenamiento, dando por resultado información sobre la aplicación glm a todas las variables (modelo full) y el valor AIC del ajuste de dicho modelo full. Se recomienda el uso del script 5 del **anexo F**.

3.3.9. Ajuste y simplificación del modelo

Se usó el Criterio de Información de Akaike con el propósito de simplificar el modelo y seleccionar la función que mejor se ajuste a los datos de campo, este modelo trabajará solo con las variables más significativas. Con la premisa de que a menor valor AIC mejor es el ajuste, se consigue el modelo final. Para el procedimiento se usó el script 6 del **anexo F**.

3.3.10. Análisis de la significancia de las variables finales

En este apartado se analizaron las variables del modelo final, con el propósito de conocer su relación con la susceptibilidad a deslizamientos de masa predichos. Este procedimiento es posible realizarlo haciendo uso del script 7 del **anexo F**.

3.3.11. Predicción del mapa de susceptibilidad

Con el modelo final se procedió a predecir la totalidad del territorio distrital, siendo el archivo final un ráster que se guardó con el nombre "Mapa_final.tif". El procedimiento se realizó empleando el script 8 del **anexo F**.

Posteriormente se almacenó y trasladó el archivo ráster al software Qgis (el cual contiene valores continuos de susceptibilidad), estos fueron clasificados por el método Natural breaks (Pourghasemi et al., 2021; Chen et al., 2019) en 4 niveles de susceptibilidad a deslizamientos en masa: bajo, medio, alto y muy alto.

3.3.12. Proceso de validación del modelo

Este apartado tuvo por finalidad medir la precisión y establecer una significancia científica sólida de los métodos empleados para el modelo.

Se analizaron solo los puntos de validación, que fueron el 20% de los puntos de la base de datos general (ver base de datos de validación en el **anexo D**) con el uso de la curva ROC. A través de dicha curva se determinó el mayor índice de Youden o "*best threshold*", que es el punto con sensibilidad y especificidad más altos, útil para conocer el umbral de exactitud diagnóstica del modelo y, el AUC – COR para conocer su capacidad discriminativa.

A la nube de puntos de validación, valores observados (binarios: 0 y 1), se agregó una columna con valores determinados usando el modelo final resultante (valores continuos entre 0 a 1). El archivo (en formato .csv) se guardará con el nombre "PUNTOS_ROC.csv", donde los puntos de inventario tendrán el encabezado "CAMPO" y los puntos del modelo tendrán "PREDICCION", a partir de estos se determinó la sensibilidad y especificidad por cada punto y se graficaron en un plano cartesiano. Es posible hacer la validación con el uso del script 9 del **anexo F**.

IV. RESULTADOS Y DISCUSIÓN

4.1. Exploración de los factores físicos

Para reforzar la capacidad de predicción de cualquier algoritmo de modelamiento, es necesaria la selección precisa de factores significativos y la depuración de factores externos (Pourghasemi et al., 2021), por ello, se extrajeron del inventario de puntos los valores de las variables cuantitativas y cualitativas (**anexo B**) para identificar aquellos factores que presenten mayor influencia.

A continuación, se muestra el análisis de correlación de los datos correspondientes a las variables de tipo cuantitativo:

Tabla 10. Matriz de correlaciones de las variables cuantitativas

	Elevac.	**Pend.**	**Long. Flujo**	**TWI**	**NDWI**	**NDVI**	**Dist. Ríos**	**Dist. Vías**	**Curv. plana**	**Curv. perfil**
Elevac.	**1,000**	-0,126	-0,035	-0,015	0,108	-0,145	0,405	0,207	0,028	-0,053
Pend.	-0,126	**1,000**	-0,050	-0,225	0,118	-0,097	-0,111	-0,047	-0,006	-0,001
Long. Flujo	-0,035	-0,050	**1,000**	**0,855**	0,072	-0,073	-0,178	0,044	-0,368	0,011
TWI	-0,015	-0,225	**0,855**	**1,000**	0,029	-0,025	-0,165	0,048	**-0,502**	0,159
NDWI	0,108	0,118	0,072	0,029	**1,000**	**-0,976**	0,095	-0,061	0,091	-0,070
NDVI	-0,145	-0,097	-0,073	-0,025	**-0,976**	**1,000**	-0,119	0,095	-0,045	0,079
Dist. Ríos	0,405	-0,111	-0,178	-0,165	0,095	-0,119	**1,000**	-0,053	0,149	-0,186
Dist. Vías	0,207	-0,047	0,044	0,048	-0,061	0,095	-0,053	**1,000**	0,110	-0,051
Curv. plana	0,028	-0,006	-0,368	**-0,502**	0,091	-0,045	0,149	0,110	**1,000**	-0,476
Curv. perfil	-0,053	-0,001	0,011	0,159	-0,070	0,079	-0,186	-0,051	-0,476	**1,000**

Existen algunas variables que presentan altas correlaciones (mayores a ±0,5), es decir son variables redundantes, por ello se analizaron por pares cuales de las variables conservar y cuales descartar.

Tabla 11. Variables con correlaciones altas

Variables	**Correlación**
Longitud de flujo - TWI	0,855
Curvatura plana - TWI	- 0,502
NDWI - NDVI	- 0,976

Debido a que la correlación entre los pares ‘Longitud de flujo – TWI’ y ‘Curvatura plana – TWI’ no superan el valor |0,900| se conservarán para ser analizadas en procesos posteriores del GLM. El par ‘NDWI – NDVI’ presenta correlación cercana a 1,000, siendo necesaria su revisión, la cual muestra que el NDVI tiene antecedentes como variable influyente en trabajos similares (Mohamed y Pourghasemi, 2020; Chen et al., 2019), mientras que el NDWI usado en el presente trabajo es una adaptación de la ecuación original (Gao, 1996; McFeeters, 1996), por lo que está supeditada a variaciones en la representación que se pretendía dar en sus inicios. De esta manera, será descartada la variable NDWI.

También, se realizó un análisis exploratorio de las variables cualitativas a través del análisis del gráfico de mosaicos que se muestra en la **figura 9**.

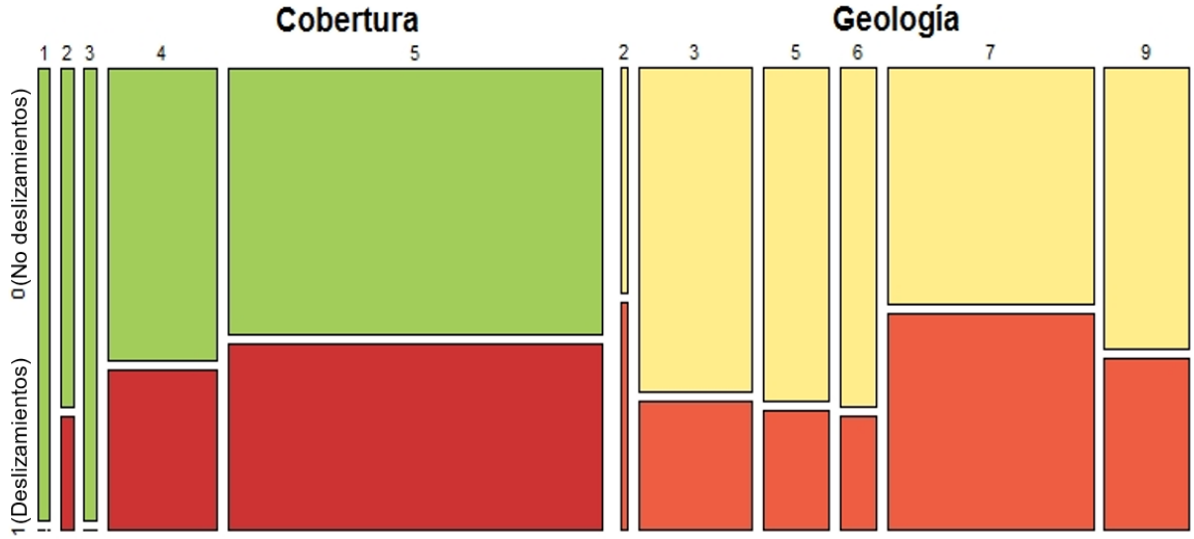


Figura 9. Gráfico de mosaicos de las variables de tipo cualitativo.

De acuerdo con la **figura 9**: el factor cobertura vegetal, con cinco categorías de cobertura, muestra concentración principalmente en la cobertura 5 (suelos desnudos), y en menor grado en la cobertura 4 (pastizal), en todas las categorías se muestran mayor cantidad de puntos estables, tal es así que, en la categoría 1 (cultivos) y 3 (humedales) hay ausencia de puntos de deslizamiento, mientras que las categorías 2 (forestal), 4 y 5 presentan puntos de deslizamientos distribuidos de forma creciente. En el factor geología, con seis categorías, la clase 7 (PE-cm2) es la que concentra mayor cantidad de puntos y en esta categoría los puntos de deslizamiento y no deslizamiento presentan equidad, en orden de cantidad de puntos muestreados le siguen la clase 3 (Qp-lu), la clase 9 (Ps-m), la clase 5 (JTr-p), 6 (PE-cm1) y 2 (Q-al), todas estas muestran una ligera concentración de los puntos en las zonas estables.

La distribución de datos cualitativos en el inventario y la concentración de puntos de deslizamiento y no deslizamiento son asimétricas, sumando a esto, la diversidad de fuentes de información, hacen conveniente conservar ambos factores.

A continuación, se muestran las variables más influyentes que fueron usadas para los procesos posteriores.

Tabla 12. Factores físicos más influyentes para ser usados en el GLM

Factor físico seleccionado	Denominación en el modelo*
Elevación	Elevacion
Pendiente	Pendiente
Longitud de flujo	LongFlujo
TWI	TWI
NDVI	NDVI
Cobertura vegetal	CoberturaVegetal
Geología	Geologia
Distancia a flujos de agua	DistanciaRios
Distancia a vías	DistanciaVias
Curvatura plana	CurvaturaPlana
Curvatura de perfil	CurvaturaPerfil

* Se usó para construir los scripts en el software RStudio que nos faciliten la manipulación de dichas variables y del modelo.

4.2. Desarrollo del modelo de susceptibilidad a deslizamientos de masa

El modelo lineal generalizado más adecuado para modelar la susceptibilidad de acuerdo con los tipos de variables de la investigación se muestra en la siguiente tabla.

Tabla 13. Selección del modelo lineal generalizado

Variable respuesta	Variables predictoras	Principal método recomendado*	Función de vínculo
Nominal / Binaria	Categóricas y continuas	Logística	$\log\left(\frac{\mu}{n-\mu}\right)$

* McCullagh y Nelder (1989), Durbán (2014) y Cayuela (2011).

4.2.1. Selección de variables y modelo

Inicialmente, se procesa el modelo lineal generalizado incluyendo todas las variables (modelo full), posteriormente, se hace el ajuste a través del análisis de devianza y el criterio de Información de Akaike (para mayor detalle, ver **anexo C**). A continuación, se muestra las variables seleccionadas para el modelo lineal generalizado final:

Tabla 14. Resumen del proceso de selección de variables para el modelo final

Variable	AIC	Modelo final	Coeficiente en el modelo
Todas	194.31	No	-
Cobertura vegetal	188.12	No	-
Geología	185.47	No	-
Curvatura plana	183.49	No	-
Distancia a vías	181.63	No	-
Longitud de flujo	179.82	No	-
Elevación	178.75	No	-
Curvatura de perfil	179.34	Sí	0.4325
Distancia a flujos de agua	180.57	Sí	-0.0029
Pendiente	181.62	Sí	0.0571
Índice diferencial de vegetación normalizado - NDVI	182.13	Sí	-4.8733
Índice de Humedad topográfico - TWI	183.61	Sí	0.2951
Independiente	-	Sí	0.4199

Procedimiento completo, ver anexo C.

El valor AIC es una métrica relativa para conocer el modelo que mejor se ajusta a los datos a través del mínimo AIC. Al descartar algunas variables de los modelos, los valores AIC alcanzaron los menores valores, pero en los modelos que excluyen las variables Curvatura de perfil, Distancia a flujos de agua, Pendiente, NDVI y TWI (posteriores a Longitud de flujo en la **tabla 14**), los valores AIC muestran incremento, por ello dichas variables no se excluyeron. Así, el modelo final quedó constituido por las cinco variables mencionadas y presenta la siguiente ecuación:

$$y = \frac{1}{1+e^{-(0.420 + 0.432X_1 - 0.003X_2 + 0.057X_3 - 4.873X_4 + 0.295X_5)}}$$

Donde:

y es el valor de susceptibilidad a deslizamientos en masas, X_1 son los valores de curvatura de perfil, X_2 son los valores de distancia a flujos de agua, X_3 son los valores de pendiente, X_4 son los valores de NDVI y X_5 los valores de TWI.

Las variables y el modelo finales son un resultado singular para la zona de estudio, no debe ser generalizado a otros lugares, pues no se obtendrán siempre las mismas variables (Ospina-Gutiérrez y Aristizábal, 2021). A continuación, se muestra el mapa de todo el territorio distrital, cuya predicción se basó en el modelo final.

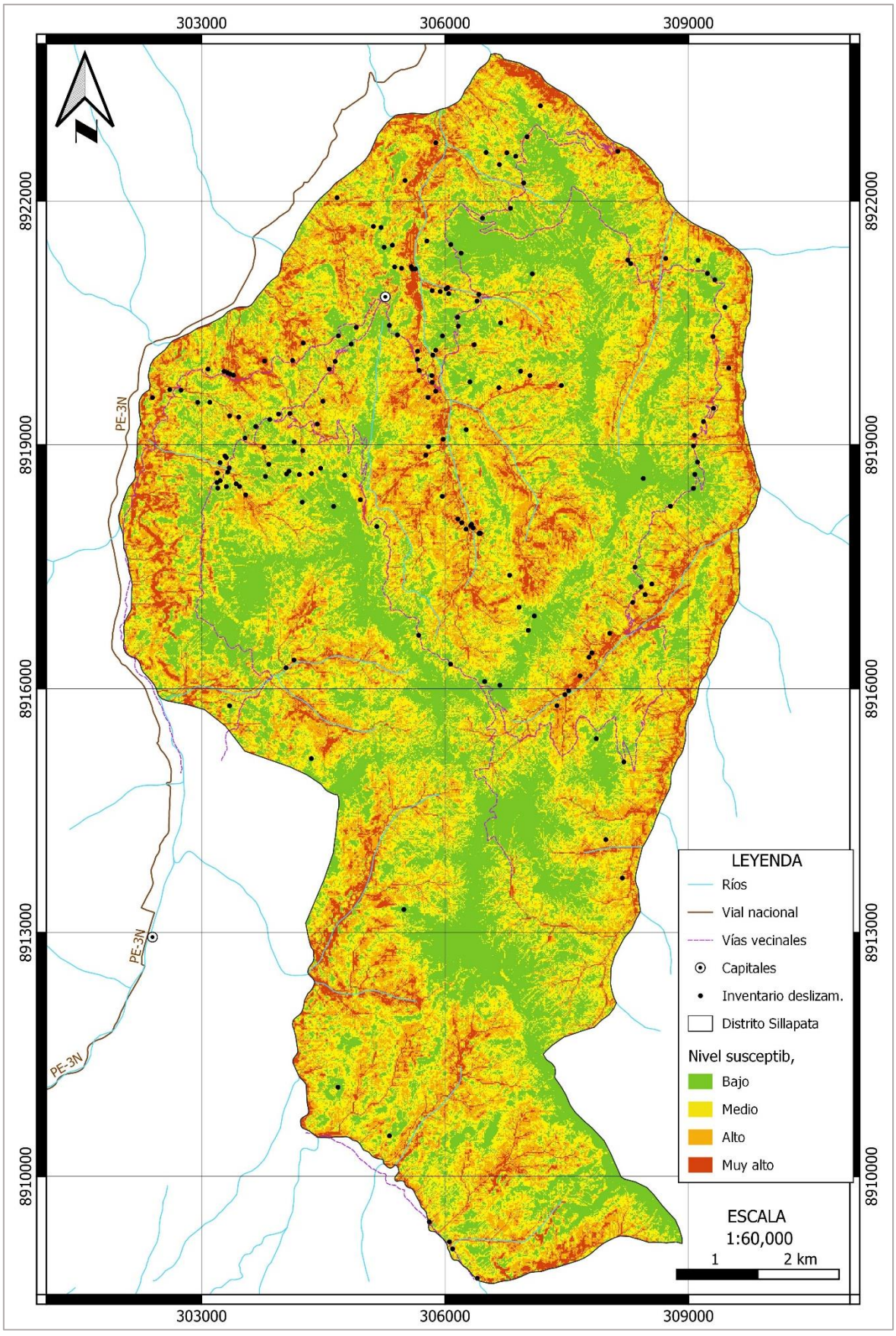


Figura 10. Mapa de susceptibilidad a deslizamientos en masa del distrito de Sillapata, usando el modelo lineal generalizado logístico final.

El ráster del modelo final de susceptibilidad a deslizamientos de masas está compuesto por 3’325’477 píxeles en la zona de estudio, cuyos valores fluctúan entre 0,0035 a 0,9999; haciendo uso de los SIG, estos valores se categorizaron con el método *Natural breaks*, como se muestra en la **figura 11**, obteniendo cuatro categorías del terreno: niveles bajos entre 0,0035 a 0,2184, niveles medios entre 0,2184 a 0,3864, niveles altos entre 0,3864 – 0,5936 y niveles muy altos entre 0,5936 – 0,9999.

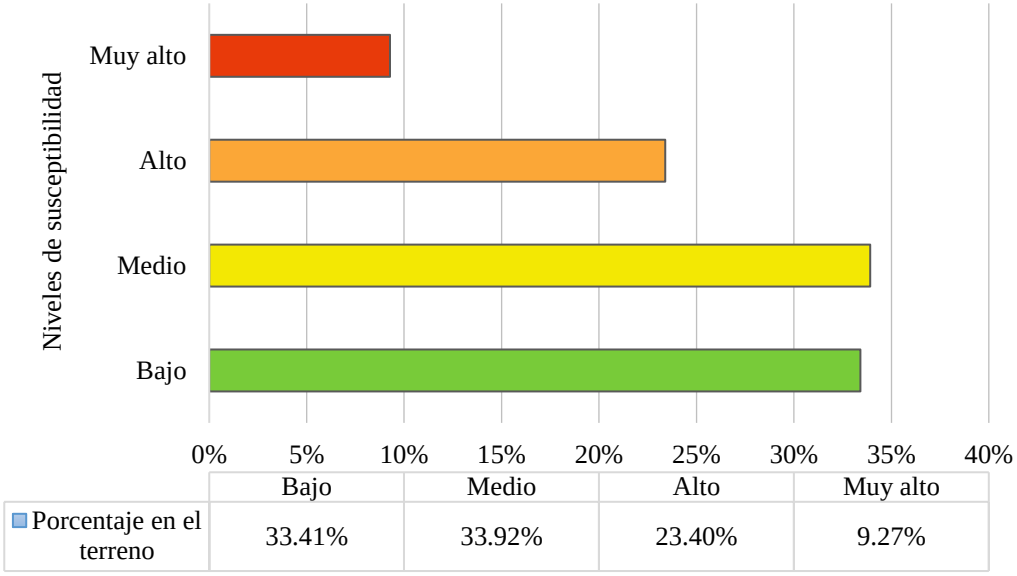


Figura 11. Porcentaje de los niveles de susceptibilidad a deslizamientos de masas obtenidos para el distrito de Sillapata.

Con el modelo de susceptibilidad a deslizamientos de masas aplicado al territorio de Sillapata se puede observar que la mayor proporción de dicho territorio presenta nivel de susceptibilidad media (33.92% con 1127853 pixeles) y la menor proporción es el nivel de susceptibilidad muy alta (9.27% con 308421 pixeles).

4.2.2. Validación del modelo

El análisis de la curva ROC del modelo determinó que el mejor umbral o índice de Youden es el punto con valor 0,422 (con un intervalo de confianza de 84% a 80%), esto significa que bajo este valor está la fracción más alta de zonas propensas a la estabilidad, que a la vez, serán correctamente identificadas con el modelo, y viceversa, sobre dicho umbral está la fracción más alta de zonas propensas a deslizamientos en masa (con presencia de susceptibilidad a deslizamientos) determinadas como tal por el modelo.

