

Diseño de geometrías de adquisición para gravimetría y magnetometría enfocado a la exploración de cuerpos intrusivos mediante el uso de Aprendizaje Profundo (Deep Learning) sobre imágenes satelitales

Tesis de posgrado presentada como requisito parcial para optar por el título de:

Magíster en Geofísica

Presentada por:

Juan Sebastián Martínez Acevedo

Director:

Sait Khurama Velásquez
MSC, Geofísica

Co-Director:

Luis Carlos Mantilla Figueroa
Ph.D., Geología

Universidad Industrial de Santander
Facultad de Ciencias
Escuela de Física y Geología
Bucaramanga, Colombia
2022

Dedicatoria

A Dios por la vida.

A mis padres Álvaro y Rocío por su guía y apoyo.

A mi hermano Santiago y demás familiares por su compañía.

A mi novia Juliana por su compañía y apoyo en mis decisiones.

Agradecimientos

Al profesor Sait Khurama por su orientación académica durante el desarrollo de mi tesis y por su colaboración en mi etapa profesional y crecimiento en mis habilidades de enseñanza.

Al profesor Luis Carlos Mantilla por su expedita y oportuna enseñanza sobre la geología de la zona de estudio de esta investigación.

Al profesor Yesid Goyes por su orientación y apoyo en el desarrollo conceptual e investigativo de mi proyecto de grado. Además, agradezco los tiempos brindados de manera oportuna.

A mi amigo y geocientífico Andrés Laverde por sus recomendaciones y tiempo brindado para que el desarrollo computacional implementado lograra buen desempeño.

A todos los profesores de mi etapa formativa y aquellos que influyeron en mi formación.

A la Universidad Industrial de Santander por su programa académico, su plan de educación subsidiado y sus incentivos a través de créditos condonables.

Juan Sebastián Martínez Acevedo

Índice

Introducción	11
1. Objetivos	13
1.1. General	13
1.2. Específicos	13
2. Marco teórico	14
2.1. Estado del arte del uso de Deep Learning en la clasificación de imágenes satelitales	14
2.2. Estado del arte de los métodos de interpolación empleados en las Geociencias . . .	16
3. Zona de estudio	18
4. Planteamiento del problema	20
4.1. Análisis de factores que afectan negativamente las adquisiciones	20
4.1.1. Coberturas vegetales	20
4.1.2. Análisis de terreno	21
4.1.3. Cuerpos de agua	21
4.2. Interpolación espacial de datos puntuales y dispersos	22
4.3. Adquisición o modelamiento de datos	22
5. Metodología y datos	22
5.1. Análisis de factores que afectan negativamente las adquisiciones	23
5.1.1. Coberturas vegetales	23
5.1.2. Análisis de terreno	26
5.1.3. Cuerpos de agua	32
5.2. Métodos de Interpolación	33
5.3. Modelamiento Geofísico	38
6. Resultados	43
6.1. Análisis de los factores que afectan las adquisiciones de gravimetría y magnetometría	43
6.1.1. Coberturas vegetales	43
6.1.2. Análisis de terreno	47
6.1.3. Cuerpos de agua	51
6.1.4. Puntos de adquisición	52
6.2. Implementación de métodos de interpolación	53
6.2.1. Comparativa de los MIE	58
6.3. Adquisición detallada de los métodos de gravimetría y magnetometría	58
7. Discusión	64
7.1. Análisis de factores	64
7.2. Determinación del MIE de mejor desempeño	66
7.3. Diseño de la geometría de adquisición	67
7.4. Flujo de trabajo	69

8. Conclusiones	70
Referencias	72
Apéndices	77
A. Análisis de terreno	77
B. Modelamiento computacional para modelo directo	79
B.1. Modelamiento en función de la densidad	79
B.2. Modelamiento en función de la susceptibilidad magnética	83

Índice de figuras

1.	Mapa geológico de la zona de estudio.	19
2.	División de la imagen satelital en ventanas de clasificación.	27
3.	Modelo de elevación digital.	28
4.	Mapa de pendientes de la zona de estudio.	29
5.	Variables de análisis de terreno.	31
6.	Ubicación de los cuerpos de agua.	32
7.	Ubicación y geología de la zona comparativa en Utah, Estados Unidos.	34
8.	Gravimetría sobre la zona comparativa en Utah, Estados Unidos.	35
9.	Magnetometría sobre la zona comparativa en Utah, Estados Unidos.	35
10.	Semivariograma para datos de gravimetría sobre la zona de interés en Utah.	37
11.	Semivariograma para datos de magnetometría sobre la zona de interés en Utah.	37
12.	Corte geológico del granito de Durania.	39
13.	Vista en planta de la ubicación de los cortes geológicos.	40
14.	Modelo de densidades en 853532 N.	42
15.	Modelo de susceptibilidad magnética en 853532 N.	42
17.	Evolución de la precisión para una RNC de clasificación de cobertura.	44
16.	Modelo de Red Neuronal Convolucional.	45
18.	Evolución de la pérdida para una RNC de clasificación de cobertura.	46
19.	Clasificación de la cobertura del suelo.	47
20.	Modelo de Red Neuronal Artificial Perceptrón Multicapa para análisis de terreno.	49
21.	Exactitud de prueba para el análisis de terreno.	50
22.	Pérdida de prueba para el análisis de terreno.	50
23.	Mapa binario de análisis de terreno.	51
24.	Zonas favorables para toma de datos.	53
25.	Métodos de interpolación en datos de gravimetría sobre la zona comparativa.	56
26.	Métodos de interpolación en datos de magnetometría sobre la zona comparativa.	57
27.	Datos de gravimetría simulados mediante un modelo computacional.	59
28.	Datos de magnetometría simulados mediante un modelo computacional.	60
29.	Resultados de interpolación sobre datos de gravimetría simulados por modela- miento computacional sobre la zona de estudio en Durania.	62
30.	Resultados de interpolación sobre datos de magnetometría simulados por modela- miento computacional sobre la zona de estudio en Durania.	63
31.	Flujo de trabajo	69

Índice de tablas

1.	Longitudes de onda de las bandas del Satélite Sentinel-2.	24
2.	Comparativa entre algunas imágenes representativas del dataset EUROSAT y de la zona de estudio.	26
3.	Descripción para la variables de análisis de terreno.	30
4.	Valores referencia para las propiedades físicas.	41
5.	Resultados de arquitecturas RNC para análisis del terreno.	48
6.	Métricas de error de los MIE.	58
7.	Métricas de error de los MIE para un modelamiento directo.	61

Glosario

Amplitud: en términos de aprendizaje profundo, hace referencia a la cantidad de datos de entrada que se ingresan o que llegan a cada capa.

Árboles: los árboles o árboles de decisión en la técnica de Random Forest son un método de aprendizaje que permite dar un valor de salida a un valor de entrada a través de reglas de decisión simples deducidas a partir de las características de los datos.

Capas Ocultas: son las capas que se encuentran entre la entrada y la salida. Aquí ocurre todo el procesamiento de la red neuronal, partiendo de los datos de entrada y generando un resultado en la capa de salida.

Cross entropy: es una función de pérdida que mide el rendimiento de un modelo de clasificación en un valor de probabilidad, es decir, da un resultado entre 0 y 1.

Dropout: es una técnica que elimina, abandona o ignora neuronas durante el entrenamiento. Este abandono solo ocurre de manera transitoria o temporal y es aleatorio.

Flatten: el flatten es una técnica donde a los resultados del MaxPooling se les elimina una dimensión. En otras palabras, no se pierden datos, solamente se reorganizan en un vector que tiene una dimensión menos que el vector de entrada.

Función de activación: es un concepto que abarca diferentes funciones matemáticas de activación (sigmoide, ReLu, tangente hiperbólica, etc.). Cada una de las funciones permiten dar un valor de salida a los datos de entrada.

Input: es la capa de inicio o de entrada de la red neuronal. Por lo tanto es la que trae la información inicial para su posterior tratamiento en las capas que componen la red neuronal.

MaxPooling: es una operación que agrupa las características más destacadas de cada una de las regio

Nodos: en Random Forest los nodos son los puntos de decisión. Cuando un nodo no tiene ramificaciones se le llama hojas (puntos finales), mientras que si el nodo tiene ramificaciones se les llama vías de decisión.

Output: es la última capa que compone la red neuronal. Por lo tanto, esta capa es la que da

el resultado del proceso de la red. Para el caso de este documento, la red da como resultado una clasificación.

Overfitting: escenario donde la red es capaz de predecir bien las muestras de datos de entrenamiento pero tiene un rendimiento deficiente y no puede generalizar bien en los datos de prueba y de validación.

Poolsize: es el número máximo o número representativo de la zona de característica y es el que asume el MaxPooling durante su proceso.

Profundidad: en términos de aprendizaje profundo, este concepto hace referencia a la cantidad de capas de peso o capas que componen una red. Entre más capas, mayor profundidad para la red.

Red Neuronal Convolutiva (RNC): Tipo de red neuronal que utiliza un algoritmo de unidad para realizar un aprendizaje supervisado sobre un conjunto de datos.

ReLU: la función de activación lineal rectificada es una función por partes que permite asignar un valor de salida a los datos de entrada entre un valor positivo o cero.

RMSprop: es un optimizador (generador de pesos durante la compilación) que busca minimizar el error modificando cada peso en la dirección negativa del gradiente.

Softmax: es una función que convierte un vector de números reales en uno de probabilidades, es decir, un vector donde sus valores suman uno. Los números de entrada pueden ser valores positivos o negativos, pero los de salida serán positivos o cero.

Resumen

Título: Diseño de geometrías de adquisición para gravimetría y magnetometría enfocado a la exploración de cuerpos intrusivos mediante el uso de Aprendizaje Profundo (Deep Learning) sobre imágenes satelitales *

Autor: Juan Sebastián Martínez Acevedo **

Palabras clave: Aprendizaje Profundo, Problema Directo, geometrías de adquisición, interpolación espacial, gravimetría, magnetometría, cuerpos intrusivos, Geofísica, Granito de Durania.

Descripción: En este estudio se presentan geometrías de adquisición enfocadas a la exploración de cuerpos intrusivos y que fueron diseñadas a partir de una aplicación de Aprendizaje Profundo sobre imágenes satelitales. Las geometrías de adquisición fueron trazadas para los métodos de gravimetría y magnetometría en la zona donde aflora el Granito de Durania, ubicado en cercanías al municipio homónimo en Norte de Santander. El diseño partió de un análisis de los factores que afectan las adquisiciones para los métodos geofísicos a implementar. Así, se consideraron las coberturas vegetales, los cuerpos de agua y la topografía como elementos negativos para una adquisición óptima en términos de logística y calidad de datos. Con este análisis se logró una clasificación de favorabilidad para la zona de estudio donde se trazaron transectas y coordenadas de adquisición.

Sin embargo, las geometrías diseñadas suponen datos puntuales y dispersos, generando un reto adicional en la obtención de imágenes geofísicas de calidad. Por esa razón, en esta investigación se implementaron Métodos de Interpolación Espacial (MIE) con el fin de obtener imágenes geofísicas de representativas para la zona de estudio. La determinación del método de mejor desempeño se realizó, en primera instancia, en una zona distinta a la de estudio donde se tuvieran datos de acceso libre. Esto último, se realizó con el fin de hacer una cuantificación de desempeño de cada MIE considerado a través de pruebas remotas y lograr un grado de certeza antes de ejecutar una adquisición sobre la zona de estudio. Así, se implementaron los métodos de Ordinary Kriging (OK), Ponderación de Distancia Inversa (IDW) y Random Forest (RF) sobre una zona en Norte América, donde el método IDW fue el de mejor desempeño de acuerdo a las métricas de error.

Una vez validados los desempeños de los MIE y de tener una geometría sobre la zona de estudio, se consideró una adquisición sobre la zona de estudio que fue desestimada por consideraciones de seguridad. Esto, sumado a que la zona de estudio no contaba con datos geofísicos de acceso libre, llevó a implementar el modelo directo en la investigación mediante la librería SimPEG disponible para Python. En esta herramienta se hizo un modelamiento de la geología de la zona y se asignaron valores a las propiedades geofísicas correspondientes a cada litología representativa. Una vez ejecutado el modelo directo, se obtuvieron datos para las geometrías diseñadas y se implementaron los MIE donde, según las métricas de error, el método IDW fue el de mejor desempeño. Los resultados de la interpolación espacial sobre datos obtenidos por modelamiento directo sugieren una aproximación real a las anomalías geofísicas de la zona. Se sugiere el flujo de trabajo presentado a futuras investigaciones y se recomienda hacer una evaluación de las geometrías y los MIE para un modelamiento inverso.

*Trabajo de grado para optar por el título de Magíster en Geofísica.

**Facultad de Ciencias e Ingenierías Fisicoquímicas. Escuelas de Física y Geología. Director MSc. Sait Khurama Velásquez, Codirector Ph.D. Luis Carlos Mantilla Figueroa.

Abstract

Title: Design of acquisition geometries for gravimetry and magnetometry focused on the exploration of intrusive bodies through the use of Deep Learning on satellite images *

Author: Juan Sebastián Martínez Acevedo **

Key Words: Deep Learning, Forward Problem, Acquisition geometries, Spatial Interpolation, gravimetry, magnetometry, intrusive bodies, Geophysics, Durania Granite.

Descripción:

In this study, acquisition geometries focused on the exploration of intrusive bodies are presented, which were designed from a Deep Learning application on satellite images. The acquisition geometries were traced for the gravimetry and magnetometry methods in the area where the Durania Granite outcrops, located near the municipality of the same name in Norte de Santander. The design was based on an analysis of the factors that led to the acquisitions for the geophysical methods to be implemented. Thus, vegetation cover, bodies of water and topography are considered as negative elements for optimal acquisition in terms of logistics and data quality. With this analysis, a favorability classification was modified for the study area where transects and acquisition coordinates were drawn.