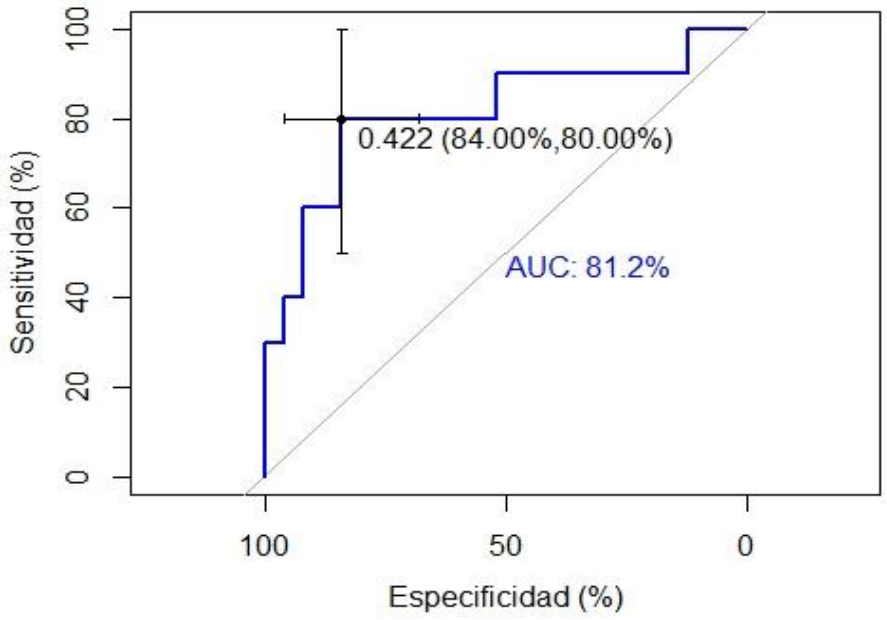


Figura 12. Curva ROC para validación del modelo de susceptibilidad a deslizamientos de masa.

El área debajo de la curva ROC (AUC-ROC) del modelo final determinó un 81,2% de capacidad de discriminar zonas susceptibles y no susceptibles a deslizamientos de masa, en comparación con otras experiencias, dicho valor muestra capacidad inferior al modelo de Pourghasemi et al. (2021) que usando GLM obtuvieron un valor AUC de 90,8%, pero a su vez es superior al modelo de Chen et al. (2019) que obtuvieron 71,8% en su modelo GLM, habiendo trabajado ambos con menor cantidad de puntos de entrenamiento (70%), lo cual evidencia que la proporción de puntos para entrenamiento y validación no influyen para generar un modelo más válido que otro.

El valor AUC-ROC resultante es posible mejorarlo a través del perfeccionamiento de ciertos procesos, por ejemplo, los factores físicos de las variables independientes deberían ser de alta y uniforme resolución desde el principio, la baja cantidad de puntos inventariados podrían ser sometidas a remuestreo de puntos (Hassangavyar et al., 2020) u optimización de hiperparámetros (Sun et al., 2020), que son técnicas nuevas que están comenzando a ser aplicadas en este tipo de modelamientos.

4.3. Determinación de la relación entre factores físicos y el modelo

Una vez determinado el modelo de susceptibilidad a deslizamientos de masas, es importante conocer el tipo de relación que guarda con las variables independientes que formaron parte del modelo final, lo cual será de utilidad como enfoque para la gestión prospectiva de riesgos, en vías de disminuir la susceptibilidad a deslizamientos de masas.

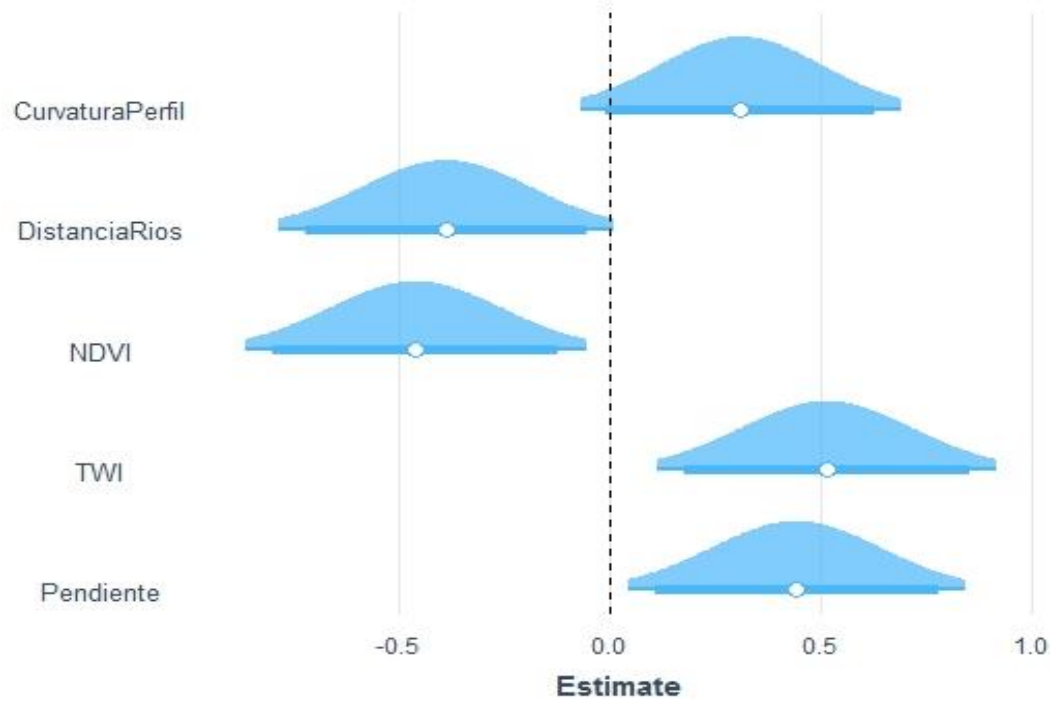


Figura 13. Gráfico de regresión de las variables del modelo final

La **figura 13** muestra que las variables del modelo final están relacionadas con la variable respuesta de distintas maneras. El NDVI y Distancia a flujos de agua, mostraron una asociación negativa con la susceptibilidad a deslizamientos de masa, significando que a valores más bajos de NDVI o a menores distancias de cuerpos de agua, existe mayor probabilidad espacial de que ocurran deslizamientos. En el caso de las variables Curvatura de perfil, Pendiente y TWI, la asociación fue positiva, es decir, para zonas con mayores valores de curvatura de perfil, grados de pendiente o valores de humedad topográfica del terreno, la probabilidad espacial de ocurrencia de deslizamientos también es mayor.

Los valores de NDVI en el terreno varían de -0,302 a 0,991, los valores negativos representan estructuras artificiales (zonas de viviendas); los valores cercanos a 0, suelos desnudos, superficies de agua y rocas; mientras los valores cercanos a 1, zonas con vegetación. La asociación negativa expuesta en este trabajo, prueba que las superficies con más vegetación corresponden a zonas con menor susceptibilidad de deslizamientos en masa y viceversa, pues Suarez (2009) menciona que la vegetación ralentiza la velocidad de infiltración y refuerza la estabilidad del suelo con sus raíces y la materia orgánica que genera; sin embargo, su influencia está sujeta a variaciones por otros factores como precipitación, pendiente y profundidad del deslizamiento, y para comprender a cabalidad la influencia de la vegetación sobre el peligro de

deslizamientos de masas en la zona de estudio, se deberá estudiar el componente temporal, es decir, el comportamiento de la precipitación, principalmente los meses de lluvias intensas.

Los valores más altos de acumulación de humedad del TWI se ubican en zonas con depresiones de drenaje y los valores más bajos de TWI coinciden con las formaciones montañosas (mayor pendiente), siendo la parte más alta donde se acumula la menor cantidad de agua. De acuerdo con la asociación positiva, aquellas zonas con mayor acumulación de humedad (depresiones y riberas de cauces) mostraron mayor susceptibilidad a deslizamientos de masa del terreno, contradiciendo la asociación inversa expuesta por Meinhardt et al. (2015), quien además explica que la influencia de factores climáticos puede producir un fuerte aumento de la presión del agua intersticial y desestabilizar las pendientes (menor TWI), lo cual incentiva la necesidad de estudiar el componente temporal (variación de parámetros climáticos y de la misma variable TWI) para conocer mejor el peligro de deslizamientos de masas. En general, se debe considerar la estacionalidad al momento del recojo de datos, para evitar variaciones por efecto de aquellas variables cuyo rango están sujetas a las variaciones temporales.

El modelo final descartó las dos variables categóricas (geología y cobertura de suelo), mientras Chen et al. (2019) determinaron la alta influencia de variables como uso de suelos y litología, con fuentes de información análogas a las variables del presente trabajo, a diferencia de la escala de adquisición de la geología usada en la presente y la litología usada en mencionada investigación (1:100 000 y 1:1 000 000 respectivamente); dado que los modelos de los antecedentes dependen de este tipo de variables categóricas, la resolución espacial de las variables iniciales y la escala de análisis deberían ser consideradas con cuidado (Mohamed y Pourghasemi, 2020) pudiendo ser fuente de error al omitir variables con potencial de influencia.

Se compararon los puntos de campo inventariados en la etapa de campo de la presente investigación con el mapa de susceptibilidad a movimientos en masa a escala regional elaborado por el CENEPRED y publicado en el SIGRID (ver: https://sigrid.cenepred.gob.pe/sigridv3/mapa), en dicho mapa se observaron zonas sin datos que abarcaron el 9% de los puntos con deslizamientos de masas confirmados por el inventario, mientras el 83% de puntos con deslizamientos correspondió a zonas con alta y muy alta susceptibilidad. En cuanto a la comparación visual de ambos mapas, se diferencian dos regiones de susceptibilidades en el mapa del CENEPRED, y en el mapa de la presente investigación la susceptibilidad presentó más variedad de matices debido a la escala de trabajo y a medida que las variables cambiaron de valor.

Finalmente, la diversidad de relaciones de tipo ‘variable independiente – modelo’ en los modelos de susceptibilidad a deslizamientos en masa de distintos escenarios montañosos del planeta (Pourghasemi et al., 2021; Londoño-Linares, 2017; Mohamed et al., 2020; Ospina, y Aristizábal, 2021; Chen et al., 2019) evidencian que estos no siempre son válidos para fines de planificación y evaluación en espacios muy locales, si no se trabajan teniendo en cuenta uniformidad de escalas entre variables y territorio (Mohamed y Pourghasemi, 2020; Lee et al., 2017), por lo que será necesaria la elección de las variables independientes basados en observaciones en campo.

V. CONCLUSIONES

De acuerdo con los análisis estadísticos, cinco factores físicos intervinieron en el modelo final de susceptibilidad a deslizamientos de masas en el distrito de Sillapata, mismos que fueron evaluados posteriormente.

Los factores físicos que intervinieron en el modelo final de susceptibilidad a deslizamientos de masa del área de estudio fueron: grados de pendiente, índice de humedad del terreno TWI, índice diferencial de vegetación normalizado NDVI, distancia a ríos y curvatura de perfil del terreno.

El modelo lineal generalizado más adecuado, de acuerdo con los tipos de variables, fue la función de enlace logística, logrando una satisfactoria capacidad de discriminar con AUC-ROC de 81.2 %, la cual representó con susceptibilidad muy alta el 9,27%, con susceptibilidad alta el 23,4%, con susceptibilidad media el 33,92% y con susceptibilidad baja el 33,41% del territorio estudiado.

Los factores físicos NDVI y distancia a ríos presentaron asociación negativa con el modelo de susceptibilidad, pues mientras menores fueron sus valores, mayor fue la susceptibilidad a deslizamientos de masas; por su parte, los factores curvatura de perfil, grados pendiente y TWI, presentaron asociación positiva, pues mientras mayores fueron sus valores, mayor fue la susceptibilidad a deslizamientos de masas.

VI. PROPUESTAS A FUTURO

Mejorar la precisión y temporalidad de los modelos digitales de elevación, haciendo uso de mejores tecnologías satelitales y de radar, con el fin de tener un insumo con una superficie de trabajo más detallada y actualizada, donde incluso se pueda detectar los fenómenos in situ, para ser utilizadas en distintos modelos de la susceptibilidad.

Aumentar el nivel de detalle y actualizar la cartografía de unidades geológicas y geomorfológicas existentes en los distritos o cuencas que presenten antecedentes de deslizamiento de masas, con la intención de caracterizar los procesos activos, lo cual permitirá aplicar a diversos modelamientos.

Realizar inventarios de deslizamientos detallados en espacio y tiempo, donde figuren datos útiles como las coordenadas del evento, tipo de movimientos de masa identificado (o en su defecto detalles fotográficos) y la fecha en las que ocurren. Esta mejora de la calidad y resolución de información, incrementaría el acceso a los sistemas de registro de deslizamientos de masa y mejoraría los estudios de predicción de los diversos tipos de movimientos en masa (sean deslizamientos, huaycos, caída, reptación, etc.). En los distritos o cuencas que presenten altas frecuencias de deslizamientos al año, esta actividad debería ser obligatoria para contar con instrumentos de planificación de ordenamiento del territorio.

En aquellas variables influenciadas por la estacionalidad temporal (factores climatológicos) como el TWI, NDVI, entre otros, la toma de datos debe realizarse en los peores escenarios (de mayor intensidad), esto mejorará no solamente la evaluación de la susceptibilidad, sino también los pronósticos de peligros y riesgos por deslizamientos en masa.

VII. REFERENCIAS BIBLIOGRÁFICAS

Beven, K. y Kirkby, M. (1979). A physically based, variable contributing area modelo f basin hydrology. Hydrological Sciences Bulletin, 24(1), 43-69.

Bonachea, J. (2006). Desarrollo, aplicación y validación de procedimientos y modelos para la evaluación de amenazas, vulnerabilidad y riesgo debido a procesos geomorfológicos [Tesis Doctoral, Universidad de Cantabria]. Repositorio institucional TDX. http://www.tesisenred.net/TDR-1124106-134112

Cayuela, L. (2011). Modelos lineales generalizados. Curso de análisis de datos ecológicos en R. https://dl.orangedox.com/z1fvF8dtPWDB2DhIEo/3-Modelos%20lineales%20generalizados.pdf

Centro Nacional de Estimación, Prevención y Reducción del Riesgo de Desastres – CENEPRED (2014). Manual para la Evaluación de Riesgos Originados por Fenómenos Naturales – 2° versión. Lima, Perú. Biblioteca Nacional del Perú. 248 p.

Cerda, J. y Cifuentes, L. (2011). Uso de curvas ROC en investigación clínica. Aspectos teóricos – prácticos. Revista Chilena de Infectología, 29(2), 138-141.

Chen, W., Pourghasemi, H., Zhang, S. y Wang, J. (2019). A comparative study of functional Data Analysis and Generalized Linear Model Data – Mining Techniques for Landslide Spatial Modeling. Elsevier. https://doi.org/10.1016/B978-0-12-815226-3.00021-1

Concejero, P. (2009). Comparación de modelos de curvas ROC para la evaluación de procedimientos estadísticos de predicción en investigación de mercados [Tesis Doctoral, Universidad Complutense de Madrid]. Repositorio Wikidot. http://concejero.wikidot.com/tesis

CRED/ UC Louvain (2020). EM-DAT Public, the International Disaster Database. https://public.emdat.be/mapping

Cubito, A., Ferrara, V. y Pappalardo, G. (2005). Landslide hazard in the Nebrodi Mountains (Northeastern Sicily). Elsevier, Geomorphology, 66, 39-379. doi:10.1016/j.geomorph.2004.09.020

Gao, B. (1996). NDWI – A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water From Space. Elsevier, 58, 257-266. https://doi.org/10.1016/S0034-4257(96)00067-3

Gerencia Regional de Planeamiento, Presupuesto y Acondicionamiento Territorial - GRPPAT Gobierno Regional Huánuco (2016). Plan de desarrollo regional concertado "Huánuco al 2021). Huánuco, Perú.

Guillard, C. y Zezere, J. (2012). Landslide susceptibility assessment and validation in the framework of municipal planning in Portugal: the case of Loures Municipality, Environ. Manage, 50, 721-735.

Hassangavyar, M., Damaneh, H., Pham, Q., Thuy Linh, N., Tiefenbacher, J. y Bach, Q. (2020). Evaluation of re-sampling methods on performance of machine learning models to predict landslide susceptibility. Revista Geocarto International. https://doi.org/10.1080/10106049.2020.1837257

Hess, D., Leshchinsky, B., Bunn, M., Mason, H. y Olsen, M. (2017). A simplified three-dimensional shallow landslide susceptibility framework considering topography and seismicity. Springer – Verlag. Doi 10.1007/s10346-017-0810-2

Instituto de Hidrología, Meteorología y Estudios Ambientales - IDEAM (2010). Leyenda Nacional de Coberturas de la Tierra. Metodología CORINE Land Cover adaptada para Colombia Escala 1:100000. Instituto de Hidrología, Meteorología y Estudios Ambientales. Bogotá, Colombia, 72 p.

Instituto Geológico, Minero y Metalúrgico - INGEMMET (1996). Mapa Geomorfológico del Perú a escala 1:100000. Lima, Perú.

Instituto Geológico, Minero y Metalúrgico - INGEMMET (2019). Informe técnico N° A6915. Evaluación geológica del sector de Sillapata - Primer reporte. Lima, Perú.

Instituto Nacional de Defensa Civil - INDECI (2018). Orientar y capacitar para un país preparado. https://www.indeci.gob.pe/preparacion/peligros/

Lacasse, F., Nadim, F. y Kalsnes, B. (2010). Living with Landslide Risk. Geotechnical Engineering Journal of the SEAGS & AGSSEA, 41 (4). https://www.researchgate.net/publication/267423148

Landinez, A. y Beltrán, D. (2019). Modelo de susceptibilidad a deslizamientos del departamento de Cundinamarca Colombia, fundamentado en las características topográficas del terreno [Tesis de grado, Universidad de La Salle]. Repositorio institucional. https://ciencia.lasalle.edu.co/ing_civil/545

Lee, S., Hong, S., Jung, H. (2017). A support vector machine for landslide susceptibility mapping in Gangwon Province, Korea. Sustainability, 9 (48). doi:10.3390/su9010048

Londoño-Linares, J. P. (2017). Cálculo de susceptibilidad a deslizamientos mediante análisis discriminante. Aplicación a escala regional. Revista Dyna, 84 (201), 278 – 289. https://doi.org/10.15446/dyna.v84n201.61385

Marquéz, M. (2002). Modelo setar aplicado a la volatilidad de la rentabilidad de las acciones: algoritmos para su identificación [Tesis Doctoral, Universitat Politècnica de Catalunya]. Repositorio institucional TDX. https://www.tesisenred.net/handle/10803/6503

Martínez, D., Albín, J., Cabaleiro, J., Pena, T., Rivera, F. y Blanco, V. (2009). El Criterio de Información de Akaike en la obtención de modelos estadísticos de rendimiento. [Conferencia]: XX Jornadas de Paralelismo, A Coruña, España.

McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open wáter features. International Journal of Remote Sensing, 17(7), 1425-1432. https://doi.org/10.1080/01431169608948714

Meinhardt, M., Fink, M. y Tunschel, H. (2014). Landslide susceptibility análisis in central Vietnam base don an incomplete landslide inventory: Comparison of a new method to calculate weighting factors by means of bivariate statistics. Revista Geomorphology. Doi 10.1016/j.geomorph.2014.12.042

Ministerio del Ambiente - MINAM (2015). Mapa Nacional de Cobertura Vegetal: memoria descriptiva. Lima, Perú. Biblioteca Nacional del Perú. 108 p.

Mohamed, A. y Pourghasemi, H. (2020). Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at Abha Basin, Asir Region, Saudi Arabia. Revista Geoscience Frontiers. https://doi.org/10.1016/j.gsf.2020.05.010

Montiel, K., Gouveia, E. y Montes, E. (2007). Influencia de la intervención antrópica en la ocurrencia de procesos de ladera. Microcuenca de la quebrada Ramos, Flanco Norandino venezolano. Terra Nueva Etapa, 23 (34), 35 – 69.

Moreno, H. A., Vélez, M. V., Montoya, J. D., & Rhenals, R. L. (2006). La lluvia y los deslizamientos de tierra en Antioquia: análisis de su ocurrencia en las escalas interanual, intraanual y diaria. Revista EIA, (5), 59-69. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1794-12372006000100005

Ospina-Gutiérrez, J. y Aristizábal, E. (2021). Aplicación de inteligencia artificial y técnicas de aprendizaje automático para la evaluación de la susceptibilidad por movimientos en masa. Revista mexicana de Ciencias Geológicas, 38(1), 43-54. http://dx.doi.org/10.22201/cgeo.20072902e.2021.1.1605

Pandey, V. (2015). Predictive landslide susceptibility assessment using the weight-of-evidence method in north-central Garhwal Himalaya. Indica Phys Geogr, 36(6), 510-536. https://doi.org/10.1080/02723646.2015.1116050

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830

Pita, S. y Pértegas, S. (2003). Pruebas diagnósticas. Unidad de Epidemiología Clínica y Bioestadística. Complejo Hospitalario-Universitario Juan Canalejo. Cad Aten Primaria, 10: 120-124.