However, the geometries assume punctual and scattered data, revealing an additional challenge in obtaining quality geophysical images. For this reason, Spatial Interpolation Methods (MIE) were implemented in this research in order to obtain representative geophysical images for the study area. The determination of the method with the best performance was carried out, in the first instance, in an area other than the study area where there was free access data. The latter was carried out in order to quantify the performance of each MIE considered through remote tests and achieve a degree of certainty before executing an acquisition on the study area. Thus, the Ordinary Kriging (OK), Inverse Distance Weighting (IDW) and Random Forest (RF) methods were implemented over an area in North America, where the IDW method was the one with the best performance according to the error metrics.

Once the performance of the SIM and having a geometry over the study area had been validated, an acquisition over the study area was considered, which was rejected due to safety considerations. This, added to the fact that the study area did not have free access geophysical data, led to the implementation of the direct model in the investigation using the SimPEG library available for Python. In this tool, a modeling of the geology of the area was made and values were assigned to the geophysical properties corresponding to each representative lithology. Once the direct model was executed, data was obtained for the designed geometries and the MIE were implemented where, according to the error metrics, the IDW method was the one with the best performance. The results of the spatial interpolation on data obtained by direct modeling suggest a real approximation to the geophysical anomalies of the area. The presented workflow is suggested for future research and it is recommended to make an evaluation of the geometries and the SIM for an inverse modeling.

* Research project to opt for the master's degree in Geophysics

** Faculty of Sciences and Physicochemical Engineering. Schools of Physics and Geology. Director MSc. Sait Khurama Velásquez, Co-director Ph.D. Luis Carlos Mantilla Figueroa.

Introducción

En la planeación geofísica se consideran aspectos como las características geológicas de la zona de interés, el cuerpo mineral o componente del subsuelo a caracterizar y los factores condicionantes del terreno que podrían afectar la adquisición de datos. La planeación determinará los métodos a emplear (en el caso de estudio de esta investigación será magnetometría y gravimetría) y el diseño de la geometría de adquisición. Esas consideraciones serán tratadas en esta investigación, donde se analizarán los factores de restricción de adquisición mediante Aprendizaje Profundo y se propondrá un método de interpolación eficaz para datos gravimétricos y magnetométricos de tipo puntual y disperso.

Los factores que pueden condicionar la geometría de adquisición a emplear se pueden analizar mediante un Aprendizaje Profundo (Redes neuronales convolucionales – RNC) sobre imágenes multispectrales del satélite Sentinel-2. Mediante este proceso, se buscará delimitar las zonas más adecuadas para llevar a cabo la toma de datos de gravimetría y magnetometría. El análisis de estos factores (p. ej. Cuerpos de agua, coberturas vegetales, topografía, zonas de acceso etc.) llevan a considerar restricciones en el muestreo, dando lugar a que la adquisición de datos sea de tipo puntual y disperso para los métodos de adquisición geofísica. Por esta razón, es necesario analizar la variación espacial de cada punto utilizando las RNC diseñadas y dar una estimación del nivel de dispersión que tendrían las geometrías planteadas para cada método antes de ir a campo.

Al tratar datos de tipo puntual y disperso, se considera un reto adicional para lograr la caracterización geofísica de una zona debido a la carencia de datos en varios sectores. Además, las áreas de estudio pueden suponer magnitudes extensas, por lo cual es necesario considerar una optimización de recursos económicos y logísticos que complementen la geometría de adquisición de datos propuesta. Por esta razón es necesario considerar técnicas de interpolación espacial de datos (p.ej. Kriging, IDW, Machine Learning) que permitan obtener una caracterización geofísica de alta calidad sobre la zona.

La zona de estudio es el Granito de Durania, ubicado en Norte de Santander (Figura 1). Este cuerpo intrusivo presenta una extensión cercana a los 97 km (Botello-Becerra et al., 2014). Tiene

tres zonas de afloramientos: Norte, centro y sur. Los propósitos de esta investigación se centrarán en la zona Norte, la cual tiene un área cercana a los 20 km². Esta zona se considera atractiva para el desarrollo de la investigación ya que las vías de acceso son escasas, la cobertura vegetal es espesa, los cuerpos de agua requieren definirse con claridad y su topografía tiene una inclinación promedio entre 15 % y 25 %. Estas razones, hacen de esta zona de estudio un punto atractivo y representativo para los objetivos de esta investigación.

De esta forma, el producto final de esta investigación se resume en: 1) el modelo de Aprendizaje Profundo empleado para identificar zonas de restricción de adquisición, 2) geometrías de adquisición para gravimetría y magnetometría en granitos, y 3) un método de interpolación eficaz para los datos puntuales y dispersos de gravimetría y magnetometría. Estos resultados implican un avance en la exploración geofísica de cuerpos intrusivos que se presenten de manera similar al Granito de Durania y pueden hacer parte de los avances que realiza la industria energética en la transición a energías limpias.

1. Objetivos

1.1. General

Diseñar geometrías de adquisición para los métodos de gravimetría y magnetometría enfocadas en la exploración de granitos mediante el uso de Aprendizaje Profundo sobre imágenes satelitales multiespectrales.

1.2. Específicos

1. Analizar mediante Aprendizaje Profundo los factores que afectan las adquisiciones de gravimetría y magnetometría como las coberturas vegetales, los cuerpos de agua y la topografía. Esto se hará para un caso de estudio en particular que será el Granito de Durania, Norte de Santander.
2. Determinar el método de interpolación de mejor desempeño sobre datos gravimétricos y magnetométricos de tipo puntual y disperso.
3. Proponer una adquisición detallada para los métodos de gravimetría y magnetometría sobre el Granito de Durania mediante Redes Neuronales Convolucionales RNC (Aprendizaje Profundo).

2. Marco teórico

2.1. Estado del arte del uso de Deep Learning en la clasificación de imágenes satelitales

El reconocimiento de objetos en imágenes y en vídeo han tenido una gran aceptación en la comunidad científica en los años más recientes debido a que, mediante el uso de métodos de Aprendizaje Automático (Machine Learning), se han logrado avances significativos y aplicables en varios campos de las ciencias, incluidas las geociencias. Dentro de los métodos de Aprendizaje Automático, destaca el Aprendizaje Profundo (Deep Learning) por su uso de redes neuronales artificiales dispuestas en niveles jerárquicos.

Estudios recientes (Krizhevsky et al., 2012, 2017; Simonyan & Zisserman, 2014; Zeiler & Fergus, 2014) han detallado el uso y funcionamiento de las redes neuronales convolucionales (Convolutional Neural Network - CNN) en la clasificación de imágenes y han dado avances relacionados a la estructura del método. El entrenamiento de este método convolucional se lleva a cabo mediante una optimización del objetivo de regresión logística multinomial usando un descenso de gradientes en grupos reducidos (Simonyan & Zisserman, 2014). En otras palabras, estas redes neuronales logran tener menos conexiones y parámetros mediante suposiciones sólidas y, en su mayoría, correctas sobre la naturaleza de la imagen (Krizhevsky et al., 2012, 2017). La capacidad de las redes neuronales convolucionales depende de la arquitectura diseñada, la cual involucra la representación de la profundidad y la amplitud (Krizhevsky et al., 2012, 2017; Simonyan & Zisserman, 2014). De hecho, Simonyan & Zisserman (2014) afirman que la profundidad resulta ser beneficiosa en términos de precisión de la clasificación del reconocimiento de imágenes a gran escala.