Pourghasemi, H., Sadhasivam, N., Amiri, M., Eskandari, S. y Santosh, M. (2021). Landslide susceptibility assessment and mapping using state-of-the art machine learning techniques. Springer, Natural Hazards. https://doi.org/10.1007/s11069-021-04732-7

Presidencia del Consejo de Ministros (2014). DS N° 034-2014-PCM. Decreto Supremo que aprueba el Plan Nacional de Gestión del Riesgo de Desastres – PLANAGERD 2014 – 2021. Diario Oficial El Peruano. https://www.peru.gob.pe/normas/docs/DS_034_2014_PCM.pdf

Schachter, P. V. (2008). Evaluación de la susceptibilidad de remociones en masa en el sector nororiente de la cuenca de Santiago mediante métodos estadísticos multivariados [Tesis de pre-grado, Universidad de Chile].

Suárez, J. (2009). Deslizamientos: Análisis Geotécnico. Edit. Universidad Industrial de Santader UIS, 01, 582 pp. https://www.erosion.com.co/deslizamientos-tomo-i-analisis-geotecnico/

Sun, D., Wen, H., Wang, D. y Xu, J. (2020). A random forest modelo f landslide susceptibility mapping base don hyperparameter optimization using Bayes algorithm. Elsevier, Geomorphology. https://doi.org/10.1016/j.geomorph.2020.107201

Tien Bui, D., Tuan, T., Hoang, N., Thanh, N., Nguyen, D., Liem, N. y Pradhan, B. (2017). Spatial prediction of rainfall-indiced landslides for the Lao Cai area (Vietnam) sing a hybrid intelligent approach of least squares support vector machine inference model and artificial bee colony optimization. Landslides 14(2), 447-458.

UNISDR - Estrategia Internacional para la Reducción de Desastres de las Naciones Unidas (2009). Terminología sobre Reducción del Riesgo de Desastres. Ginebra, Suiza.

Wang, Y., Fang, Z., Wang, M., Peng, L. y Hong, H. (2020). Comparative study of landslide susceptibility mapping with different recurrent neural networks. Elsevier, Computer and Geosciences. https://doi.org/10.1016/j.cageo.2020.104445

ANEXOS

Anexo A. Inventario de puntos georreferenciados en etapa de campo

Nombre	Desliza-miento	X	Y	Z	Fecha	Hora
CMM1	SI	304205	8918634	3738	26/02/2021	11:58:30
CMM2	NO	304466	8918712	3791	26/02/2021	03:51:42
CMM3	SI	304087	8919386	3608	26/02/2021	07:57:25
CMM4	SI	304844	8920239	3514	26/02/2021	09:16:56
CMM5	SI	304425	8919255	3630	26/02/2021	09:40:26
CMM6	SI	304046	8918641	3695	26/02/2021	10:19:30
CMM7	SI	304247	8918930	3692	26/02/2021	11:26:11
CMM8	SI	304574	8919931	0	26/02/2021	12:40:26
CMM9	SI	303787	8918612	3656	26/02/2021	04:01:35
CMM10	SI	303189	8918537	3571	26/02/2021	04:44:11
CMM11	SI	303229	8918560	3569	26/02/2021	04:46:17
CMM12	SI	303321	8918665	3562	26/02/2021	04:52:11
CMM13	SI	303340	8918716	3563	26/02/2021	04:55:42
CMM14	SI	303287	8918863	3553	26/02/2021	05:04:11
CMM15	SI	303839	8919313	3578	26/02/2021	05:19:36
CMM16	SI	305674	8916657	3945	26/02/2021	10:17:27
CMM17	SI	306675	8916043	4015	26/02/2021	10:36:11
CMM18	SI	306437	8917912	3687	26/02/2021	12:58:33
CMM19	SI	306428	8917917	3677	26/02/2021	01:02:34
CMM20	SI	306417	8917910	3677	26/02/2021	01:04:31
CMM21	SI	306346	8917978	3648	26/02/2021	01:20:32
CMM22	SI	306315	8918007	3628	26/02/2021	01:28:06
CMM23	SI	306205	8918042	3623	26/02/2021	01:44:23
CMM24	SI	305762	8918873	3566	26/02/2021	02:39:49
CMM25	SI	305976	8919070	3511	26/02/2021	02:57:15
CMM26	SI	305884	8919660	3462	26/02/2021	03:17:47
CMM27	SI	305836	8919853	3438	26/02/2021	03:29:23
CMM28	SI	305680	8919914	3456	26/02/2021	03:44:23
CMM29	SI	305412	8920352	3456	26/02/2021	03:58:31
CMM30	SI	305313	8920469	3463	26/02/2021	04:04:30
CMM31	SI	303457	8919342	3514	27/02/2021	10:50:15
CMM32	SI	303345	8919357	3495	27/02/2021	10:54:44
CMM33	SI	302394	8919581	3185	27/02/2021	01:29:41
CMM34	SI	303276	8919905	3343	27/02/2021	01:59:53
CMM35	SI	303318	8919888	3345	27/02/2021	02:06:24
CMM36	SI	303351	8919871	3350	27/02/2021	02:16:40
CMM37	SI	303394	8919857	3347	27/02/2021	02:21:36
CMM38	SI	303775	8920032	3364	27/02/2021	02:40:41
CMM39	SI	307432	8919731	3769	27/02/2021	11:21:32
CMM40	SI	307045	8919853	3682	27/02/2021	11:41:20
CMM41	SI	306662	8919704	3597	27/02/2021	12:11:26

Nombre	Desliza-miento	X	Y	Z	Fecha	Hora
CMM42	SI	306012	8920919	3363	27/02/2021	01:48:09
CMM43	SI	306031	8920934	3363	27/02/2021	01:51:39
CMM44	SI	305940	8920885	3369	27/02/2021	02:04:34
CMM45	SI	305841	8920897	3369	27/02/2021	02:15:50
CMM46	SI	305633	8921164	3318	27/02/2021	02:39:31
CMM47	SI	305595	8921162	3302	27/02/2021	02:51:06
CMM48	SI	308202	8915099	3940	27/02/2021	09:25:48
CMM49	SI	308184	8913670	3841	27/02/2021	10:53:54
CMM50	SI	307523	8915975	3820	27/02/2021	12:28:56
CMM51	SI	307662	8916157	0	27/02/2021	12:44:34
CMM52	SI	307809	8916442	3813	27/02/2021	13:08:15
CMM53	SI	308028	8916681	3801	27/02/2021	13:22:17
CMM54 carretera	SI	308465	8917158	0	27/02/2021	13:41:49
CMM55	SI	308412	8917251	3793	27/02/2021	13:45:02
CMM56	SI	308547	8917288	0	28/02/2021	13:46:38
CMM57	SI	309110	8918788	3709	28/02/2021	14:44:49
CMM58	SI	309060	8918987	0	28/02/2021	14:52:00
CMM59 margen vía	SI	309072	8919119	3667	28/02/2021	14:55:12
CMM60	SI	309183	8919285	3667	28/02/2021	15:03:54
CMM61	SI	309300	8920332	3577	28/02/2021	15:28:11
CMM62	SI	308250	8921273	3536	28/02/2021	16:20:58
CMM63 Saporumi	SI	308128	8922611	3249	28/02/2021	10:48:26
CMM64	SI	306761	8922594	3372	28/02/2021	11:21:15
CMM65	SI	306969	8922221	3415	28/02/2021	12:34:56
CMM66	SI	306163	8920460	3458	28/02/2021	14:00:49
CMM67	SI	306092	8909108	3528	28/02/2021	07:45:58
SMM1	NO	304076	8918675	3695	28/02/2021	11:44:36
SMM2	NO	304240	8918296	3868	28/02/2021	02:11:53
SMM3	NO	304626	8918243	3972	28/02/2021	02:41:10
SMM4	NO	304763	8918624	3882	28/02/2021	03:26:20
SMM5	NO	304354	8918647	3773	28/02/2021	04:06:25
SMM6	NO	304142	8919036	3452	28/02/2021	07:49:26
SMM7	NO	303543	8918384	3643	28/02/2021	04:12:13
SMM8	NO	303467	8918490	3612	01/03/2021	04:20:38
SMM9	NO	303200	8918467	3606	01/03/2021	04:31:13
SMM10	NO	303303	8918843	3563	01/03/2021	05:00:26
SMM11	NO	303535	8919083	3566	01/03/2021	05:11:43
SMM12	NO	303667	8919226	3572	01/03/2021	05:15:33
SMM13	NO	303948	8919379	3569	01/03/2021	05:24:06
SMM14	NO	306487	8916088	4039	01/03/2021	10:26:58
SMM15	NO	307025	8916718	4056	01/03/2021	11:20:27
SMM16	NO	307100	8916895	4087	01/03/2021	11:27:47
SMM17	NO	306795	8917396	3912	01/03/2021	12:25:30
SMM18	NO	306157	8918088	3609	01/03/2021	01:51:27

Nombre	Desliza-miento	X	Y	Z	Fecha	Hora
SMM19	NO	305966	8918367	3630	01/03/2021	02:02:44
SMM20	NO	305795	8918978	3563	01/03/2021	02:47:41
SMM21	NO	305789	8919584	3473	01/03/2021	03:15:57
SMM22	NO	304494	8919536	3600	01/03/2021	09:30:58
SMM23	NO	303767	8918972	3608	01/03/2021	10:22:30
SMM24	NO	303101	8919523	3425	01/03/2021	11:15:26
SMM25	NO	302951	8919522	3388	01/03/2021	11:36:54
SMM26	NO	302746	8919673	3339	01/03/2021	12:48:17
SMM27	NO	303078	8919932	3321	01/03/2021	01:48:04
SMM28	NO	302607	8919679	3291	01/03/2021	01:06:36
SMM29	NO	304120	8920038	3427	01/03/2021	03:12:45
SMM30	NO	304253	8920253	3446	01/03/2021	03:24:51
SMM31	NO	304685	8920341	3457	02/03/2021	03:32:28
SMM32	NO	304905	8920445	3462	02/03/2021	03:47:18
SMM33	NO	308442	8918586	3959	02/03/2021	10:30:27
SMM34	NO	306260	8919187	3566	02/03/2021	12:35:49
SMM35	NO	306306	8919775	3549	02/03/2021	12:50:28
SMM36	NO	306359	8920231	3494	02/03/2021	01:03:41
SMM37	NO	306154	8920571	3478	02/03/2021	01:16:22
SMM38	NO	306045	8920860	3391	02/03/2021	01:34:55
SMM39	NO	305378	8921189	3418	02/03/2021	03:16:44
SMM40	NO	305351	8921456	3404	02/03/2021	03:32:35
SMM41	NO	305248	8921431	3401	02/03/2021	03:39:47
SMM42	NO	305211	8921674	3381	02/03/2021	03:53:15
SMM43	NO	305506	8922253		02/03/2021	
SMM44	NO	305885	8922717		02/03/2021	
SMM45	NO	305116	8921688	3376	02/03/2021	04:01:32
SMM46	NO	304668	8922041		02/03/2021	
SMM47	NO	304957	8918324	0	02/03/2021	09:02:15
SMM48	NO	306067	8916303	0	02/03/2021	09:03:15
SMM49	NO	307863	8915385	0	02/03/2021	09:15:08
SMM50	NO	304138	8916353	0	02/03/2021	
SMM51	NO	307980	8914144	3896	02/03/2021	11:10:55
SMM52	NO	307378	8915791	3914	02/03/2021	12:16:26
SMM53	NO	307476	8915928	0	02/03/2021	12:22:38
SMM54	NO	307773	8916388	0	02/03/2021	12:59:25
SMM55	NO	308313	8917062	3784	02/03/2021	13:36:40
SMM56	NO	308776	8918245	0	02/03/2021	14:28:45
SMM57	NO	309077	8918637	3739	02/03/2021	14:39:26
SMM58	NO	309307	8919449	3613	02/03/2021	15:10:41
SMM59	NO	309495	8919945	0	02/03/2021	15:21:00
SMM60	NO	309447	8920693	0	02/03/2021	15:36:19
SMM61	NO	309117	8921270	0	02/03/2021	15:53:13
SMM62	NO	308718	8921293	3560	02/03/2021	16:02:02

Nombre	Desliza-miento	X	Y	Z	Fecha	Hora
SMM63	NO	308286	8921229	3547	02/03/2021	16:19:36
SMM64	NO	307175	8923171	3292	03/03/2021	10:55:58
SMM65	NO	306870	8922551	0	03/03/2021	11:08:10
SMM66 camino	NO	306506	8922597	3325	03/03/2021	11:31:31
SMM67 viviendas	NO	306669	8922448	3396	03/03/2021	12:16:03
SMM68	NO	306806	8921910	3445	03/03/2021	12:46:40
SMM69 altura Torretorre	NO	306460	8921789	3456	03/03/2021	12:57:18
SMM70 Pampa Pampa	NO	306071	8921464	3450	03/03/2021	13:18:03
SMM71	NO	306197	8921358	3450	03/03/2021	13:20:14
SMM72	NO	306416	8920851	3427	03/03/2021	13:43:12
SMM73	NO	306394	8920769	3444	03/03/2021	13:48:58
SMM74 Cinculpampa	NO	305967	8920340	3436	03/03/2021	14:04:38
SMM75	NO	305885	8920165	3395	03/03/2021	14:09:27
SMM76	NO	305661	8920152	3425	03/03/2021	14:20:31
SMM77	NO	303827	8918761	3646	03/03/2021	15:30:43
SMM78	NO	303345	8915790	0	03/03/2021	06:12:17
SMM79	NO	304352	8915140	0	03/03/2021	06:14:40
SMM80 Matriz rocosa	NO	305492	8913284	0	03/03/2021	06:21:45
SMM81	NO	306054	8909195	3527	03/03/2021	07:43:05
SMM82	NO	306397	8908747	0	03/03/2021	08:02:01
SMM83	NO	305158	8917995		03/03/2021	
SMM84	NO	305314	8910499	0	04/03/2021	08:48:21
SMM85	NO	304680	8911097		04/03/2021	
SMM86	NO	305583	8921193		04/03/2021	
SMM87	NO	305776	8921508		04/03/2021	
SMM88	NO	305464	8921172		04/03/2021	
SMM89	NO	304039	8916257		04/03/2021	
SMM90	NO	303195	8918654		04/03/2021	
SMM91	NO	303227	8918780		04/03/2021	
SMM92	NO	304646	8920027			
SMM93	NO	306912	8917002			
SMM94	NO	305657	8920054			
SMM95	NO	305841	8919770			
SMM96	NO	307075	8921104			
SMM97	NO	306684	8920500			
CP Buena vista	NO	308340	8917495	0		13:56:07
CP Colpa	NO	309059	8918463	3750		09:54:37
CP Ocrospata	NO	307010	8922790	3399		11:03:30
CP Progreso	NO	305805	8909441	3501		07:31:31
CP Huaruhuayin	NO	309324	8921031	3566		09:39:45
IE Huaruhuayin	NO	309231	8921110	3558		15:48:16
Troca abierta1	NO	303303	8918488	3618		04:27:13

Nombre	Desliza-miento	X	Y	Z	Fecha	Hora
Trocha abierta	NO	303425	8918521	3612		04:22:42
Manantial	NO	306326	8918023	3635		01:32:03
Quebrada Ancahuachanan	NO	306258	8917963	3613		01:39:25
Quebrada Yacuragra	NO	305849	8920104	3380		14:13:42
Acequia Shiula	NO	306930	8919905	3678		11:56:14

Anexo B. Base de datos general de los factores físicos extraídos con puntos de campo

X	Y	Inv. campo*	Elevacion	Pendiente	Longitud Flujo	TWI	NDWI	NDVI	Cobertura Vegetal	Geologia	Distancia Rios	Distancia Vias	Curvatura Plana	Curvatura Perfil
304205	8918634	1	3759	25.377	548.008	6.855	-0.628	0.786	5	5	325.240	454.491	-0.700	1.220
304466	8918712	0	3814	28.459	123.744	5.218	-0.545	0.693	4	5	344.828	312.500	-0.069	0.571
304087	8919386	1	3609	11.794	25.000	5.478	-0.479	0.628	5	9	50.000	27.951	-0.088	0.552
304844	8920239	1	3525	19.686	190.533	6.921	-0.603	0.748	5	7	367.636	0.000	-1.536	1.024
304425	8919255	1	3670	13.518	175.000	7.170	-0.548	0.779	5	9	0.000	27.951	-0.224	1.056
304046	8918641	1	3715	22.596	229.810	6.041	-0.614	0.764	5	5	277.545	530.330	-0.307	-0.307
304247	8918930	1	3712	18.988	106.066	5.538	-0.623	0.789	5	5	35.355	218.303	-0.640	0.640
304574	8919931	1	3549	20.161	855.330	9.916	-0.381	0.542	5	7	0.000	37.500	-0.748	-0.108
303787	8918612	1	3677	13.719	70.711	5.545	-0.578	0.722	5	5	189.159	407.162	-0.250	0.390
303189	8918537	1	3595	20.846	17.678	4.184	-0.499	0.614	5	9	164.412	0.000	-0.128	-0.128
303229	8918560	1	3594	22.192	50.000	5.032	-0.502	0.621	5	9	137.500	0.000	0.000	0.000
303321	8918665	1	3587	10.417	35.355	5.318	-0.439	0.542	5	9	35.355	12.500	-0.160	-0.160
303340	8918716	1	3585	11.873	115.533	6.793	-0.494	0.671	5	9	12.500	12.500	0.000	-0.640
303287	8918863	1	3566	19.529	47.855	4.949	-0.532	0.708	4	9	162.500	72.887	-0.614	1.306
303839	8919313	1	3605	19.827	0.000	3.546	-0.415	0.547	5	9	295.011	12.500	0.640	-0.640
305674	8916657	1	3990	19.471	214.277	6.656	-0.433	0.561	5	5	280.903	0.000	-0.605	0.675
306675	8916043	1	4030	11.528	0.000	4.116	-0.460	0.574	5	5	408.121	95.197	0.763	-0.517
306437	8917912	1	3705	15.793	386.244	8.145	-0.510	0.687	5	7	162.980	1115.025	-0.704	-0.704
306428	8917917	1	3701	19.471	414.277	7.770	-0.488	0.641	5	7	150.520	1102.554	0.535	-0.745
306417	8917910	1	3701	17.249	476.409	9.682	-0.528	0.693	5	7	150.520	1102.554	-1.401	0.519
306346	8917978	1	3656	33.855	459.987	7.269	-0.635	0.792	5	7	79.057	1034.181	-0.750	0.530
306315	8918007	1	3641	23.541	200.888	6.912	-0.594	0.766	5	7	51.539	1002.886	0.246	0.886
306205	8918042	1	3628	20.243	541.053	8.311	-0.482	0.615	4	7	27.951	909.756	-0.693	1.227
305762	8918873	1	3577	22.347	233.211	7.475	-0.371	0.501	5	7	0.000	731.864	-0.675	-0.035
305976	8919070	1	3529	14.069	0.000	3.910	-0.508	0.735	5	7	12.500	588.696	1.280	1.920