Estos avances en la clasificación automática de las imágenes han tenido impacto en aplicaciones relacionadas a las ciencias de la Tierra mediante el análisis de imágenes satelitales multiespectrales. Las aplicaciones se han llevado a cabo en investigaciones como ocurrencia de deforestación e identificación de patrones de uso del terreno (Gardner & Nichols, 2017; Rakshit et al., 2018) donde se requiere generar etiquetados múltiples para las imágenes satelitales. De esta forma, el re-

conocimiento de las zonas de una imagen satelital resulta posible mediante las suposiciones sólidas que se incluyan en el entrenamiento del método.

Por ejemplo, Rakshit et al. (2018) encontraron una correlación relevante entre la agricultura con carreteras, cultivos y sitios habitados. De igual forma, Gardner & Nichols (2017) encontraron, entre otras cosas, las etiquetas y correlaciones apropiadas para clasificar una zona de minería ilegal apropiadamente. Paralelamente, Gadiraju et al. (2020) hacen su análisis de tipos de cultivos (p. ej. Maíz, soya, algodón, etc.) partiendo de la consideración al índice de vegetación de diferencia normalizada (NDVI por sus siglas en inglés) afirmando que, aunque es una herramienta útil para cuantificar y caracterizar la vegetación, no representa un índice representativo en la diferenciación de cultivos, pero si resulta ser un buen parámetro de entrada para que el algoritmo identifique una zona de vegetación. Otras investigaciones, han desarrollado algoritmos enfocados en la identificación de edificaciones (Prathap & Afanasyev, 2018) o incluso en la identificación de múltiples etiquetas como minas ilegales, construcciones de pozos, cuerpos de agua desarrollo de agricultura, deslizamientos de tierra, entre otros (Pritt & Chern, 2017)

Estas investigaciones previas y los resultados presentados en ellas apuntan a un desarrollo amplio de la herramienta de Aprendizaje Profundo en el reconocimiento de imágenes satelitales. Esto lleva a considerar una aplicación eficaz en el reconocimiento de factores de restricción de datos geofísicos, donde se pueden delimitar las zonas donde factores como las coberturas vegetales espesas, los cuerpos de agua y las vías de acceso restrinjan la adquisición de datos geofísicos. Delimitar estas zonas de restricción pueden, por ejemplo, permitir identificar zonas donde la biomasa resulta mayor y, por ende, caracterizarla como una zona de restricción ambiental donde no es posible adquirir un dato geofísico. Con estas zonas de restricción se pueden diseñar geometrías de adquisición apropiadas para el método geofísico a emplear.

2.2. Estado del arte de los métodos de interpolación empleados en las Geociencias

Al tener zonas de restricción en la adquisición de datos geofísicos, se considera que el resultado de la adquisición será datos puntuales y dispersos. Además, adquirir datos geofísicos sobre una zona de gran extensión resulta en una práctica donde no se optimizan los recursos económicos y logísticos. Así, la necesidad de encontrar métodos matemáticos y computacionales que permitan enlazar estos datos se convierte en una prioridad luego de realizar la adquisición correspondiente. Los Métodos de Interpolación Espacial (MIE) han tenido avances considerables en la solución de problemas multidisciplinarios y de retos como los son los datos puntuales y dispersos. Sin embargo, estos esfuerzos tentativos no han llegado a solucionar el desafío totalmente (Chen et al., 2019).

Los MIE incluyen métodos geoestadísticos, determinísticos, estocásticos, interpolantes, no interpolantes, etc. Estos métodos se han llevado a un concepto multidisciplinario y han permitido desarrollar investigaciones en áreas como recursos hídricos, ciencias ambientales, ciencias de la Tierra, agricultura, uso del suelo, probabilidad y estadística, ecología, ingeniería, limnología, hidrocarburos, entre otros (Li & Heap, 2014). En general, y de acuerdo a Li & Heap (2014), la mayoría de los MIE tienen como principio de funcionamiento la siguiente ecuación:

$$\hat{Z}(X_0) = \sum_{i=1}^n \lambda_i Z(X_i) \quad (1)$$

Donde $\hat{Z}(X_0)$ corresponde al valor a estimar en la posición λ_i , n es el número de muestras disponibles, $Z(X_i)$ corresponde al peso asignado al punto de muestra y X_i el valor observado en la posición. En términos generales, el principio de los MIE se basa en la relación entre la distancia entre puntos, es decir, el valor a determinar en una zona no muestreada se verá mayormente afectado por los puntos muestreados más cercanos que por los puntos más lejanos.

Tradicionalmente, los métodos más empleados en la interpolación espacial de datos son los métodos que en este documento se clasifican como clásicos, donde se incluyen Kriging (método geoestadístico), ponderación de distancia inversa (IDW por sus siglas en inglés; método determi-

nístico) y Spline (método estocástico). Algunos autores han desarrollado estudios comparativos entre estos métodos para diferentes casos de estudio y han determinado que los métodos geoestadísticos, usualmente, son los de mejor desempeño (Wang et al., 2017). Esto apuntaría a destacar el Kriging dentro de los métodos clásicos por su desempeño en estudios. Cabe destacar que, en los estudios donde destaca el Kriging, no se señala un reto de datos puntuales y dispersos. Adicionalmente, los métodos de la familia Kriging abarcan varios como el ordinario, el co-kriging, el indicador, entre otros. De esta familia, y de acuerdo a estudios comparativos de esta familia, el de mejor desempeño es el ordinario (OK) (de Amorim Borges et al., 2016).

Otros métodos de interpolación se han desarrollado en años recientes debido al incremento de procesos de aprendizaje automático (Machine Learning). Dentro de los MIE de aprendizaje automático más conocidos destacan el Random Forest (RF) y el Cubist. Similar a los estudios de comparación de desempeño de la familia Kriging, los MIE derivados del aprendizaje automático se han evaluado y se ha señalado al Random Forest como el de mejor desempeño (Houborg & McCabe, 2018).

Por otro lado, el caso particular de los datos a interpolar en este estudio es el de los datos dispersos y puntuales. Este tipo de datos ha sido un desafío constante en la interpolación espacial. Esto ha llevado a que se desarrollen MIE más novedosos que, entre otros factores, han sido impulsados por los avances tecnológicos en términos de computación. Algunos autores se han esforzado por solucionar el desafío que sugiere la interpolación de este tipo de datos y han desarrollado y sugerido métodos y algoritmos. Uno de esos métodos es el de Muestreo Compresivo Bayesiano (BCS), el cual se basa en la suposición de que una señal (p. ej. Propiedad geológica) varía en tiempo-espacio y puede ser reconstruida a partir de información parcial mediante una suma ponderada. (Shi & Wang, 2021; Zhao et al., 2018; Wang et al., 2017). Otro método desarrollado para el caso de datos particular de este estudio es el de inversión basada en la regularización del modelado (SSRI). Este método se basa en la suavización de la densidad de datos a partir de las transformadas dispersas y el rendimiento de comprensión de cada una (p. ej. Fourier, ondícula) (Chen et al., 2019). A pesar de estos avances, estos métodos se han empleado sobre datos 1D (caso del BCS) o solamente

sobre datos provenientes de ciertos métodos (e.g., sísmica, GPR, sismos; caso del SSRI), dejando una consideración adicional a estudios futuros donde se logre implementar estas interpolaciones a otros escenarios como el que se trabajará en esta investigación.