X	Y	Inv. campo*	Elevacion	Pendiente	Longitud Flujo	TWI	NDWI	NDVI	Cobertura Vegetal	Geologia	Distancia Rios	Distancia Vias	Curvatura Plana	Curvatura Perfil
305884	8919660	1	3470	22.724	145.711	6.229	-0.295	0.406	5	7	45.069	35.355	0.345	-0.295
305836	8919853	1	3451	20.833	72.855	5.284	-0.600	0.735	5	7	37.500	37.500	0.634	-0.006
305680	8919914	1	3468	19.686	815.686	10.105	-0.405	0.527	5	7	35.355	12.500	-1.024	0.256
305412	8920352	1	3482	18.580	176.777	6.014	-0.581	0.759	5	7	238.812	12.500	0.295	-0.345
305313	8920469	1	3486	22.347	122.855	5.717	-0.559	0.705	5	7	352.003	12.500	0.602	-0.038
303457	8919342	1	3533	15.020	0.000	3.841	-0.542	0.732	5	9	27.951	270.416	0.512	-0.128
303345	8919357	1	3509	18.883	53.033	4.985	-0.515	0.678	4	9	45.069	320.400	-0.489	0.791
302394	8919581	1	3211	28.303	100.888	5.784	-0.415	0.574	5	2	106.066	194.454	-0.640	0.640
303276	8919905	1	3365	25.573	225.000	6.207	-0.551	0.693	5	7	146.309	37.500	-0.052	-0.052
303318	8919888	1	3367	28.622	242.678	6.176	-0.634	0.783	5	7	91.001	35.355	0.645	0.645
303351	8919871	1	3375	28.303	0.000	3.145	-0.480	0.631	4	7	63.738	27.951	0.545	-0.095
303394	8919857	1	3376	30.573	359.099	7.563	-0.501	0.645	5	7	12.500	51.539	0.000	0.640
303775	8920032	1	3385	35.796	100.000	5.050	-0.350	0.515	5	7	53.033	87.500	0.603	-0.037
307432	8919731	1	3793	20.807	50.000	5.103	-0.583	0.741	5	7	51.539	1336.156	0.000	0.000
307045	8919853	1	3697	22.407	223.744	6.908	-0.610	0.792	5	7	87.500	933.408	-0.221	-0.221
306662	8919704	1	3615	21.042	112.500	5.784	-0.469	0.605	4	3	62.500	672.681	0.000	0.000
306012	8920919	1	3379	8.207	55.178	6.072	-0.519	0.695	4	3	0.000	282.843	-0.197	0.443
306031	8920934	1	3379	2.919	2398.160	14.780	-0.486	0.639	5	3	12.500	265.165	-1.280	-0.640
305940	8920885	1	3379	21.826	155.178	6.006	-0.636	0.828	2	3	50.000	350.223	-0.646	1.274
305841	8920897	1	3380	25.179	35.355	4.379	-0.504	0.647	5	3	50.000	415.519	-0.937	0.343
305633	8921164	1	3341	0.810	590.165	12.102	-0.541	0.739	5	3	25.000	513.870	0.000	0.000
305595	8921162	1	3341	11.115	100.000	7.043	-0.688	0.855	5	3	12.500	516.751	0.000	3.200
308202	8915099	1	3974	25.815	173.744	5.960	-0.492	0.626	5	7	482.830	0.000	0.416	-0.224
308184	8913670	1	3867	19.180	569.454	9.459	-0.672	0.798	5	9	75.000	1319.683	-0.666	-0.026
307523	8915975	1	3862	7.429	125.888	7.336	-0.580	0.725	5	7	12.500	341.184	0.000	0.640
307662	8916157	1	3858	34.175	247.487	5.621	-0.493	0.651	4	7	45.069	488.141	0.160	1.440
307809	8916442	1	3836	29.687	447.487	8.181	-0.530	0.669	4	7	39.528	475.986	-1.250	1.310
308028	8916681	1	3817	20.202	128.033	5.828	-0.573	0.708	4	7	117.925	200.000	0.113	-0.527

X	Y	Inv. campo*	Elevacion	Pendiente	Longitud Flujo	TWI	NDWI	NDVI	Cobertura Vegetal	Geologia	Distancia Rios	Distancia Vias	Curvatura Plana	Curvatura Perfil
308465	8917158	1	3770	31.462	237.132	6.483	-0.538	0.704	4	7	212.132	80.039	0.763	-0.517
308412	8917251	1	3818	34.216	35.355	4.010	-0.372	0.506	5	7	246.539	0.000	0.142	-0.498
308547	8917288	1	3755	32.821	35.355	4.063	-0.526	0.687	4	7	212.867	103.078	0.551	0.551
309110	8918788	1	3737	20.780	78.033	5.692	-0.505	0.648	5	7	289.935	0.000	-0.320	-0.320
309060	8918987	1	3723	23.771	677.665	8.141	-0.500	0.645	5	7	106.800	0.000	-1.280	0.640
309072	8919119	1	3711	18.018	354.810	7.581	-0.538	0.687	4	7	25.000	12.500	-1.120	1.440
309183	8919285	1	3684	24.095	317.678	7.356	-0.535	0.709	4	7	206.155	17.678	-0.092	-0.732
309300	8920332	1	3606	24.723	200.000	6.135	-0.359	0.481	5	6	312.500	0.000	0.000	-0.640
308250	8921273	1	3568	13.129	773.896	9.254	-0.385	0.520	5	6	12.500	0.000	-0.584	-0.584
308128	8922611	1	3278	27.758	87.500	5.247	-0.224	0.359	5	6	12.500	0.000	-0.775	1.785
306761	8922594	1	3407	32.670	37.500	4.356	-0.559	0.710	5	3	187.500	176.777	0.558	-0.082
306969	8922221	1	3448	8.049	0.000	4.482	-0.615	0.776	5	3	63.738	12.500	0.448	-0.832
306163	8920460	1	3481	17.249	12.500	4.389	-0.455	0.587	5	3	238.812	0.000	0.151	0.151
306092	8909108	1	3551	18.748	106.066	5.552	-0.574	0.742	5	3	53.033	12.500	0.000	0.000
304076	8918675	0	3714	21.016	582.475	7.394	-0.693	0.840	5	5	279.508	512.957	-1.009	0.911
304240	8918296	0	3899	11.310	0.000	4.135	-0.577	0.723	5	5	308.727	728.011	1.280	-0.640
304626	8918243	0	3986	6.505	0.000	4.697	-0.480	0.627	4	5	406.971	375.000	0.352	-0.928
304763	8918624	0	3894	19.925	155.178	6.105	-0.561	0.691	4	5	438.214	212.867	-0.128	-0.128
304354	8918647	0	3798	27.414	184.099	5.821	-0.554	0.699	4	5	336.573	388.104	1.075	-0.205
304142	8919036	0	3693	6.303	25.000	5.828	-0.601	0.735	5	9	91.001	182.003	0.000	-0.640
303543	8918384	0	3658	12.870	1771.014	12.371	-0.627	0.768	5	5	12.500	347.536	-0.022	-0.662
303467	8918490	0	3630	13.518	37.500	5.743	-0.688	0.826	5	5	0.000	225.347	-0.288	-0.288
303200	8918467	0	3614	10.888	0.000	4.174	-0.656	0.780	3	9	225.347	63.738	0.384	0.384
303303	8918843	0	3573	21.120	0.000	3.477	-0.550	0.694	4	9	150.000	55.902	0.640	-0.640
303535	8919083	0	3589	10.142	95.711	6.326	-0.589	0.748	5	9	27.951	35.355	0.000	0.000
303667	8919226	0	3596	11.447	47.855	5.732	-0.628	0.776	5	9	159.099	27.951	0.256	-0.384
303948	8919379	0	3591	13.518	25.000	5.050	-0.598	0.754	5	9	173.656	37.500	0.096	0.096
306487	8916088	0	4050	9.922	0.000	4.269	-0.575	0.716	4	5	312.500	50.000	1.408	-1.152

X	Y	Inv. campo*	Elevacion	Pendiente	Longitud Flujo	TWI	NDWI	NDVI	Cobertura Vegetal	Geologia	Distancia Rios	Distancia Vias	Curvatura Plana	Curvatura Perfil
307025	8916718	0	4068	9.922	106.066	6.215	-0.514	0.654	4	7	395.285	832.166	0.128	0.128
307100	8916895	0	4102	9.160	17.678	5.044	-0.452	0.598	4	7	440.348	1031.155	0.000	0.000
306795	8917396	0	3920	19.600	150.888	5.956	-0.479	0.649	5	7	209.538	1257.788	0.095	0.095
306157	8918088	0	3629	19.827	106.066	5.492	-0.508	0.652	4	7	53.033	874.017	-0.337	0.943
305966	8918367	0	3640	20.311	229.810	6.159	-0.546	0.690	4	7	107.529	771.464	0.113	-0.527
305795	8918978	0	3584	18.034	17.678	4.341	-0.653	0.819	2	7	75.000	701.115	0.902	-1.018
305789	8919584	0	3525	30.465	88.388	4.848	-0.495	0.631	5	7	152.069	160.078	-1.182	0.098
304494	8919536	0	3636	31.414	360.876	6.110	-0.453	0.622	4	7	215.058	145.774	0.157	0.157
303767	8918972	0	3628	16.398	242.678	6.745	-0.552	0.700	4	9	39.528	106.800	0.591	-0.049
303101	8919523	0	3431	13.129	460.876	8.285	-0.590	0.746	5	9	111.803	301.040	-0.358	0.922
302951	8919522	0	3405	9.632	35.355	5.398	-0.648	0.805	5	9	53.033	275.284	-0.640	0.640
302746	8919673	0	3356	13.518	70.711	5.561	-0.747	0.869	1	5	17.678	27.951	-0.160	-0.160
303078	8919932	0	3352	25.815	197.855	6.085	-0.527	0.677	5	9	294.215	62.500	0.644	-1.276
302607	8919679	0	3306	26.325	67.678	5.021	-0.539	0.719	5	2	45.069	45.069	-0.452	0.188
304120	8920038	0	3440	30.217	137.500	5.551	-0.684	0.836	5	7	50.000	25.000	-0.338	-0.338
304253	8920253	0	3392	42.179	123.744	4.704	-0.515	0.697	5	3	63.738	97.628	1.201	-1.359
304685	8920341	0	3468	12.003	181.066	6.907	-0.702	0.846	5	7	264.870	0.000	-0.960	0.320
304905	8920445	0	3474	17.631	25.000	4.771	-0.687	0.819	5	3	152.069	25.000	0.075	0.075
308442	8918586	0	3968	3.337	0.000	5.368	-0.452	0.618	4	7	637.377	431.023	1.120	-0.160
306260	8919187	0	3606	17.433	0.000	3.684	-0.350	0.490	5	7	212.500	575.543	1.129	-0.151
306306	8919775	0	3564	16.254	366.421	8.102	-0.555	0.698	4	3	300.520	359.905	0.685	-0.595
306359	8920231	0	3516	23.172	70.711	4.984	-0.592	0.750	5	3	12.500	189.159	-0.640	0.640
306154	8920571	0	3489	17.383	42.678	5.073	-0.575	0.731	4	3	279.508	25.000	0.118	0.758
306045	8920860	0	3408	22.383	17.678	4.106	-0.665	0.832	2	3	50.000	273.004	-0.066	-0.066
305378	8921189	0	3433	20.912	12.500	4.181	-0.635	0.824	5	3	206.912	441.765	0.697	0.057
305351	8921456	0	3415	18.534	0.000	3.619	-0.480	0.646	5	3	200.390	657.647	2.683	-0.517
305248	8921431	0	3429	14.471	65.533	5.672	-0.542	0.685	5	3	274.146	643.477	0.094	-0.546
305211	8921674	0	3391	16.236	0.000	3.759	-0.524	0.765	5	3	215.783	788.392	1.570	-0.350

X	Y	Inv. campo*	Elevacion	Pendiente	Longitud Flujo	TWI	NDWI	NDVI	Cobertura Vegetal	Geologia	Distancia Rios	Distancia Vias	Curvatura Plana	Curvatura Perfil
305506	8922253	0	3311	22.046	0.000	3.430	-0.556	0.714	5	7	123.744	686.477	1.036	0.396
305885	8922717	0	3151	30.173	65.533	4.678	-0.456	0.602	5	7	50.000	796.182	-0.440	-0.440
305116	8921688	0	3382	13.541	0.000	3.949	-0.422	0.571	5	3	123.111	888.292	1.655	-0.265
304668	8922041	0	3156	23.585	0.000	3.355	-0.562	0.717	5	3	159.099	1334.693	0.957	-0.963
304957	8918324	0	3969	32.784	70.711	4.575	-0.537	0.687	4	5	226.385	62.500	0.547	-0.093
306067	8916303	0	4022	15.812	224.632	6.965	-0.507	0.638	5	5	309.485	0.000	-0.480	0.800
307863	8915385	0	4037	16.942	47.855	5.324	-0.554	0.696	4	7	400.000	88.388	-0.408	0.872
304138	8916353	0	3657	20.474	12.500	4.204	-0.629	0.771	5	9	25.000	336.573	-0.672	1.888
307980	8914144	0	3913	21.715	314.277	7.184	-0.594	0.740	5	9	100.000	902.947	0.504	-0.776
307378	8915791	0	3933	17.433	0.000	3.684	-0.389	0.521	5	7	88.388	225.347	1.942	-0.618
307476	8915928	0	3875	10.142	70.711	6.645	-0.535	0.683	4	7	17.678	322.587	-0.512	0.768
307773	8916388	0	3853	33.849	217.678	5.816	-0.604	0.737	5	7	90.139	540.833	-0.734	0.546
308313	8917062	0	3800	22.204	88.388	5.820	-0.457	0.620	5	7	35.355	0.000	-0.232	1.048
308776	8918245	0	3810	15.961	95.711	5.723	-0.440	0.577	5	7	245.586	0.000	-0.148	-0.148
309077	8918637	0	3773	18.988	53.033	4.979	-0.624	0.765	5	7	437.500	27.951	-0.337	0.943
309307	8919449	0	3637	18.318	0.000	3.631	-0.509	0.659	4	7	63.738	25.000	0.561	-0.079
309495	8919945	0	3523	28.646	47.855	4.740	-0.571	0.727	4	6	83.853	150.000	0.131	0.771
309447	8920693	0	3552	20.595	136.244	6.144	-0.711	0.899	2	6	237.500	88.388	0.000	0.000
309117	8921270	0	3527	20.050	228.033	6.424	-0.639	0.801	5	6	330.009	128.695	0.570	-0.710
308718	8921293	0	3506	32.676	186.244	5.535	-0.606	0.794	5	6	39.528	123.744	-0.414	-0.414
308286	8921229	0	3572	21.924	12.500	4.129	-0.510	0.646	5	6	35.355	12.500	-0.106	0.534
307175	8923171	0	3322	22.666	112.500	5.702	-0.541	0.699	4	6	246.221	238.812	0.411	-0.229
306870	8922551	0	3416	16.611	85.355	6.133	-0.648	0.786	5	3	170.018	67.315	-1.244	1.956
306506	8922597	0	3348	14.069	286.244	7.861	-0.456	0.592	5	3	197.642	395.285	-0.358	0.922
306669	8922448	0	3402	17.874	125.888	5.960	-0.609	0.749	5	3	35.355	200.390	0.709	0.069
306806	8921910	0	3480	21.603	35.355	4.551	-0.553	0.720	5	7	261.008	12.500	0.320	-0.960
306460	8921789	0	3488	19.911	434.099	8.525	-0.489	0.648	5	7	488.141	39.528	-0.640	0.640
306071	8921464	0	3478	9.922	30.178	5.368	-0.551	0.734	5	7	452.942	0.000	-0.128	-0.128

X	Y	Inv. campo*	Elevacion	Pendiente	Longitud Flujo	TWI	NDWI	NDVI	Cobertura Vegetal	Geologia	Distancia Rios	Distancia Vias	Curvatura Plana	Curvatura Perfil
306197	8921358	0	3486	4.355	113.388	7.499	-0.662	0.803	1	7	466.704	51.539	-0.384	0.256
306416	8920851	0	3463	21.924	30.178	4.534	-0.405	0.529	5	3	25.000	12.500	0.132	0.772
306394	8920769	0	3461	15.812	17.678	4.480	-0.703	0.846	3	3	50.000	37.500	-0.800	0.480
305967	8920340	0	3451	2.291	0.000	5.745	-0.745	0.861	1	3	178.973	137.500	0.000	-0.640
305885	8920165	0	3422	15.218	37.500	5.214	-0.659	0.805	3	3	25.000	135.208	-0.192	-0.192
305661	8920152	0	3432	31.138	555.698	7.141	-0.410	0.553	5	7	62.500	50.000	-0.348	-0.348
303827	8918761	0	3669	11.143	17.678	4.844	-0.615	0.742	5	5	37.500	287.772	0.000	0.000
303345	8915790	0	3556	16.682	83.211	6.033	-0.624	0.772	5	9	139.754	127.475	-0.524	0.756
304352	8915140	0	3967	15.277	0.000	3.824	-0.492	0.635	5	5	225.347	973.476	0.384	-0.256
305492	8913284	0	4000	8.089	0.000	4.477	-0.524	0.656	4	5	270.416	1341.233	0.049	-0.591
306054	8909195	0	3559	16.977	281.066	7.296	-0.564	0.744	5	3	17.678	17.678	-0.158	0.482
306397	8908747	0	3558	15.354	292.310	7.076	-0.643	0.800	5	3	50.000	39.528	0.148	0.148
305158	8917995	0	3982	13.873	30.178	5.023	-0.409	0.571	5	5	312.750	79.057	0.096	0.096
305314	8910499	0	3631	23.001	35.355	4.481	-0.548	0.715	5	5	176.777	441.942	0.113	0.753
304680	8911097	0	3665	10.142	17.678	4.940	-0.574	0.719	4	3	250.312	608.276	0.000	-1.280
305583	8921193	0	3341	7.641	113.388	7.173	-0.614	0.775	5	3	12.500	537.645	0.000	1.920
305776	8921508	0	3392	20.311	25.000	4.618	-0.411	0.570	5	3	185.826	230.489	1.280	-0.640
305464	8921172	0	3409	29.264	70.711	4.714	-0.641	0.783	5	3	134.629	469.707	-0.083	-0.723
304039	8916257	0	3640	12.750	53.033	5.398	-0.686	0.820	3	9	0.000	197.642	0.000	1.280
303195	8918654	0	3549	31.352	65.533	4.813	-0.584	0.730	5	9	39.528	72.887	0.000	0.000
303227	8918780	0	3550	21.414	0.000	3.462	-0.581	0.725	5	9	72.887	100.000	0.352	-0.928
304646	8920027	0	3559	27.995	60.355	4.767	-0.335	0.446	5	7	100.000	12.500	0.145	-0.495
306912	8917002	0	4078	14.615	37.500	5.256	-0.440	0.585	4	7	308.727	1026.904	0.000	0.640
305657	8920054	0	3459	28.630	53.033	4.517	-0.638	0.776	5	7	132.877	25.000	0.000	0.000
305841	8919770	0	3447	18.534	287.132	7.510	-0.665	0.800	5	7	25.000	25.000	-1.280	1.280
307075	8921104	0	3623	16.891	0.000	3.718	-0.602	0.741	5	7	200.000	439.638	1.147	-0.773
306684	8920500	0	3581	31.899	88.388	4.792	-0.600	0.761	5	3	279.508	335.876	-0.480	0.800
308340	8917495	0	3839	10.417	260.355	7.310	-0.353	0.479	5	7	341.184	0.000	0.128	0.128

X	Y	Inv. campo*	Elevacion	Pendiente	Longitud Flujo	TWI	NDWI	NDVI	Cobertura Vegetal	Geologia	Distancia Rios	Distancia Vias	Curvatura Plana	Curvatura Perfil
309059	8918463	0	3784	16.398	55.178	5.541	-0.407	0.508	5	7	559.715	0.000	-0.595	0.685
307010	8922790	0	3422	7.209	259.099	8.059	-0.405	0.526	5	6	312.750	17.678	-0.640	0.640
305805	8909441	0	3535	8.660	42.678	5.794	-0.499	0.630	5	3	51.539	12.500	0.000	0.000
309324	8921031	0	3580	9.697	30.178	5.679	-0.463	0.632	5	6	379.556	12.500	-0.602	0.038
309231	8921110	0	3577	22.204	0.000	3.422	-0.534	0.684	5	6	416.083	12.500	0.776	-0.504
303303	8918488	0	3633	20.202	35.355	4.624	-0.602	0.775	5	9	151.038	100.778	0.851	-1.069
303425	8918521	0	3623	11.365	642.310	8.794	-0.408	0.505	5	9	25.000	167.705	-0.517	0.763
306326	8918023	0	3650	28.029	183.211	6.683	-0.537	0.743	5	7	62.500	1017.426	0.338	-0.302
306258	8917963	0	3640	15.961	138.388	6.080	-0.551	0.702	5	7	0.000	945.797	0.148	0.148
305849	8920104	0	3410	5.878	206.954	7.890	-0.498	0.692	5	3	0.000	146.309	0.480	0.480
306930	8919905	0	3699	14.069	0.000	3.910	-0.688	0.827	1	7	184.560	794.611	0.640	-0.640

* DM indica presencia (1) o ausencia (0) de deslizamientos en masa.
De esta base de datos se particionará el 80% para puntos de entrenamiento del modelo y 20% para puntos de validación del modelo.