Considerando lo mencionado anteriormente, en este estudio se trabajará con el método OK, considerado como el MIE de mejor desempeño dentro de los métodos clásicos (de Amorim Borges et al., 2016). Además, se utilizarán otros métodos como el IDW, y Random Forest (RF).

3. Zona de estudio

El área considerada como caso de estudio para el tema propuesto es un cuerpo intrusivo en Norte de Santander: Granito de Durania (Figura 1). El sector tiene elementos de interés a solucionar mediante el desarrollo de la investigación: es una zona que cuenta con zonas de difícil acceso, zonas boscosas y de vegetación espesa, topografía irregular y cuerpos de agua (p. ej. Laguna la Barca).

Geográficamente, esta zona está asociada al Macizo de Santander, una gran masa montañosa que se ubica en los departamentos de Santander y Norte de Santander, y que hace parte del Sistema Orogénico Andino en el sector de los Andes del Norte, donde tienen lugar los cuerpos montañosos de Venezuela, Colombia y Ecuador (Mantilla et al., 2011). El área del Macizo de Santander se compone por un grupo de rocas metamórficas precámbricas y paleozoicas, rocas ígneas mesozoicas y rocas sedimentarias del paleozoico superior (Rojas & Cramer, 2013). En esta zona se han reconocido diferentes pulsos magmáticos. Dentro de estos reconocimientos, el más reciente data en el Mioceno (aproximadamente entre 10.9 y 8.4 Ma) y ha sido caracterizado con composiciones riódacíticas, dacíticas y cuarzomonzoníticas (Mantilla et al., 2009, 2011). El basamento metamórfico del Macizo de Santander se presenta intruido por batolitos del Grupo Plutónico de Santander y granitoides del Silúrico como el Granito de Durania (Alarcón et al., 2020).

Este cuerpo intrusivo, caracterizado como un granito blanco moscovítico por Ward et al. (1973), se originó como un producto de la inyección de magmas de origen mantélico en la corteza, donde además ocurrieron procesos de diferenciación magmática y contaminación cortical. Esta etapa de formación se dató mediante el método geocronológico U-Pb en circones y se determinó una edad

de 442.6 Ma (+7.4 / -6.0 Ma), correspondiente al Silúrico (Botello-Becerra et al., 2014). Posterior a esta etapa de formación, este cuerpo intrusivo tuvo emplazamiento, según Rojas & Cramer (2013), en el Cretácico.

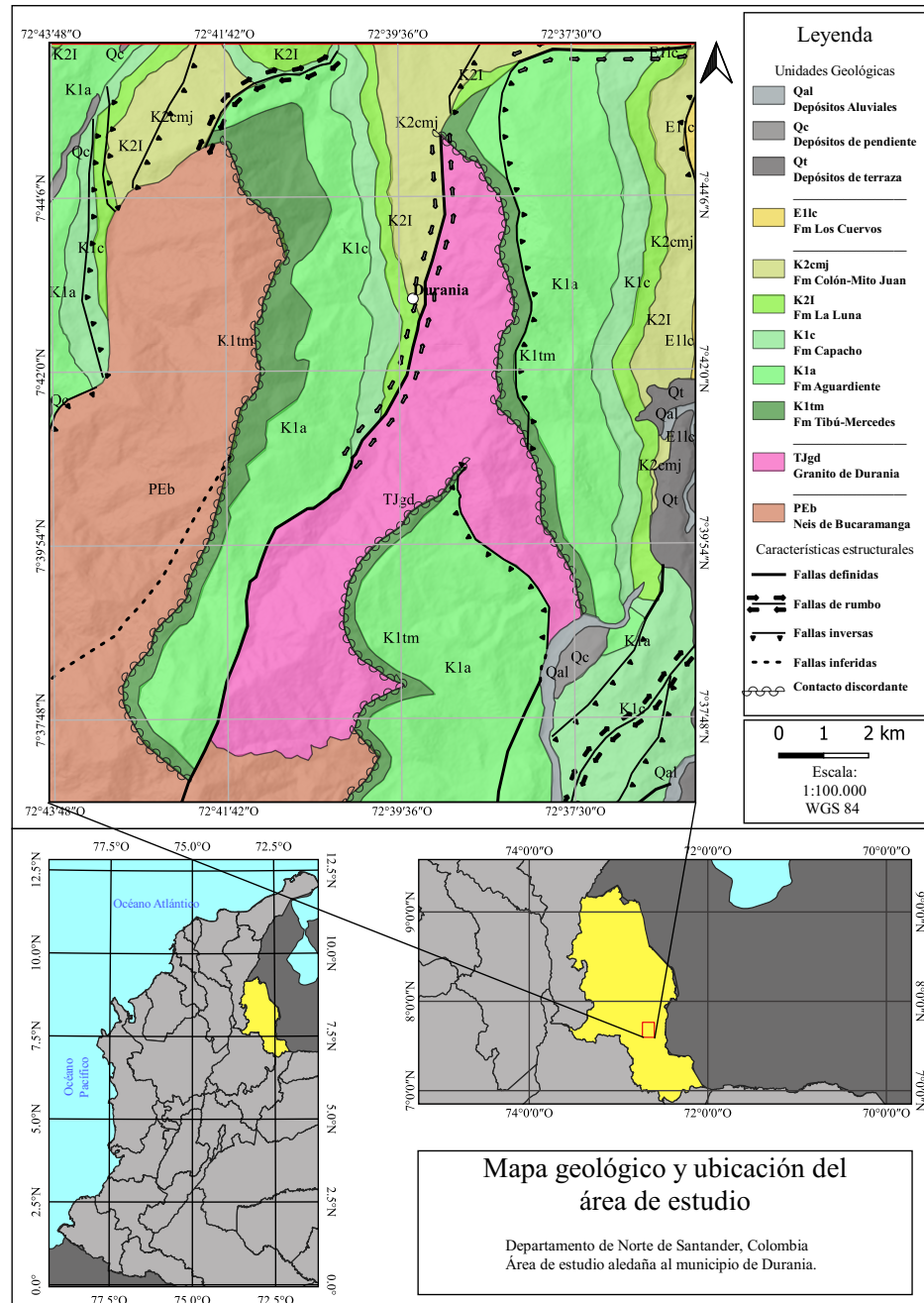


Figura 1: Mapa geológico de la zona de estudio.

Área de estudio en cercanías al municipio de Durania, Norte de Santander. El área tiene una extensión de 269.8 km². Adaptado de la plancha 98 del Servicio Geológico Colombiano (Fúquen et al., 2011).