Anexo C. Ajuste del modelo lineal generalizado

Se ajustó la base de datos de entrenamiento al modelo lineal generalizado, para ello, en primer lugar, el glm fue corrido con todas las variables (modelo general o full). Las variables del análisis de devianza de la **tabla 15** fueron adheridos secuencialmente; las variables Pendiente, Longitud de Flujo, TWI, NDVI y Distancia a ríos son las variables que muestran mayor relación P (> |Chi|) < 0,05 con la variable dependiente, están marcadas con asteriscos de acuerdo al grado de significancia.

Tabla 15. Test de razón de verosimilitudes usando todas las variables en GLM

Análisis de Devianza					
	Gl.	Devianza	Gl. Residual	Dev. residual	P (> |Chi|)
NULL			140	189,46	
Elevación	1	0,0045	139	189,46	0,94549
Pendiente	1	3,5775	138	185,88	0,05857 .
LongFlujo	1	4,8655	137	181,01	0,02740 *
TWI	1	4,1989	136	176,81	0,04045 *
NDVI	1	3,5568	135	173,26	0,05930 *
CoberturaVegetal	4	3,2854	131	169,97	0,51125
Geología	5	7,2882	126	162,68	0,20007
DistanciaRios	1	3,5476	125	159,14	0,05963 .
DistanciaVias	1	0,4607	124	158,68	0,49729
CurvaturaPlana	1	0,7883	123	157,89	0,37463
CurvaturaPerfil	1	1,5773	122	156,31	0,20916
Códigos de significancia: 0 ‘***’ 0,001 ‘**’ 0,01 ‘*’ 0,05 ‘.’ 0,1 ‘ ’ 1					

Fuente: Elaboración propia, datos obtenidos en el software R.

La **tabla 16**, muestra el efecto de los factores más significativos de acuerdo con su valor z, estas son las variables Pendiente, NDVI y Distancia a ríos, que tienen efectos significativos y presentan valores z con signo positivo.

Tabla 16. Resumen del ajuste de los datos al modelo lineal generalizado

	Estimación	Error std.	Valor z	Pr(>|z|)
(Intercept)	1,040e+00	4.799+03	0,000	0,9998
Elevación	1,928e-04	1,421e-03	0,136	0,8920
Pendiente	5,681e-02	3,100e-02	1,833	0,0669 .
Longitud de Flujo	-6,002e-04	1,361e-03	-0,441	0,6593
TWI	3,514e-01	2,498e-01	1,407	0,1595
NDVI	-4,820e+00	2,427e+00	-1,986	0,0470 *
Cobertura Vegetal2	-3,395e-02	3,721e+03	0,000	1,0000
Cobertura Vegetal3	-9,044e-01	3,547e+03	0,000	0,9998
Cobertura Vegetal4	1,510e+01	2,716e+03	0,006	0,9956
Cobertura Vegetal5	1,518e+01	2,716e+03	0,006	0,9955
Geología 3	-1,688e+01	3.956e+03	-0,004	0,9966
Geología 5	-1,635e+01	3.956e+03	-0,004	0,9967
Geología 6	-1,893e+01	3.956e+03	-0,005	0,9962
Geología 7	-1,645e+01	3.956e+03	-0,004	0,9967

Geología 9	-1,617e+01	3.956e+03	-0,004	0,9967
Distancia a ríos	-3,186e-03	1,799e-03	-1,771	0,0766 .
Distancia vías	-2,682e-04	5,719e-04	-0,469	0,6390
Curvatura plana	-7,956e-01	3,921e-01	-0,203	0,8392
Curvatura de perfil	3,964e-01	3,160e-01	1,255	0,2096
Códigos de significancia: 0 ‘***’ 0,001 ‘**’ 0,01 ‘*’ 0,05 ‘.’ 0,1 ‘ ’ 1				
Devianza nula: 189,46 sobre 140 grados de libertad (Gl.)				
Devianza residual: 156,31 sobre 122 grados de libertad (Gl.)				
AIC: 194,31				
Número de iteraciones de puntuación de Fisher: 16				

Fuente: Elaboración propia, datos obtenidos en el software R.

Posteriormente al análisis de todas las variables con el modelo lineal generalizado, se procederá a hacer la selección de variables para el modelo final, a través de análisis AIC paso a paso (función stepwise en R). A continuación, se muestra el resumen del proceso de selección de las variables para el modelo final, usando el Criterio de Información de Akaike como único estimador.

Tabla 17. Selección del mejor modelo de susceptibilidad a deslizamientos de masas

Modelo inicial:						
y ~ Elevación + Pendiente + Longitud de flujo + TWI + NDVI + Cobertura vegetal + Geología + Distancia a ríos + Distancia a vías + Curvatura plana + Curvatura de perfil						
N°	Iteración	Gl.	Devianza	Gl. residual	Dev. residual	AIC
01				122	156,3096	194,3096
02	- Cobertura vegetal	4	1,8147	126	158,1243	188,1243
03	- Geología	5	7,3409	131	165.4652	185,4652
04	- Curvatura Plana	1	0,0243	132	165,4895	183,4895
05	- Distancia a vías	1	0.1439	133	165,6335	181,6335
06	- Longitud de flujo	1	0,1865	134	165,8200	179,8200
07	- Elevación	1	0,9252	135	166,7452	178,7452
Modelo final:						
y ~ Pendiente + TWI + NDVI + Distancia a ríos + Curvatura de perfil						

Fuente: Elaboración propia, datos obtenidos en el software R.

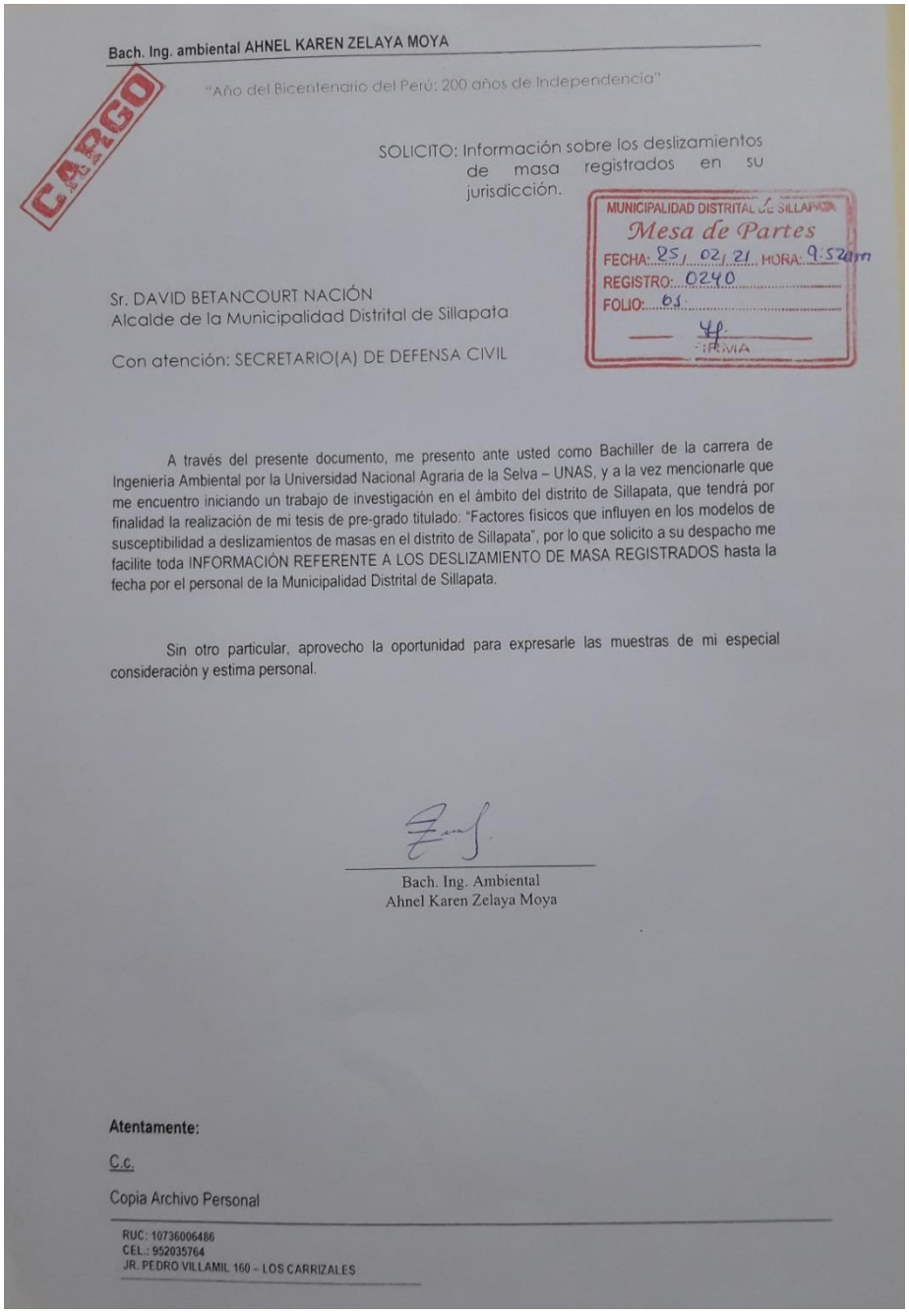
La **tabla 17** muestra que el algoritmo realizó siete pasos (*stepwise*) para lograr una función que corresponda con el menor AIC, alcanzando dicha condición con cinco variables: Pendiente, Índice de humedad del Terreno (TWI), Índice Diferencial de Vegetación Normalizado (NDVI), Distancia a Ríos y Curvatura de perfil del terreno.

Anexo D. Valores de validación para el análisis ROC por punto

Orden punto	Umbral	Sensibilidad	Especificidad
1	--	1	0
2	0.064	1	0.04
3	0.073	1	0.08
4	0.093	1	0.12
5	0.118	0.9	0.12
6	0.138	0.9	0.16
7	0.149	0.9	0.2
8	0.152	0.9	0.24
9	0.160	0.9	0.28
10	0.169	0.9	0.32
11	0.176	0.9	0.36
12	0.184	0.9	0.4
13	0.190	0.9	0.44
14	0.192	0.9	0.48
15	0.222	0.9	0.52
16	0.258	0.8	0.52
17	0.265	0.8	0.56
18	0.273	0.8	0.6
19	0.297	0.8	0.64
20	0.323	0.8	0.68
21	0.331	0.8	0.72
22	0.348	0.8	0.76
23	0.383	0.8	0.8
24	0.422	0.8	0.84
25	0.448	0.7	0.84
26	0.458	0.6	0.84
27	0.480	0.6	0.88
28	0.514	0.6	0.92
29	0.553	0.5	0.92
30	0.621	0.4	0.92
31	0.706	0.4	0.96
32	0.779	0.3	0.96
33	0.830	0.3	1
34	0.854	0.2	1
35	0.868	0.1	1
36	--	0	1

Los puntos fueron extraídos de la base de datos de validación (20% del total de datos).

Anexo E. Fotografías



Bach. Ing. ambiental AHNEL KAREN ZELAYA MOYA

CARGO

"Año del Bicentenario del Perú: 200 años de Independencia"

SOLICITO: Información sobre los deslizamientos de masa registrados en su jurisdicción.

MUNICIPALIDAD DISTRITAL DE SILLAPATA
Mesa de Partes
FECHA: 25/02/21 HORA: 9:52am
REGISTRO: 0240
FOLIO: 01
FIRMA

Sr. DAVID BETANCOURT NACIÓN
Alcalde de la Municipalidad Distrital de Sillapata

Con atención: SECRETARIO(A) DE DEFENSA CIVIL

A través del presente documento, me presento ante usted como Bachiller de la carrera de Ingeniería Ambiental por la Universidad Nacional Agraria de la Selva – UNAS, y a la vez mencionarle que me encuentro iniciando un trabajo de investigación en el ámbito del distrito de Sillapata, que tendrá por finalidad la realización de mi tesis de pre-grado titulado: "Factores físicos que influyen en los modelos de susceptibilidad a deslizamientos de masas en el distrito de Sillapata", por lo que solicito a su despacho me facilite toda INFORMACIÓN REFERENTE A LOS DESLIZAMIENTO DE MASA REGISTRADOS hasta la fecha por el personal de la Municipalidad Distrital de Sillapata.

Sin otro particular, aprovecho la oportunidad para expresarle las muestras de mi especial consideración y estima personal.

Bach. Ing. Ambiental
Ahnel Karen Zelaya Moya

Atentamente:

C.c.

Copia Archivo Personal

RUC: 10736006486
CEL.: 952035764
JR. PEDRO VILLAMIL 160 – LOS CARRIZALES

Solicitud de información temática enviada a la M.D. Sillapata.

Equipo GPS capturando un punto de zona estable SMM1.

Georreferenciación de la ubicación de un área de deslizamiento (código CMM4) al margen de una vía vecinal

Georreferenciación de punto de deslizamiento CMM27, con fecha 27/02/2021.

Georreferenciación de un punto sin deslizamiento SMM28, con fecha 28/02/2021

Georreferenciación de corona de deslizamiento CMM39, con fecha 01/03/2021

Georreferenciación de punto de deslizamiento CMM50 cercana a una quebrada, con fecha 02/03/2021

Georreferenciación de punto de deslizamiento CMM63 en el sector denominado Saporumi, con fecha 03/03/2021

Georreferenciación de un punto sin deslizamiento SMM74 en el caserío Cinculpampa, con fecha 03/03/2021.

Georreferenciación de un punto de deslizamiento CMM67 en el margen de la vía vecinal, en el caserío Progreso, con fecha 04/03/2021

Vista de zona georreferenciada de un punto sin deslizamiento SMM82, en la zona alta del caserío Progreso, con fecha 04/03/2021

Anexo F. Scripts utilizados para la obtención de resultados

Script 1. Generación de nube de puntos del inventario

```
# Leer el csv
csv_ruta <- 'D:/RESULTADOS/INVENTARIO_PUNTOS.csv'
deslizamientos <-  read.csv(csv_ruta)
head(deslizamientos)
# Codificar la respuesta
deslizamientos$y <- c()
deslizamientos[deslizamientos[,'DESLIZAMIENTO']=='SI', 'y'] <- 1
deslizamientos[deslizamientos[,'DESLIZAMIENTO']=='NO', 'y'] <- 0

# Coordenadas en este y norte (sistena UTM)
deslizamientos$easting <- deslizamientos$X
deslizamientos$northing <- deslizamientos$Y

# Convertir los datos a vector
utm18sCRS <- st_crs(32718) # Defining a CRS with sf
class(utm18sCRS)
utm18sCRS
locations_landslide  <-  st_as_sf(deslizamientos,  coords  =  c("easting",
"northing"), crs = utm18sCRS)
st_crs(locations_landslide)
# Graficar la información espacial
ggplot() +
geom_sf(data = locations_landslide) +
ggtitle("Map of Plot Locations")
```

Script 2. Extracción de factores físicos en nube de puntos

```
#Leer raster con variables predictoras
raster_nom <- 'D:/RESULTADOS/12VARIABLES.tif'
raster_ds <- terra::rast(raster_nom) # terra package

#Renombrar bandas
names(raster_ds) <- c('Elevacion', 'Pendiente', 'LongFlujo', TWI, 'NDWI',
'NDVI', 'CoberturaVegetal', 'Geologia', 'DistanciaRios', 'DistanciaVias',
'CurvaturaPlana', 'CurvaturaPerfil')
names(raster_ds)
plot(raster_ds)

#Extraer valores de píxel
df_extr <- terra::extract(raster_ds, terra::vect(locations_landslide))
head(df_extr)

#Eliminar columna
df_extr <- subset(df_extr, select = -c(ID))
head(df_extr)

#Unir conjuntos de datos
landslide <- cbind(locations_landslide, df_extr)
```

```
'landslide <- landslide[,c('X', 'Y', 'y', 'Elevacion', 'Pendiente',
'LongFlujo', TWI, 'NDWI', 'NDVI', 'CoberturaVegetal', 'ResistenciaSuelo',
'DistanciaRios', 'DistanciaVias', 'CurvaturaPlana', 'CurvaturaPerfil')]
head(landslide)

#Remover geometría
landslide = as.data.frame(st_drop_geometry(landslide))
head(landslide)

#Convertir variable a factor
landslide$CoberturaVegetal <- as.factor(landslide$CoberturaVegetal)
landslide$Geología <- as.factor(landslide$Geología)
str(landslide)

#Guardar conjunto de datos
write.csv(landslide, 'D:/RESULTADOS/Variables_Sillapata.csv')
```

Script 3. Análisis exploratorio de variables cuantitativas y cualitativas

```
#Matriz de correlación (var.cuantitativas)
pairs(landslide[,c('Elevacion', 'Pendiente', 'LongFlujo', 'TWI', 'NDWI',
'NDVI', 'DistanciaRios', 'DistanciaVias', 'CurvaturaPlana',
'CurvaturaPerfil')],pch=20)
matriz_cuantitativa <- cor(landslide[,c('Elevacion', 'Pendiente',
'LongFlujo', 'TWI', 'NDWI', 'NDVI', 'DistanciaRios', 'DistanciaVias',
'CurvaturaPlana', 'CurvaturaPerfil')])
write.csv(matriz_cuantitativa,"D:/RESULTADOS/Matrizcuantitativa.csv")

#Hacer un gráfico de mosaico (var.cualitativas)
#Factor CoberturaVegetal
Cobertura<-table(landslide$CoberturaVegetal,as.factor(landslide$y))
Cobertura
mosaicplot(Cobertura,color=c("#A2CD5A","#CD3333"))
#Factor Geologia
Geología<-table(landslide$Geologia,as.factor(landslide$y))
Geología
mosaicplot(Geología,color=c("#FFEC8B","#EE5C42"))

#Descartar las variables redundantes: Descartar NDWI por ser muy similar
a NDVI
landslide<- landslide[,c("X", "Y", "y",'Elevacion', 'Pendiente',
'LongFlujo', 'TWI', 'NDVI', 'CoberturaVegetal', 'Geologia',
'DistanciaRios', 'DistanciaVias', 'CurvaturaPlana', 'CurvaturaPerfil')]
str(landslide)
```

Script 4. Particionar la base de datos en puntos de entrenamiento y puntos de validación

```
#Crear una partición de puntos (entrenamiento y validación)
library(caret)
#el tamaño de muestra será de 80%
trainIndex=createDataPartition(landslide$y, p=0.80)$Resample1
```

```
entrenam_landslide=landslide[trainIndex, ] # conjunto entrenamiento
valid_landslide= landslide[-trainIndex, ] # conjunto de test
#Guardarlos
write.csv(entrenam_landslide, 'D:/RESULTADOS/puntos_entrenamiento.csv')
write.csv(valid_landslide, 'D:/RESULTADOS/puntos_validacion.csv')

#Quitar coordenadas XY a puntos de entrenamiento
entrenam_landslide <- entrenam_landslide[,c('y', 'Elevacion', 'Pendiente',
'LongFlujo',    'TWI',    'NDVI',    'CoberturaVegetal',    'Geologia',
'DistanciaRios', 'DistanciaVias', 'CurvaturaPlana', 'CurvaturaPerfil')]
str(entrenam_landslide)
```

Script 5. Ajuste de los datos a la función logit del modelo lineal generalizado

```
# Modelo con todas las variables
full <- glm(y ~ ., data = entrenam_landslide, family = binomial(link =
"logit"))
summary(full)
anova(full, test = "Chi")
plot(full)
```

Script 6. Ajuste y simplificación del modelo final con el AIC

```
#Evaluación con el AIC
library(MASS)
step <- stepAIC(full, trace = TRUE)
step$anova

#Generación de coeficientes del modelo final
step$coefficients
```

Script 7. Análisis de significancia de las variables finales

```
#Análisis de significancia
library(jtools)
plot_summs(step, scale = TRUE, plot.distributions = TRUE, inner_ci_level
= .9)
library(sjPlot)
plot_model(step)
```

Script 8. Predicción del mapa de susceptibilidad final

```
#Predicción para el área
prediccion <- subset(raster_ds, c('Pendiente', 'TWI', 'NDVI',
'DistanciaRios', 'CurvaturaPerfil'))
plot(prediccion)

Raster_pred <- predict(prediccion, step, na.rm = TRUE, type="response")
Raster_pred
plot(Raster_pred)

#Guardar raster
rst_out <- 'D:/RESULTADOS/Mapa_final.tif'
```