La extensión de Granito de Durania se presenta dividida en tres zonas (norte, centro y sur) que abarcan una longitud de 97 km (Botello-Becerra et al., 2014). La zona norte se ubica en cercanías al municipio de Durania, la zona sur a los municipios de Pamplona y Cácuta, y la zona central se encuentra en una ubicación en medio de las dos zonas anteriores, en cercanías a la vereda de San Miguel Alto. De estas tres ubicaciones, este estudio se centrará en la zona norte (Figura 1), la cual está en contacto con rocas sedimentarias del Cretácico (Formaciones Colón Mito-Juan, Aguardiente, Tibú-Mercedes y rocas metamórficas del Proterozoico (Neis de Bucaramanga).

4. Planteamiento del problema

4.1. Análisis de factores que afectan negativamente las adquisiciones

4.1.1. Coberturas vegetales

Dentro de los factores que afectan las adquisiciones geofísicas están las coberturas vegetales que, a grandes rasgos, se pueden clasificar como espesas o ligeras. Un ejemplo de una cobertura vegetal espesa es un bosque, mientras que para una ligera la referencia es una zona de pastos o cultivos.

El principal efecto que tienen las coberturas vegetales sobre las adquisiciones es la localización de los puntos mediante el posicionamiento GPS. El efecto que tienen un bosque o una zona rodeada de árboles es la generación de multitrayectos de la señal de posicionamiento, lo que eventualmente resultará en la principal fuente de error del equipo. Teniendo en cuenta esto, y que se desea diseñar una geometría de adquisición con una etapa de precampo, es importante delimitar las zonas de bosque que afectarían el posicionamiento del equipo GPS que está incorporado al magnetómetro y el GPS adicional que se planea usar para la adquisición de gravimetría.

Otra de las consideraciones a tener en cuenta con las coberturas vegetales es el acceso. Es posible encontrarse con coberturas vegetales de difícil acceso o de acceso restringido como zonas

de cultivo. Por esa razón, se consideran este tipo de zonas como posibles puntos de adquisición, pero con la eventualidad de encontrarse con un acceso restringido.

4.1.2. *Análisis de terreno*

Dentro de las consideraciones a tener en cuenta durante el análisis de factores que afectan las adquisiciones, también se considera importante los efectos asociados al terreno. Dentro de este análisis surge un factor relevante en la investigación: la topografía. Se puede decir que es un factor que puede dificultar el acceso a un punto de medición por pendientes pronunciadas. Además, la topografía es un efecto importante a tener en cuenta en el método de gravimetría debido a que el relieve pronunciado cercano a un punto de medición (colinas o valles), puede suponer un efecto en la medición que llega, aunque en raras ocasiones, a superar 1 mGal.

Adicional a esto, se considera que topografías inestables o muy pronunciadas pueden implicar un incremento en los tiempos empleados en la adquisición, afectando directamente la logística de un estudio en zona. Por estas razones, se considera necesario tener presente este factor al momento de diseñar las geometrías en conjunto a los demás factores.

4.1.3. *Cuerpos de agua*

En cuanto a los cuerpos de agua, como ríos y lagos, es necesario considerar la capacidad de los equipos disponibles y los efectos que pueden tener respecto a cuerpos de agua. Por ejemplo, cuando se hacen adquisiciones de gravimetría sobre lagos, se tiene que hacer una consideración especial a la corrección por placa de Bouguer cambiando la densidad de roca por la densidad menos la densidad de agua ($\rho - 1000 \text{ kg m}^{-3}$). Además, es necesario conocer las profundidades estimadas de estos cuerpos de agua para poder corregir la medición. De esta forma, y considerando que la disponibilidad de equipos se limita a adquisiciones terrestres, las zonas muy cercanas o sobre los cuerpos de agua no se tendrán en consideración para evitar cualquier efecto a la medición que pueda incluirse.

4.2. Interpolación espacial de datos puntuales y dispersos

Limitar las zonas de adquisición en función de los factores mencionados anteriormente y de los recursos económicos y logísticos, conlleva a tener un caso de datos puntuales y dispersos. Este tipo de datos supone un desafío en la obtención de imágenes geofísicas de calidad.

A pesar de que algunos estudios se han centrado en solucionar este problema (Chen et al., 2019; Wang et al., 2017), solo lo han hecho para geofísica geofísica 1D. Esto hace que los retos de esta investigación se centren en la implementación de los métodos tradicionales (OK y IDW) y algunos más recientes surgidos del Machine Learning (RF). Con esto se busca que, a partir de las geometrías diseñadas con el análisis de factores, se logre implementar los métodos tradicionales y de acceso libre para generar imágenes geofísicas representativas para los casos de estudio que involucren datos puntuales y dispersos.

4.3. Adquisición o modelamiento de datos

Los propósitos de esta investigación se centran, en términos generales, en el diseño de las geometrías de adquisición. Así, es necesario considerar una evaluación de estas geometrías a partir de adquisiciones geofísicas en campo. Sin embargo, también se hace una consideración al trabajo remoto ante cualquier eventualidad que pueda restringir una adquisición. Es por eso que las geometrías que se diseñarán no apuntan a una ejecución o evaluación única sino que se plantean con el fin de permitir una ejecución a través del modelo por problema inverso o por el problema directo.

5. Metodología y datos

En esta sección se describirán y abordarán las etapas llevadas a cabo para solucionar cada uno de los planteamientos problema. Además, se detallará el origen de los datos tratados para cada una de las etapas. Respecto a esto último, cabe mencionar que esta investigación se centró en el uso de

datos y recursos de acceso libre o, en su defecto, hechos por el autor.

5.1. Análisis de factores que afectan negativamente las adquisiciones

5.1.1. Coberturas vegetales

Con el fin de analizar las coberturas vegetales o el uso del suelo sobre la zona de estudio, se realizó una clasificación del uso del suelo mediante Aprendizaje Profundo. Para esto, se implementaron 4 etapas: 1) obtención de imágenes satelitales sobre la zona de estudio, 2) obtención de un set libre de datos útiles para el entrenamiento, 3) ajuste de la imagen satelital en función del grupo de datos de entrenamiento, y 4) diseño de una arquitectura de Aprendizaje Profundo que se entrene a partir del set de datos obtenidos en el anterior paso y que clasifique las imágenes de la zona de estudio. Esta cuarta etapa se abordará en la sección de resultados.

Etapas 1: Imágenes satelitales

Las imágenes satelitales empleadas en esta investigación corresponden a la misión Sentinel-2 ESA (imágenes son cortesía del Servicio Geológico de Estados Unidos, USGS). Este proyecto comprende de dos satélites idénticos dispuestos en órbita polar en fases de 180° entre ellos, donde el objetivo principal es monitorear la variabilidad de las condiciones de superficie terrestre y costeras. Las imágenes adquiridas por esta misión brindan diferentes resoluciones de acuerdo a las longitudes de onda de cada una de las trece bandas que lo componen y que se detallan en la Tabla 1.