Script 9. Validación con el AUC ROC del modelo final

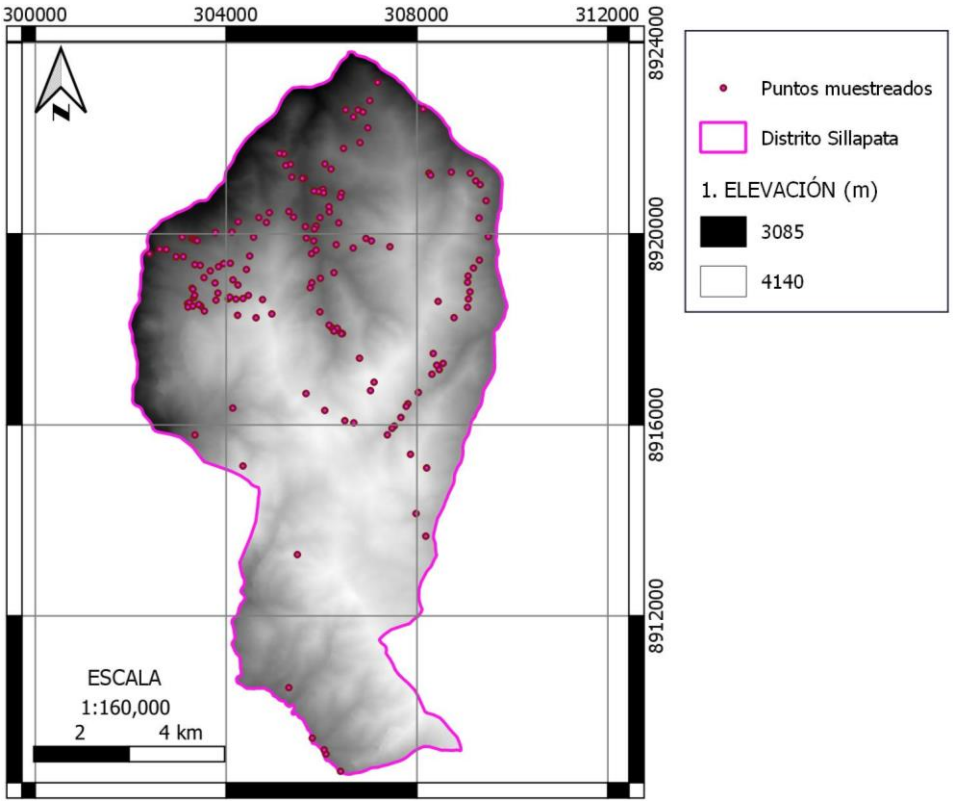
```
#Acondicionar la base de datos de validación para el análisis posterior
valid_landslide <- valid_landslide[,c('X', 'Y', 'y')]
str(valid_landslide)

#Georeferenciar puntos validación
# easting and northing (UTM coordinates)
valid_landslide$easting <- valid_landslide$X
valid_landslide$northing <- valid_landslide$Y

# Converto to vector sf
utm18sCRS <- st_crs(32718) # Defining a CRS with sf
class(utm18sCRS)
utm18sCRS

valid_georef <- st_as_sf(valid_landslide, coords = c("easting",
"northing"), crs = utm18sCRS)
st_crs(valid_georef)
head(valid_georef)

#Extraer pixel a puntos
puntosROC <- terra::extract(Raster_pred, terra::vect(valid_georef))
head(puntosROC)
puntosROC <- subset(puntosROC, select = -c(ID)) #Quitar el ID
head(puntosROC)

# Unir base de datos
DatosROC <- cbind(valid_georef, puntosROC)

DatosROC <- DatosROC[,c('y', 'lyr1')]
head(DatosROC)

#Remover geometría
Datos_valid = as.data.frame(st_drop_geometry(DatosROC))
# rename bands
names(Datos_valid) <- c('CAMPO', 'PREDICCION')
head(Datos_valid)
write.csv(Datos_valid, 'D:/ROC/PUNTOS_ROC.csv')
```

```
#Cargar y llamar el paquete necesario.
library("pROC")

#Crear las curvas sobre las que se trabajará.
dfcor<-read.csv('D:/ROC/PUNTOS_ROC.csv')
roc<-roc(dfcor$CAMPO,dfcor$PREDICCION)
roc1<-roc(dfcor$CAMPO,dfcor$PREDICCION,smooth=TRUE) #Curva ROC suavizada
plot(roc)
plot(roc1)
auc(roc)
coords(roc,transpose = TRUE)
coords(roc,transpose = FALSE)
```

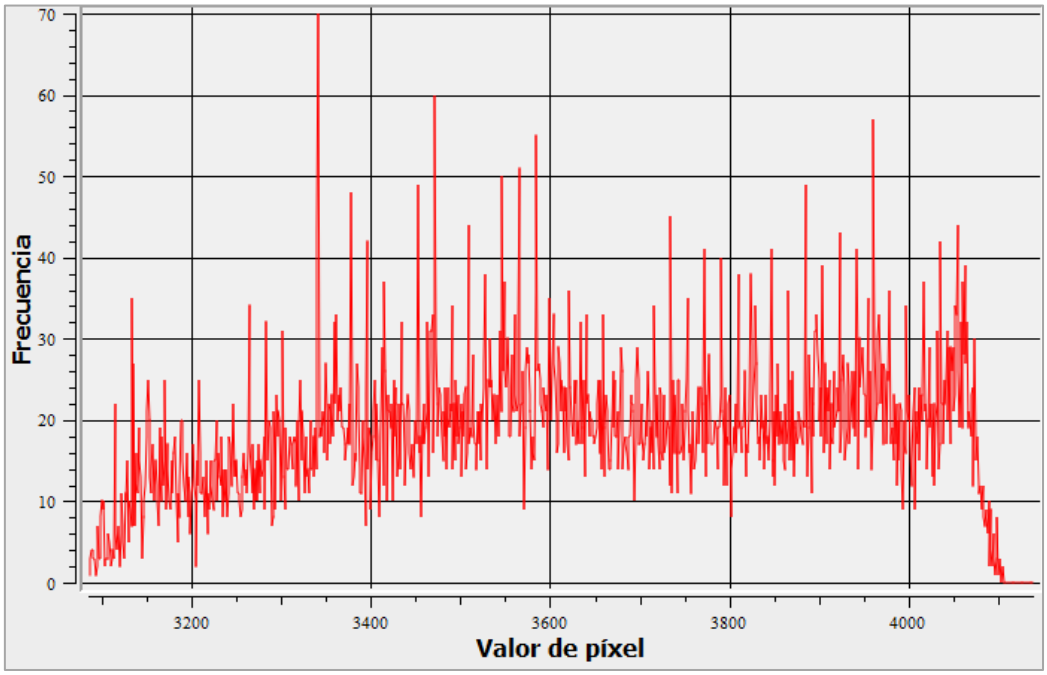
```
#Guardar la tabla de datos del Análisis
ANALISISROC<- coords(roc,transpose = FALSE)
write.csv(ANALISISROC,"D:/ROC/ANALISISROC.csv")

#Mejorar apariencia de la curva.
#insertar AUC y IC
rocobj   <-   plot.roc(dfcor$CAMPO,dfcor$PREDICCION,main="Intervalos   de
confianza", percent=TRUE,ci=TRUE,print.auc=TRUE)

#Creación de objeto para plotear el IC por defecto al 95% CI. Boots Default:
2000.
ciobj <- ci.se(rocobj, specificities=seq(0, 100, 5))

#insertamos el IC
plot(ci(rocobj, of="thresholds", thresholds="best")) # add one threshold

#Insertar el punto de corte sobre la gráfica con su IC y personalizar
umbral<-plot.roc(dfcor$CAMPO,col="blue",dfcor$PREDICCION, main="Intervalo
de    confianza    del    umbral",    percent=TRUE,ci=TRUE,print.auc=TRUE,
of="thresholds",thresholds="best",print.thres="best",xlab="Especificidad
(%)",ylab="Sensitividad (%)", print.thres.pattern= "%.3f (%.2f%%,%.2f%%)",
auc=TRUE)

#Potencia estadística de tu objeto.
power.roc.test(rocobj)
```

Anexo G. Distribución espacial y frecuencia de factores físicos empleados

G.1. Factor de elevación

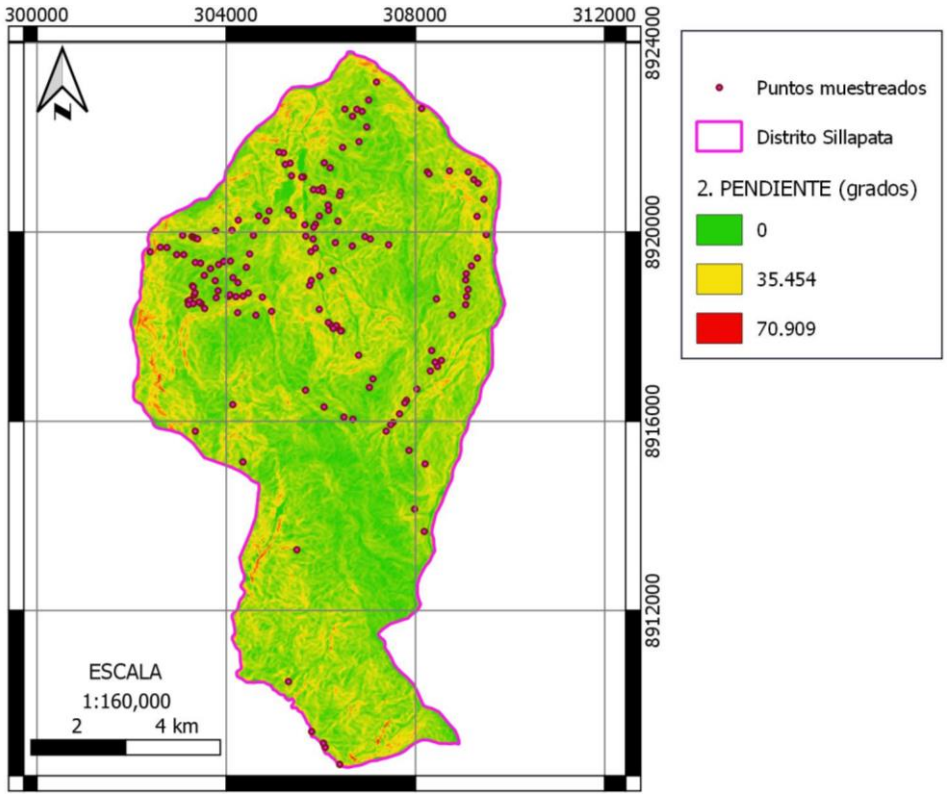


Mapa de elevación del distrito Sillapata

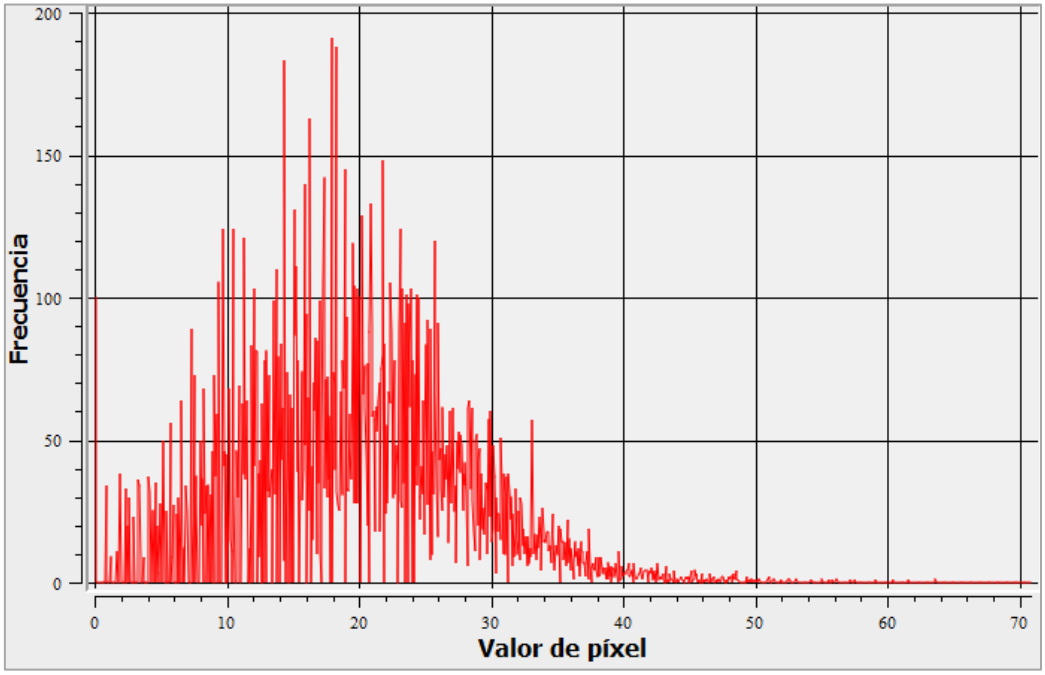


Histograma de valores de elevación en el área de estudio

G.2. Factor de grados de pendiente

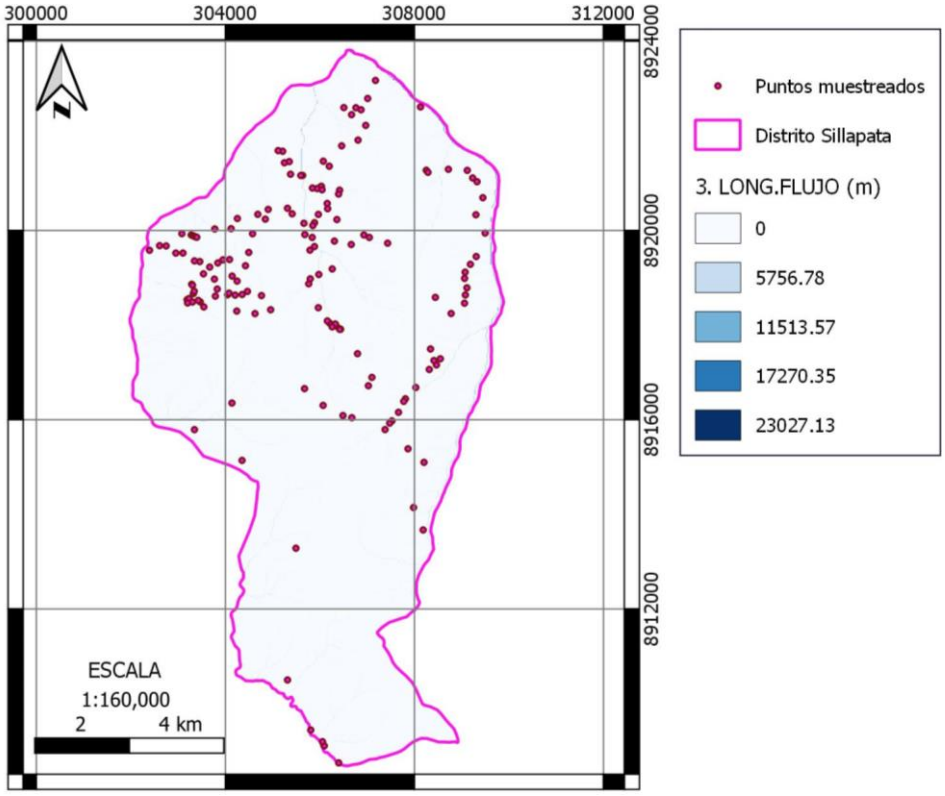


Mapa de pendientes en grados del distrito Sillapata

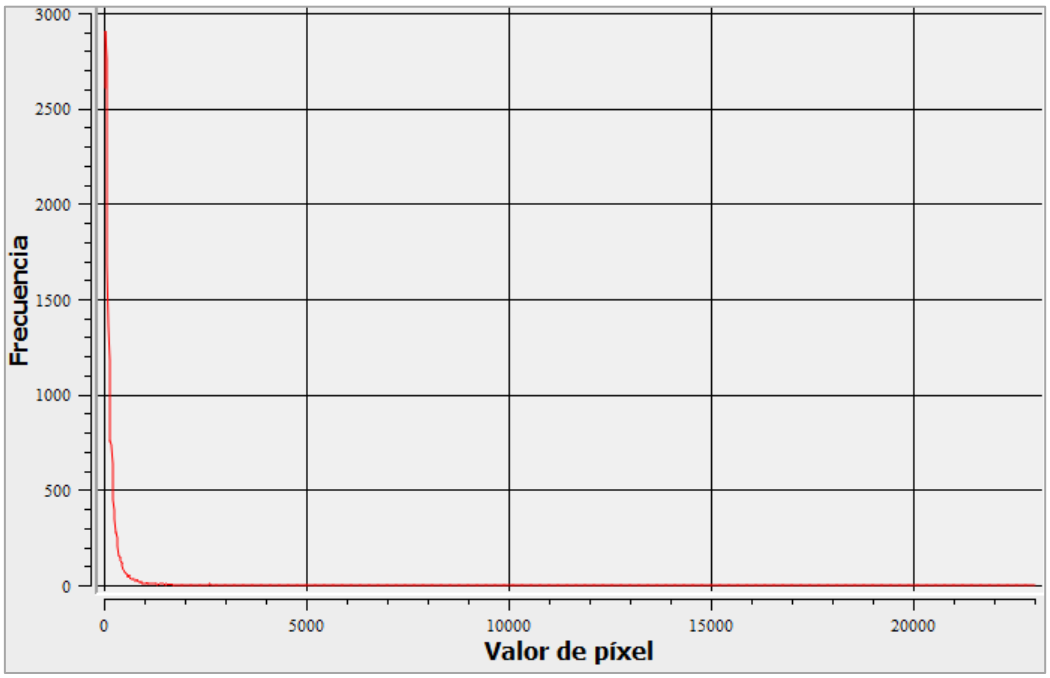


Histograma de valores de grados de pendiente en el área de estudio

G.3. Factor de longitud de flujos

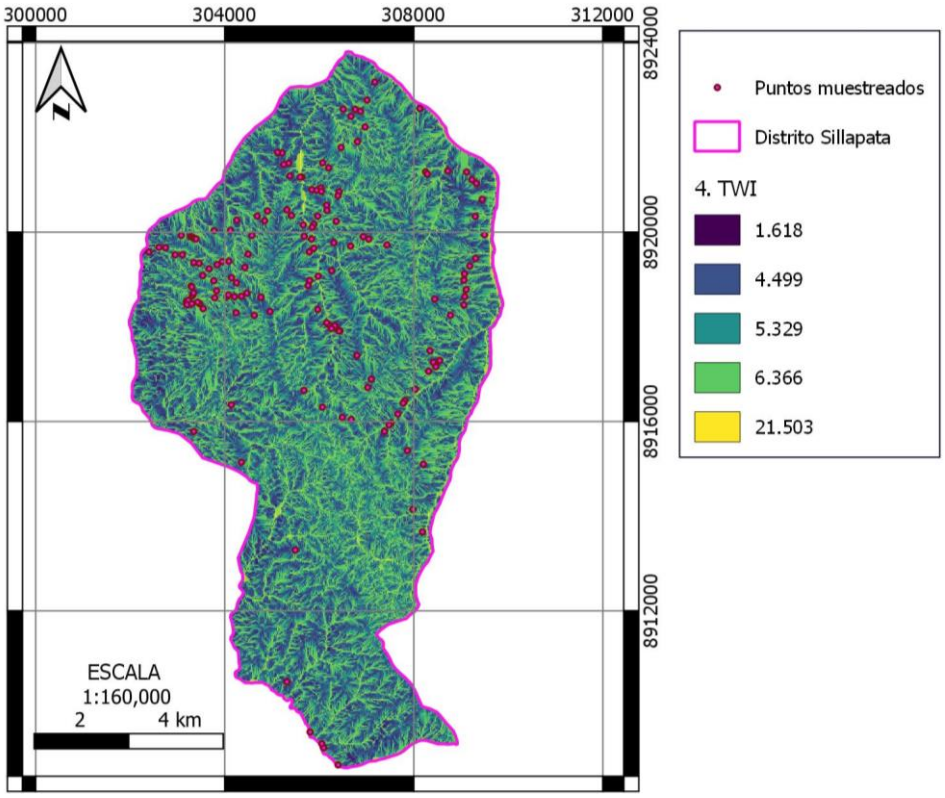


Mapa de longitud de flujos del distrito Sillapata

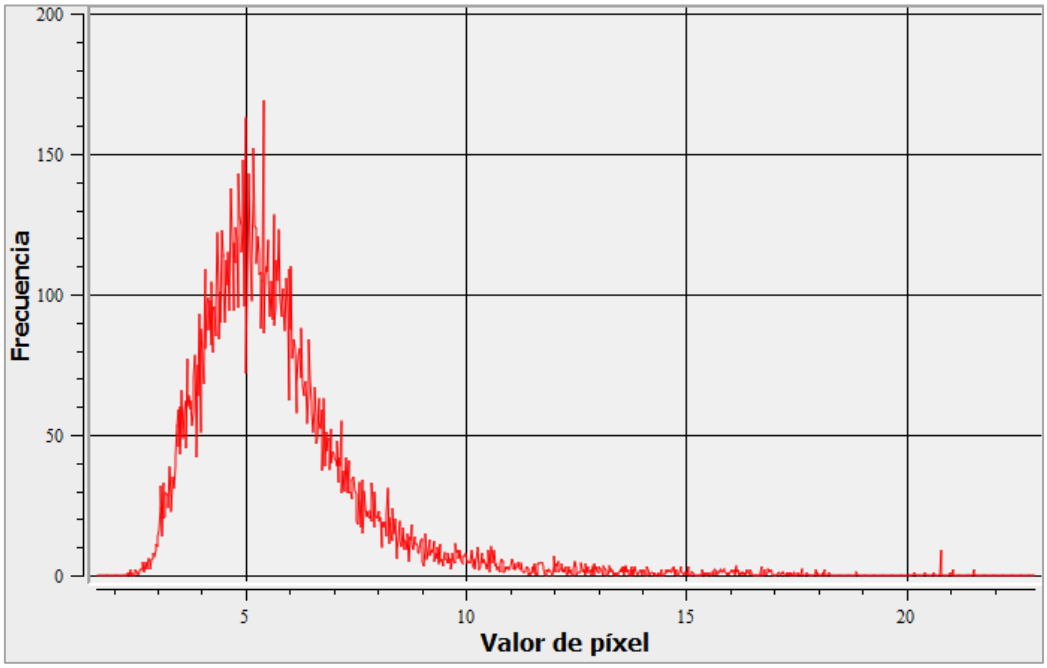


Histograma de valores de longitud de flujos en el área de estudio

G.4. Factor de índice de humedad topográfica - TWI

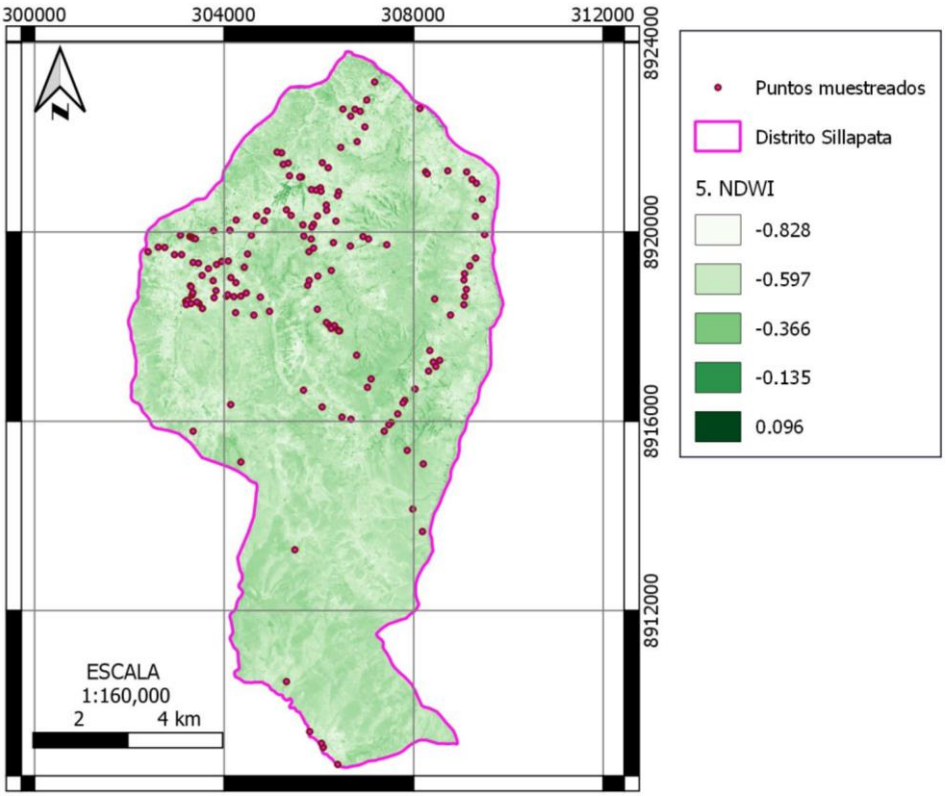


Mapa del índice de humedad topográfica TWI del distrito Sillapata

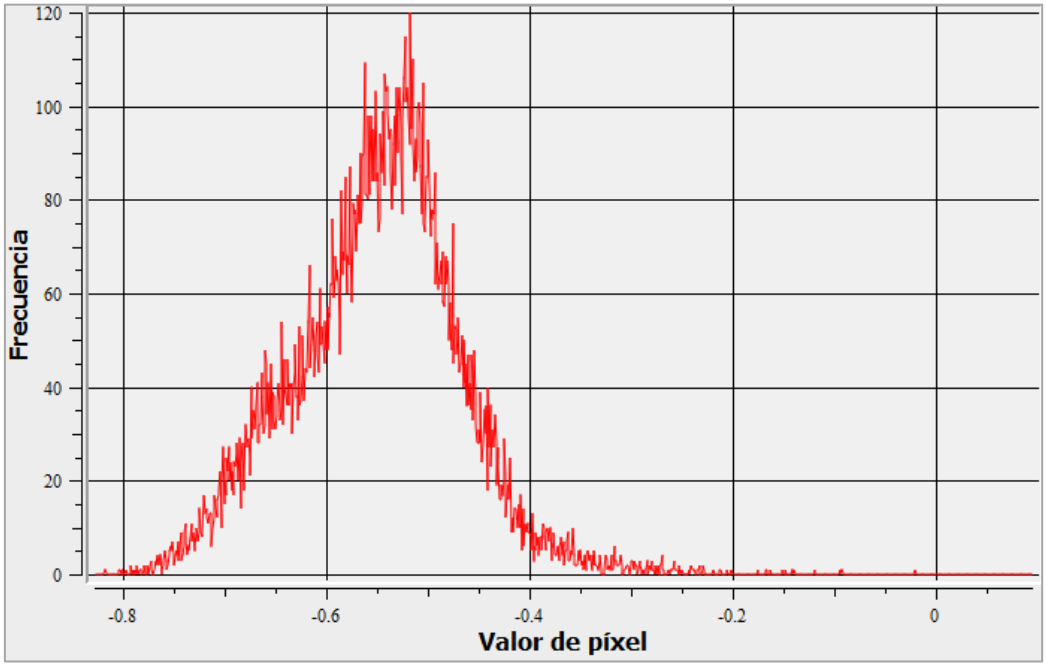


Histograma de valores del índice de humedad topográfica en el área de estudio

G.5. Factor del índice diferencial de agua normalizado - NDWI

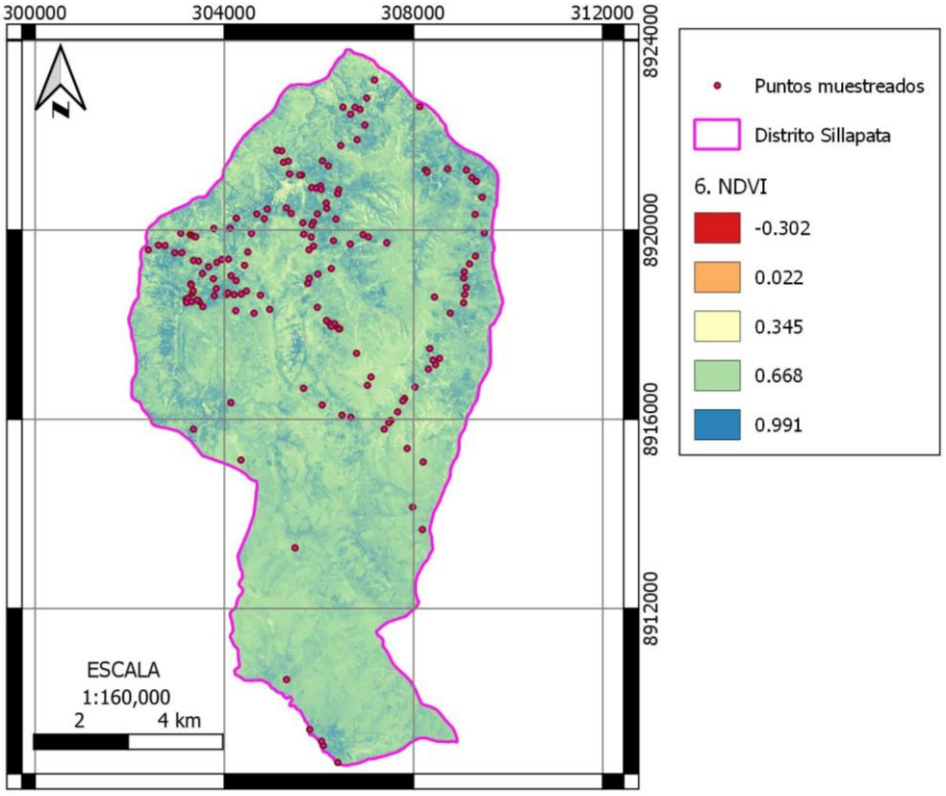


Mapa del índice diferencial de agua normalizado NDWI del distrito Sillapata

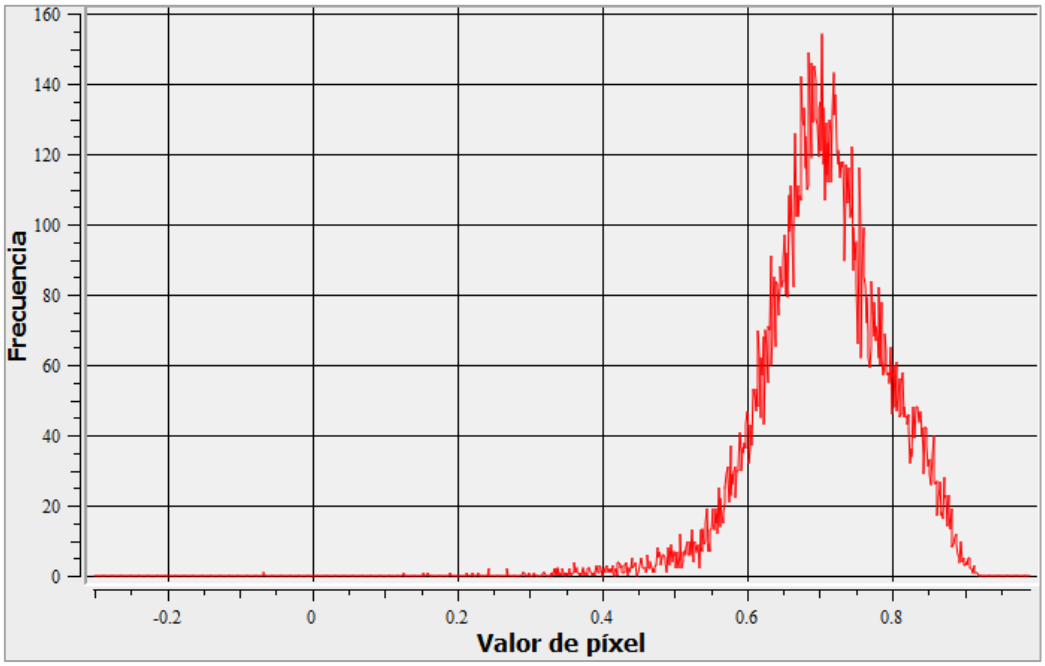


Histograma de valores del índice diferencial de agua normalizado en el área de estudio

G.6. Factor del índice diferencial de vegetación normalizado - NDVI

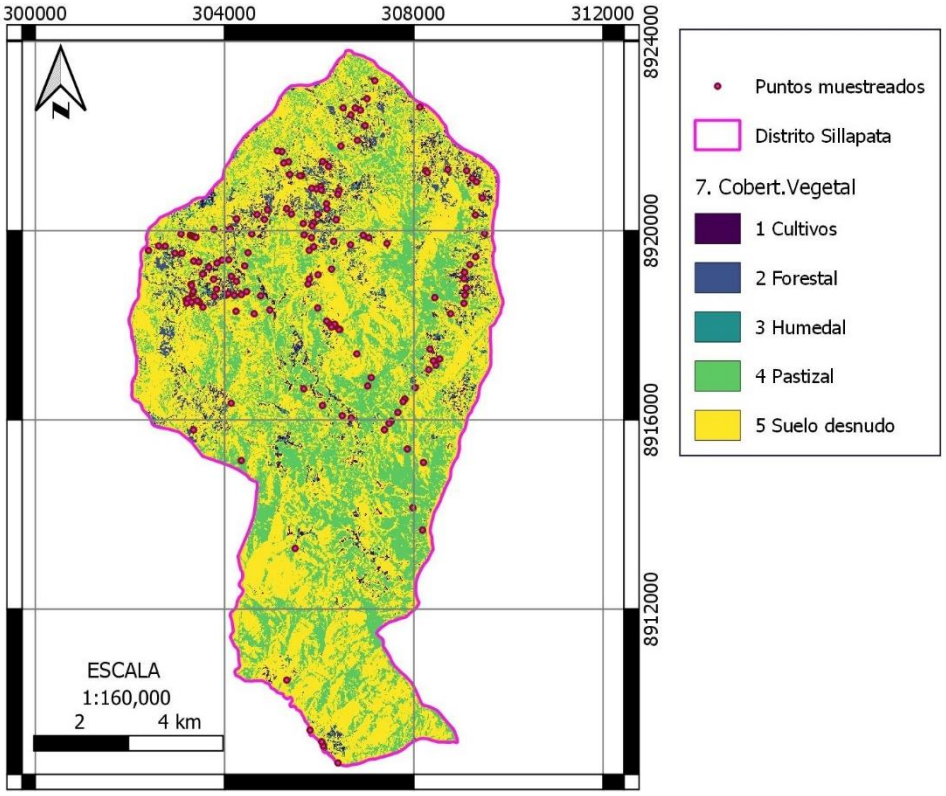


Mapa del índice diferencial de vegetación normalizado NDVI del distrito Sillapata

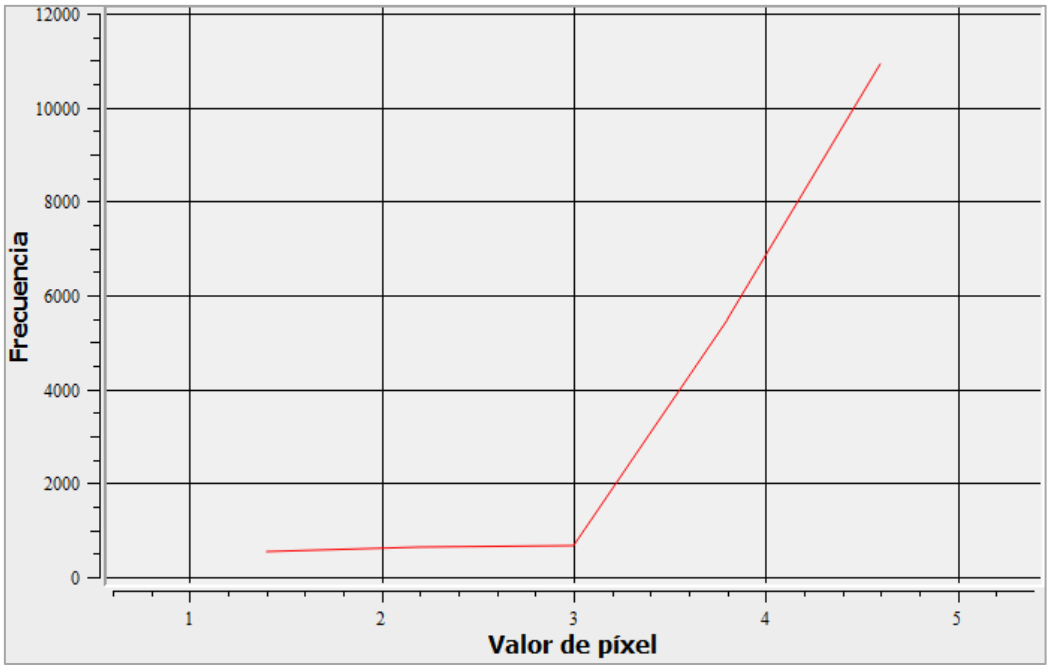


Histograma de valores del índice diferencial de vegetación normalizado en el área de estudio