Así, se obtuvieron las imágenes del satélite Sentinel-2 desde el portal *Earth Explorer* del Servicio Geológico de Estados Unidos (USGS). Las necesidades de la investigación implican una fecha de adquisición relativamente reciente y de buena visibilidad dado que el porcentaje de cobertura de nubes puede ser un factor adverso en la clasificación de la cobertura del suelo. De esta forma, la adquisición de la misión durante el 4 de febrero de 2021 se presentó como la mejor dentro del rango de fechas estipulado debido a que cuenta con una cobertura de nubes menor al 3 %. Esta imagen fue recortada de acuerdo a la región de interés y las dimensiones fueron de 1472 x 1856 x 3, siendo la última dimensión las 3 capas RGB. Estas dimensiones equivalen a 14.713 x 18.553 m.

	Longitud de onda (μm)	Resolución (m)
Banda 1 - Aerosol	0.43 - 0.45	60
Banda 2 - Azul	0.45 - 0.52	10
Banda 3 - Verde	0.54 - 0.57	10
Banda 4 - Rojo	0.65 - 0.68	10
Banda 5 - Red edge 1	0.69 - 0.71	20
Banda 6 - Red edge 2	0.73 - 0.74	20
Banda 7 - Red edge 3	0.77 - 0.79	20
Banda 8 - Infrarrojo cercano (NIR) 1	0.78 - 0.90	10
Banda 8A - Infrarrojo cercano (NIR) 2	0.85 - 0.87	20
Banda 9 - Vapor de agua	0.93 - 0.95	60
Banda 10 - Cirrus	1.36 - 1.39	60
Banda 11 - SWIR 1	1.56 - 1.65	20
Banda 12 - SWIR 2	2.10 - 2.28	20

Tabla 1: Longitudes de onda de las bandas del Satélite Sentinel-2.

Etapas 2: Set de datos de entrenamiento

Se consideró el dataset de acceso libre del proyecto EUROSAT (Helber et al., 2018, 2019). En este proyecto, se analizaron y clasificaron imágenes del Sentinel-2 adquiridas sobre Europa. Este proyecto cuenta con dos conjuntos de datos: 1) RGB, donde están las bandas 4, 3 y 2 codificadas como imágenes jpeg y 2) ALL, donde están las 13 bandas en el rango de valor original (float32). Para los propósitos de esta investigación se trató solo la combinación RGB cuyo dataset cuenta con 27.000 imágenes georreferenciadas y etiquetadas dentro de 10 clases: cultivo anual, bosque, vegetación herbácea, carretera, industrial, pasto, cultivo permanente, residencial, río y mar-lago. Este conjunto de datos cuenta con unas dimensiones individuales para cada imagen de 64 píxeles de ancho x 64 píxeles de alto x 3 capas de profundidad (RGB).

Es necesario destacar que, al no encontrar un set de datos libres con etiquetas para imágenes del Sentinel-2, la diferencia entre el set de datos EUROSAT y la zona de estudio se considera una limitante en esta investigación. Teniendo en cuenta que el set de datos corresponde a una latitud distinta se pueden encontrar variaciones ecológicas, climáticas, ambientales, tipo de flora, entre otros. Así, de las clases con las que cuenta el proyecto EUROSAT, solo se consideraron las de interés para los propósitos de esta investigación. Esto permitió que el set de datos se redujera

en tamaño ya que se desestimaron las etiquetas industrial y carreteras (esta última debido a las diferencias de infraestructura). De esta forma se contó con 22000 imágenes distribuidas de la siguiente manera: 3000 de cultivo anual, 3000 de bosque, 3000 de vegetación herbácea, 2000 de pastos, 2500 de cultivos permanentes, 3000 de residencial, 2500 de ríos y 3000 de mar-lago.

En la Tabla 2 se observan algunas imágenes comparativas entre el set de datos clasificado de Europa y algunas imágenes clasificadas manualmente dentro de las etiquetas de interés con el fin de hacer una observación comparativa. Cabe mencionar que, para esta comparación, se garantizó que el tamaño de las imágenes de la zona de estudio fueran iguales a las de el set de datos clasificado, es decir 64 x 64 x 3. En general, la comparación entre los sectores se considera, a priori, como útil para los propósitos de la investigación.

Una vez entrenada y ejecutada la RNC en la sección de resultados, se podrán considerar los efectos que tuvieron las diferencias entre los grupos de datos. Sin embargo, desde esta observación inicial, se destaca la diferencia de anchura de los ríos. Además, el tamaño de las lagunas, o cuerpos de agua presentes en la zona de estudio, se considera menor a la de la etiqueta de entrenamiento. Esto sugiere, en principio, que es necesario abordar la presencia de cuerpos de agua desde otra perspectiva que se mencionará adelante en la sección de metodología para este factor.

Etapas 3: Ajuste de los grupos de datos

Con el fin de que la imagen satelital de la zona de estudio pudiera ser clasificada, se dividió la imagen sobre la zona de interés en una grilla compuesta por ventanas de análisis. Se garantizó que estas ventanas contaran con la misma dimensión que el grupo de datos de EUROSAT. Así, las dimensiones resultaron en 64 x 64 x 3 (ventanas cuadradas de aproximadamente 640 m de lado). De esta forma, la imagen que cubre la zona resultó en una grilla de 23 x 29 para un total de 667 imágenes a clasificar con la red convolucional previamente entrenada (Figura 2).

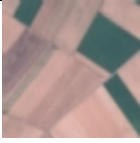

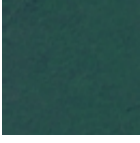

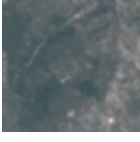

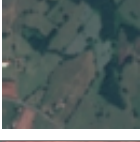

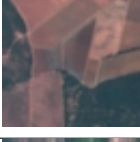
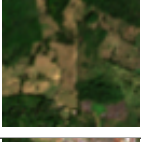
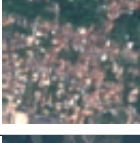

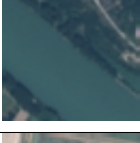
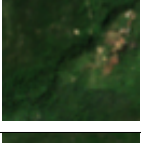
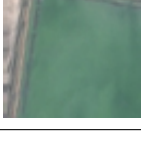

Etiqueta	EUROSAT	Zona de Estudio
Cultivo anual		
Bosque		
Vegetación herbácea		
Pasto		
Cultivo permanente		
Residencial		
Río		
Mar - lago		

Tabla 2: Comparativa entre algunas imágenes representativas del dataset EUROSAT y de la zona de estudio.

5.1.2. *Análisis de terreno*

En cuanto al análisis de terreno, se considera que la zona tiene un interés para el caso de estudio debido a que cuenta con picos, altos, lomas, cuchillas, entre otros (Figura 3). Así, para abordar el factor correspondiente a los efectos del terreno se realizó: 1) clasificación binaria sobre un mapa de pendientes generado desde el modelo de elevación digital, 2) eliminación del sesgo

univariable, 3) extracción de datos de entrenamiento y 4) clasificación a partir de una Red Neuronal Artificial Perceptrón Multicapa (RNAPM). Esta última etapa se presenta en la sección de resultados correspondiente al tema de análisis de terreno.