G.7. Factor de la cobertura vegetal

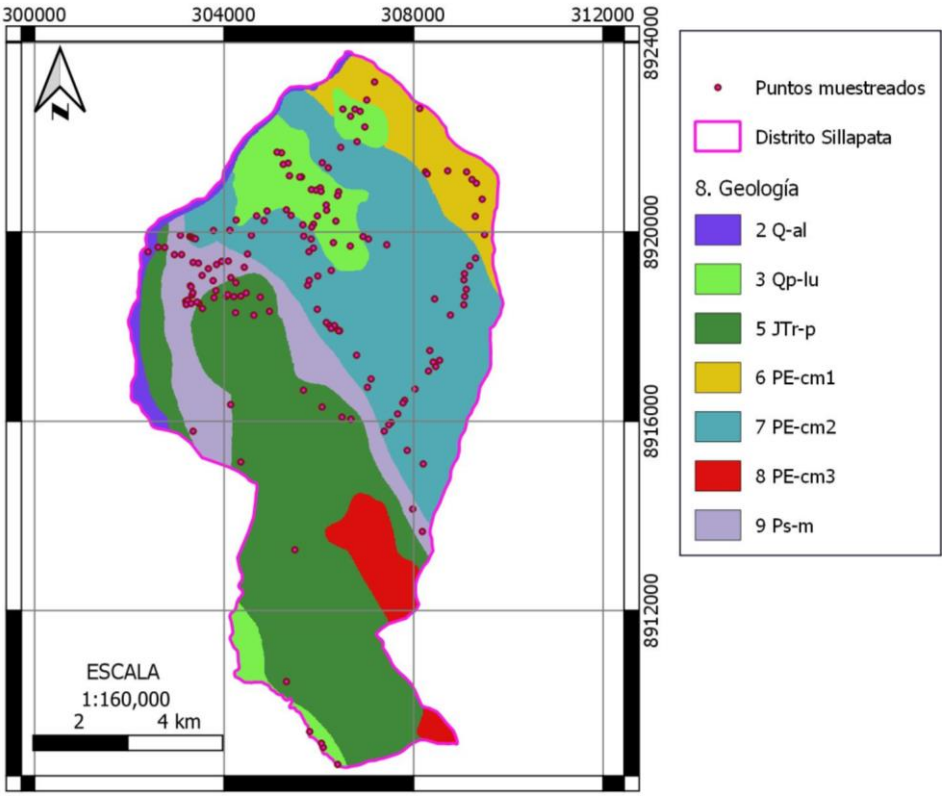


Mapa de la cobertura vegetal del distrito Sillapata

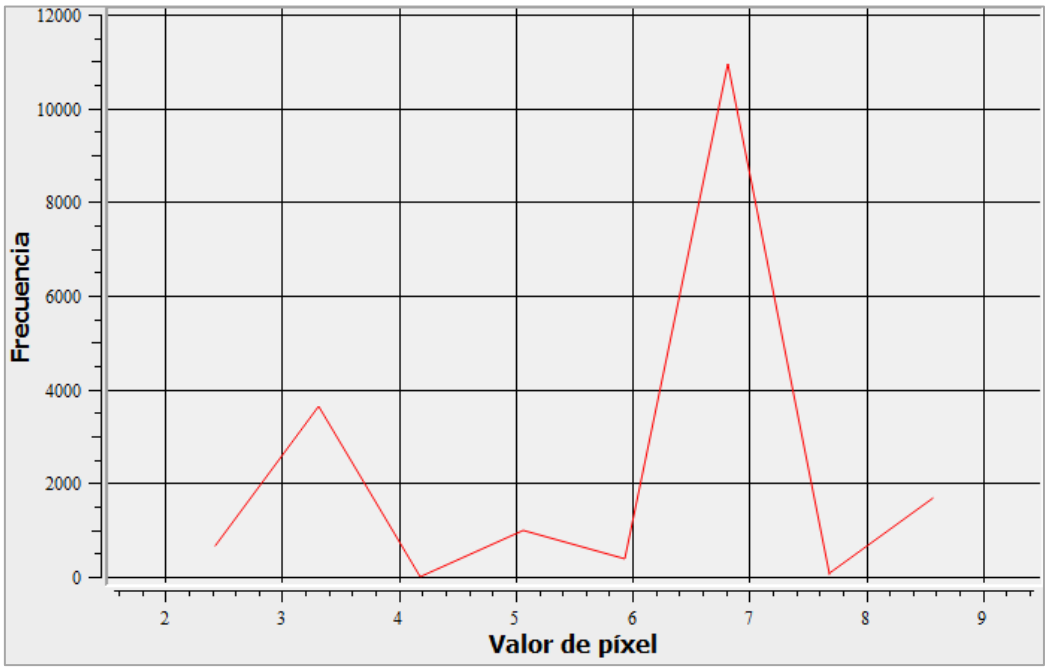


Histograma de valores asignados a la cobertura vegetal en el área de estudio

G.8. Factor de la geología

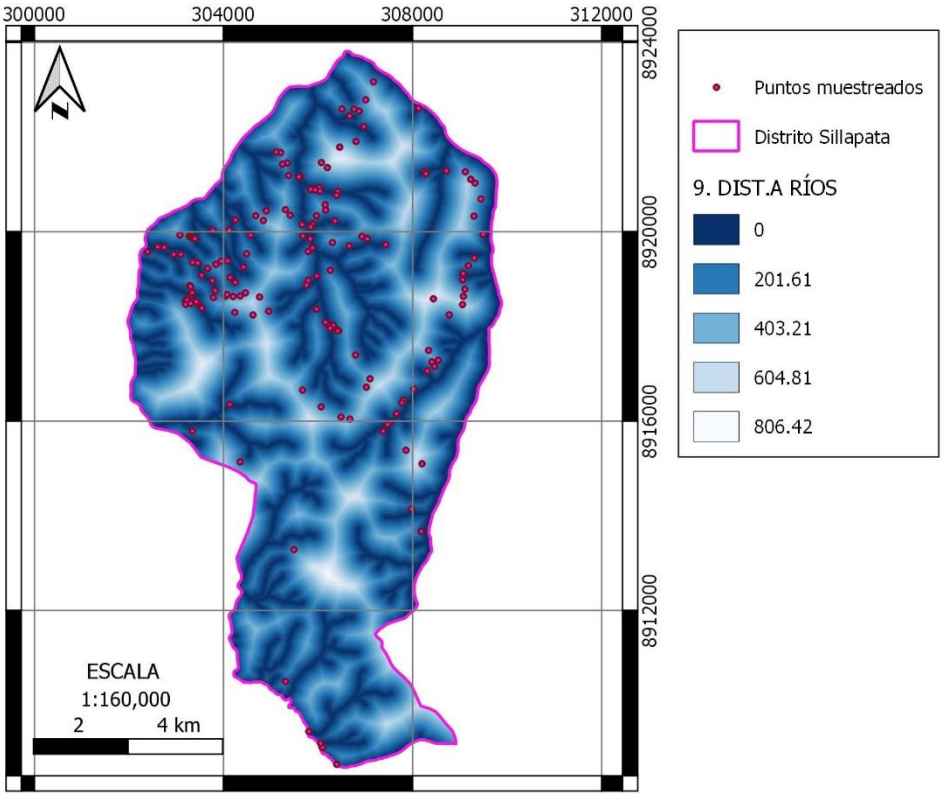


Mapa de formaciones geológicas del distrito Sillapata

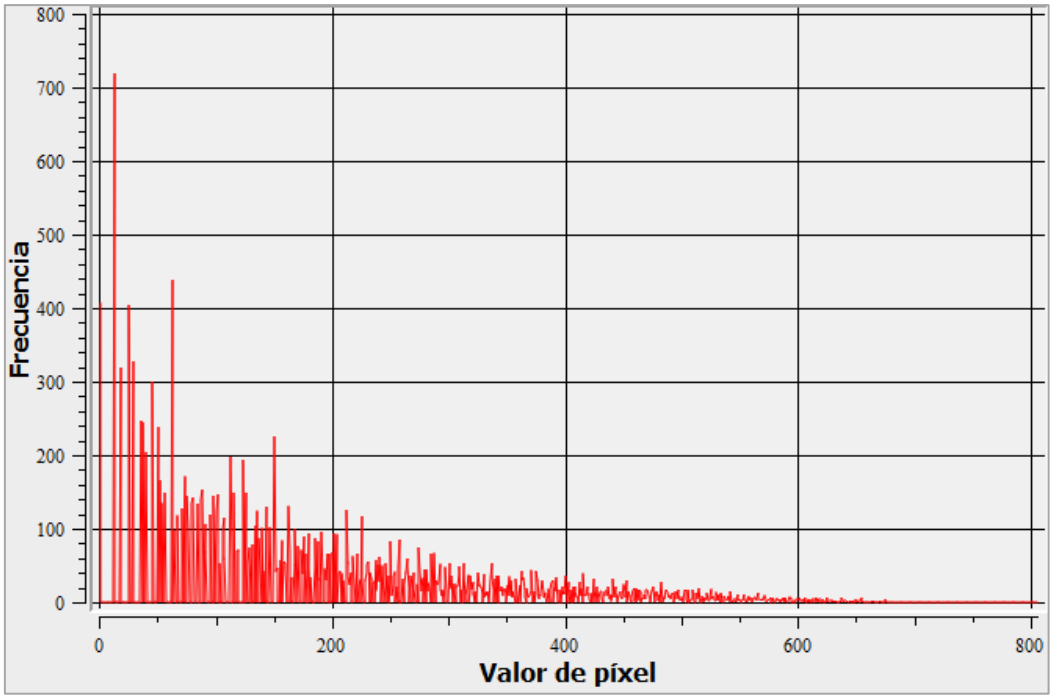


Histograma de valores asignados a formaciones geológicas en el área de estudio

G.9. Factor de distancia a ríos

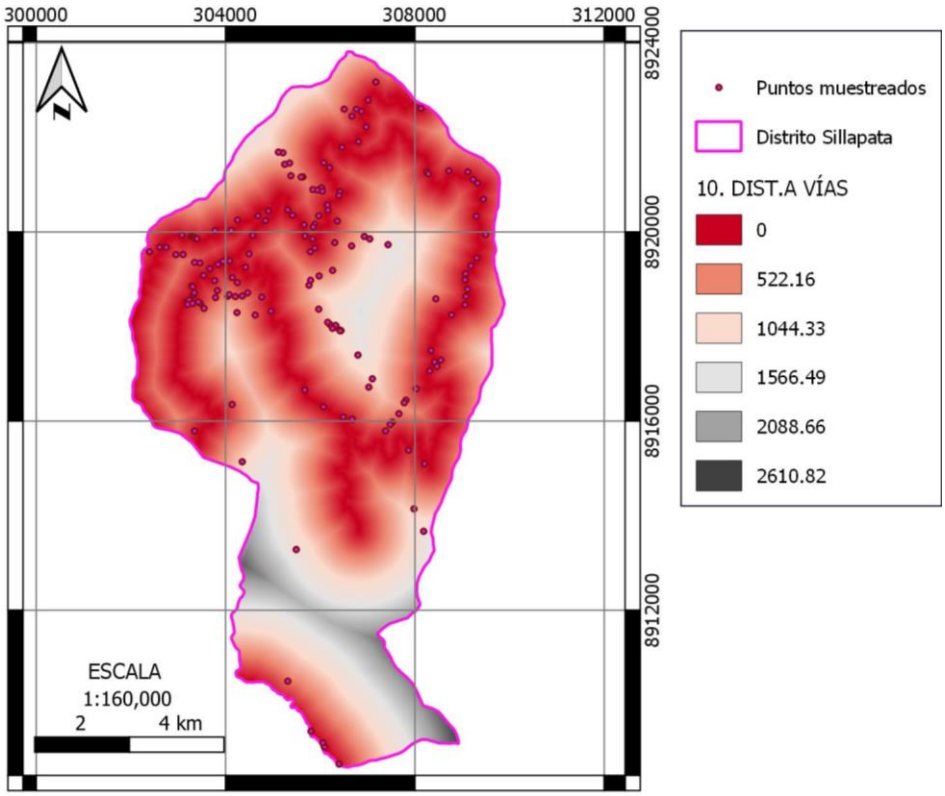


Mapa de distancia a ríos y quebradas del distrito Sillapata

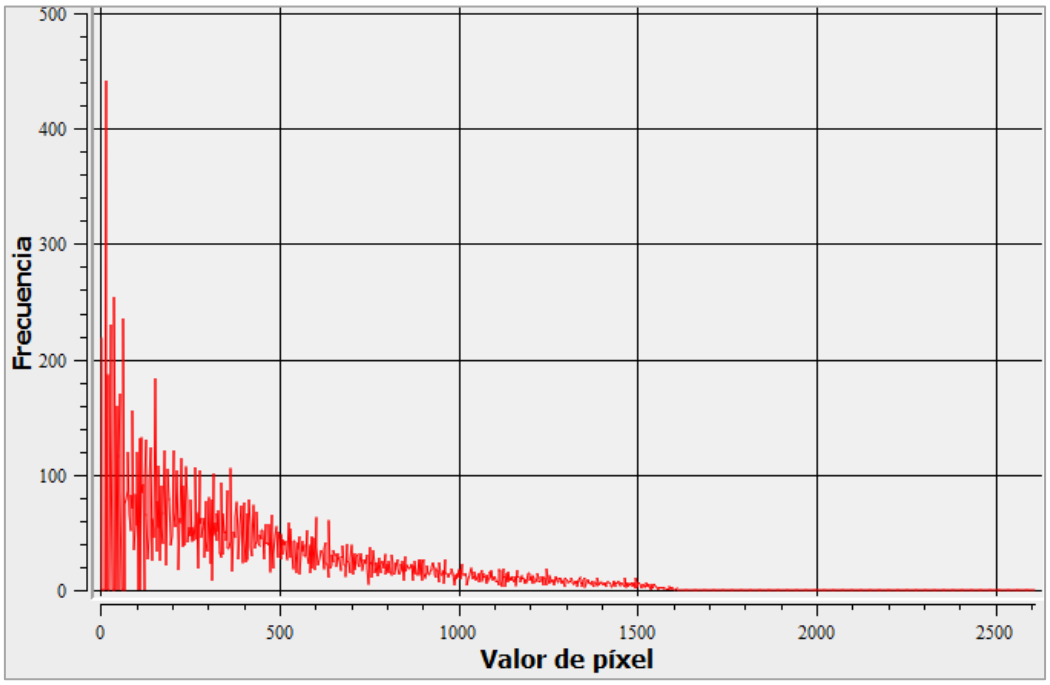


Histograma de valores de distancia a ríos y quebradas en el área de estudio

G.10. Factor de distancia a vías

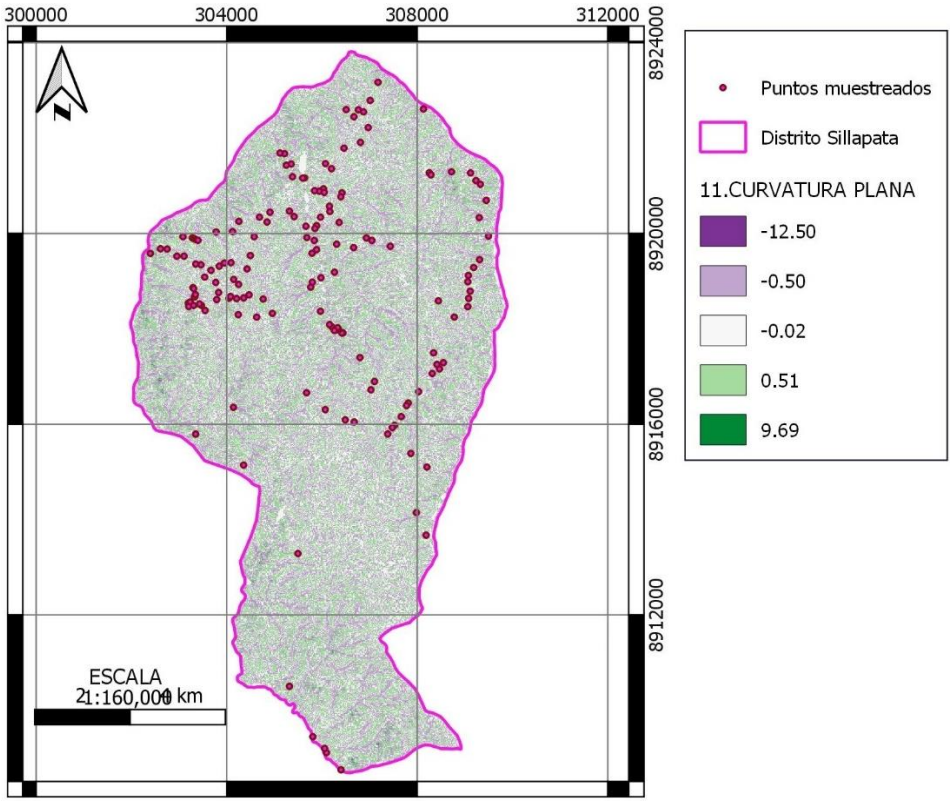


Mapa de distancia a vías del distrito Sillapata

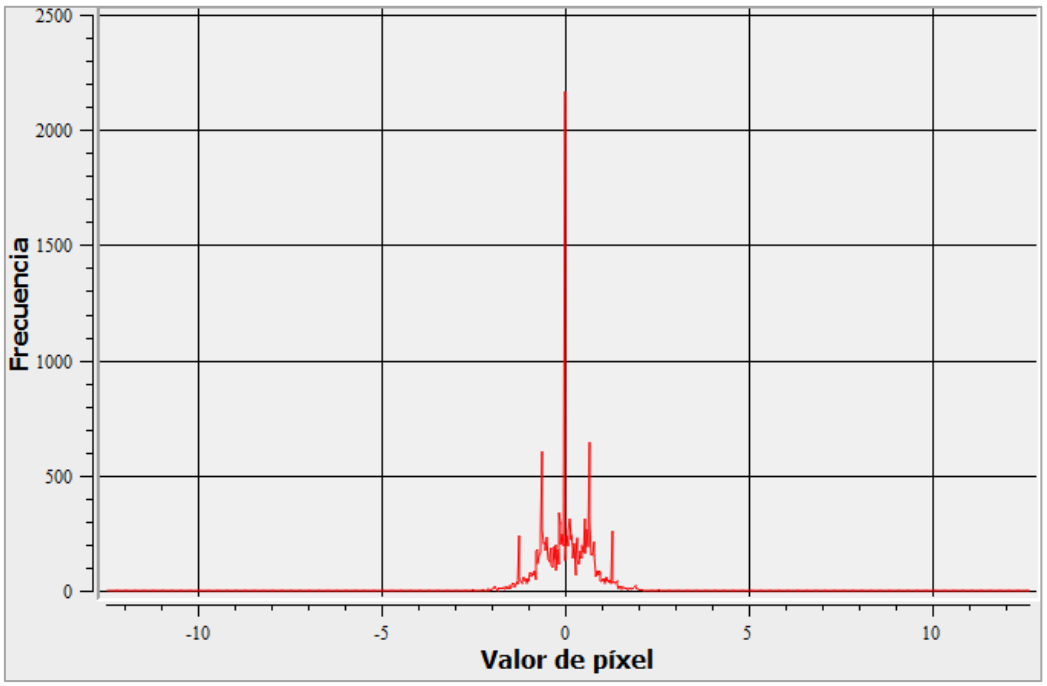


Histograma de valores de distancia a vías en el área de estudio

G.11. Factor de curvatura plana

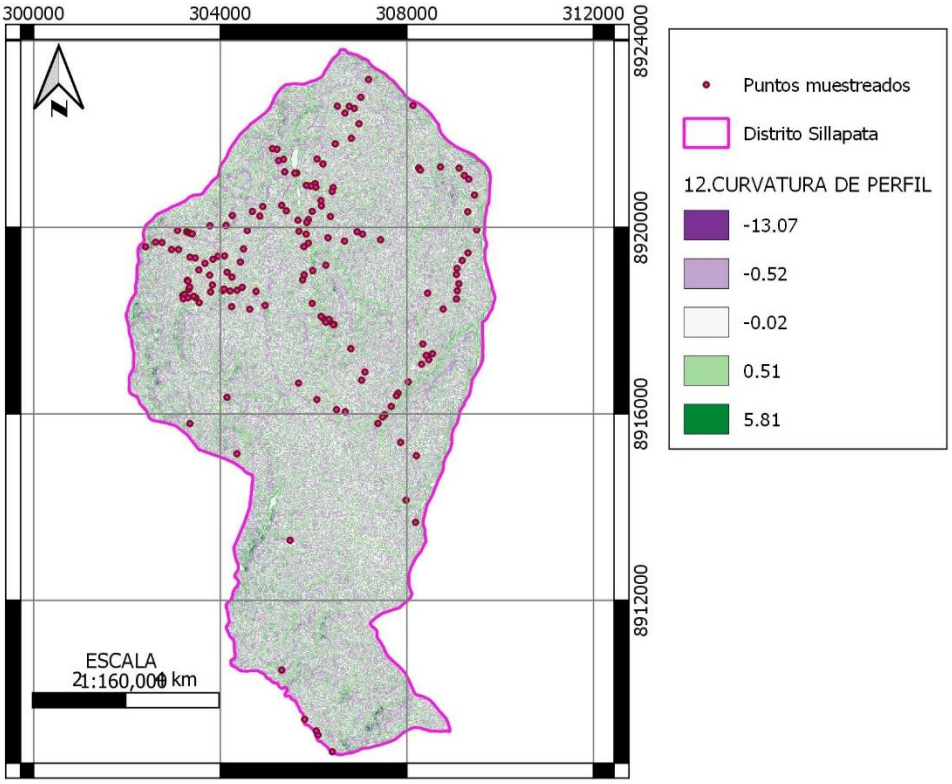


Mapa de curvatura plana del distrito Sillapata

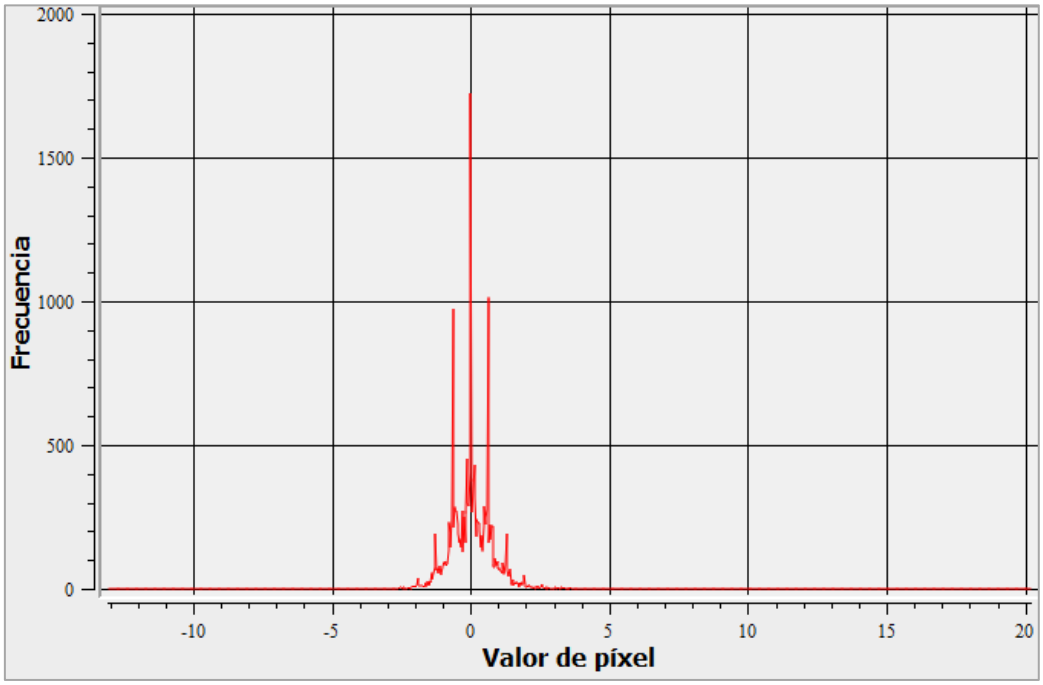


Histograma de valores de curvatura plana en el área de estudio

G.12. Factor de curvatura de perfil



Mapa de curvatura de perfil del distrito Sillapata



Histograma de valores de curvatura de perfil en el área de estudio