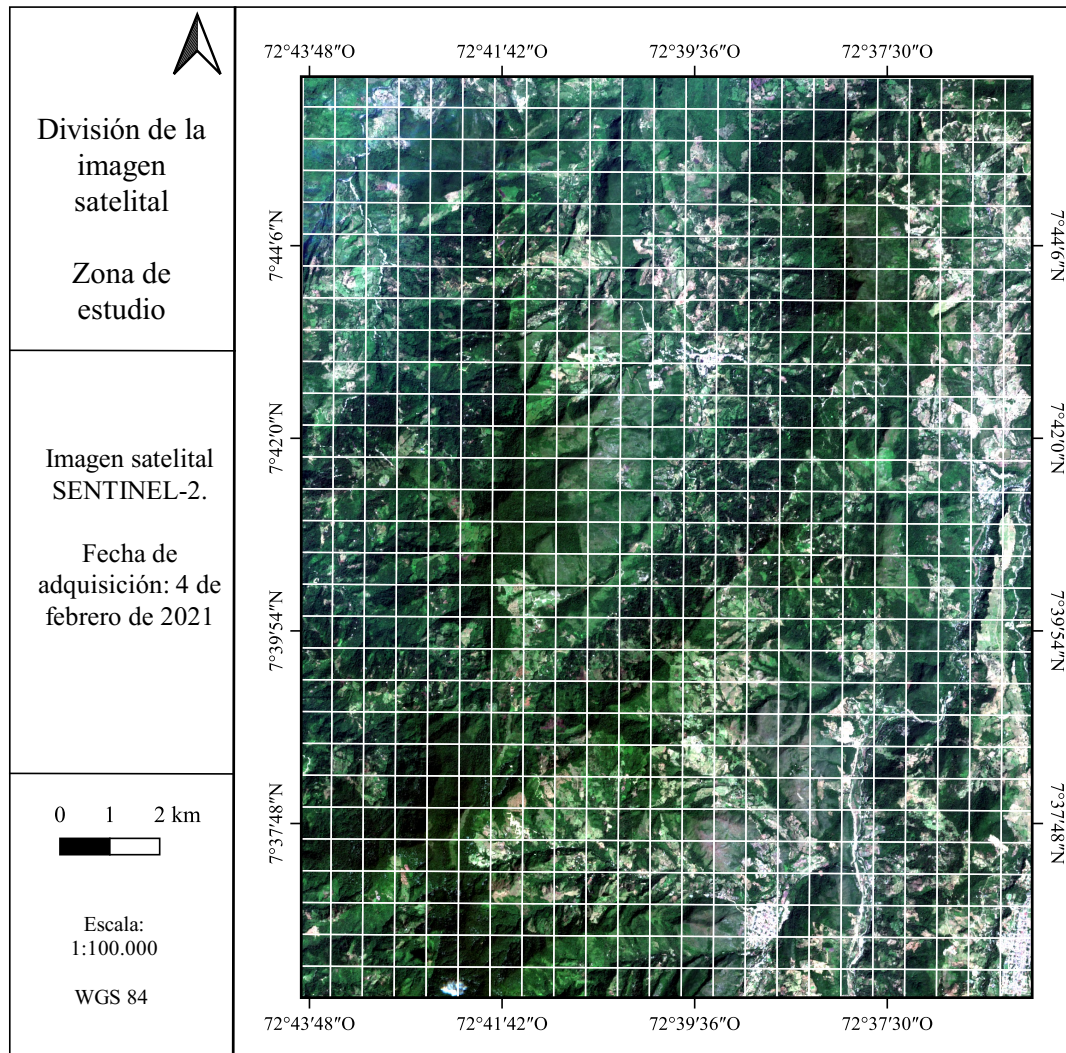


Figura 2: División de la imagen satelital en ventanas de clasificación.
Las ventanas de clasificación de borde blanco tienen dimensiones de 64 x 64 x3, dando un total de 667 ventanas.

Etapa 1: clasificación binaria para un mapa de pendientes

Para realizar el análisis de terreno, se obtuvo el modelo de elevación digital (DEM), el cual es producto del proyecto ALOS-PALSAR liderado por la Agencia Japonesa de Exploración Aeroespacial. Este proyecto consistió en un satélite (ALOS) lanzado en 2006 que, mediante su sensor

PALSAR (Phased Array type L-band Synthetic Aperture Radar), recolectó imágenes de radar en escenas de 50 x 70 km en todo el planeta en períodos de 45 días hasta el año 2011.

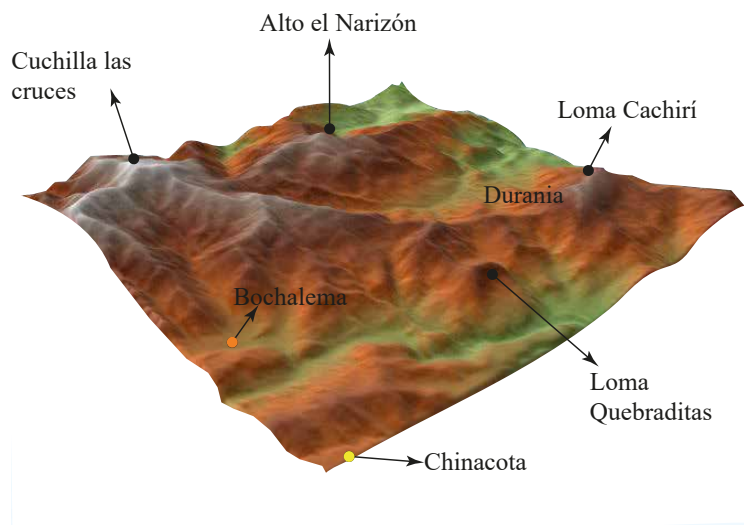


Figura 3: Modelo de elevación digital.

Modelo tridimensional de la elevación presente en la zona de estudio para un área de 269.8 km². La exageración vertical se estableció en 1.5x para visualizar mejor las orografías señaladas. La información georreferenciada se obtuvo del portal Colombia en Mapas perteneciente al Instituto Geográfico Agustín Codazzi (IGAC). Las planchas (de escala 1:25.000) consideradas fueron: 98IB, 98ID, 98IIA, 98IIC, 98IIIB y 98IVA.

A partir del DEM y gracias a los algoritmos fue posible calcular el mapa de pendientes asociado a la zona de estudio. Sobre este mapa de pendientes se realizó una clasificación binaria que se desarrolló de la siguiente manera: 1) se distribuyeron 500 puntos de manera aleatoria, 2) se clasificaron con un valor de 1 los puntos que, por su rango de pendiente (menor a 22°), se consideran óptimos para la toma de datos de buena calidad y 3) se clasificaron con un valor de cero los puntos que se consideran como no válidos para la toma de datos por el efecto que tendría la pendiente en el dato.

En la Figura 4 se evidencian los rangos de pendientes elevadas en tonalidades rojizas, mientras que los rangos de pendientes inferiores se ven en tonalidades verdes. Además, se evidencia el resultado de la clasificación binaria (1: puntos blancos, 0: puntos negros).

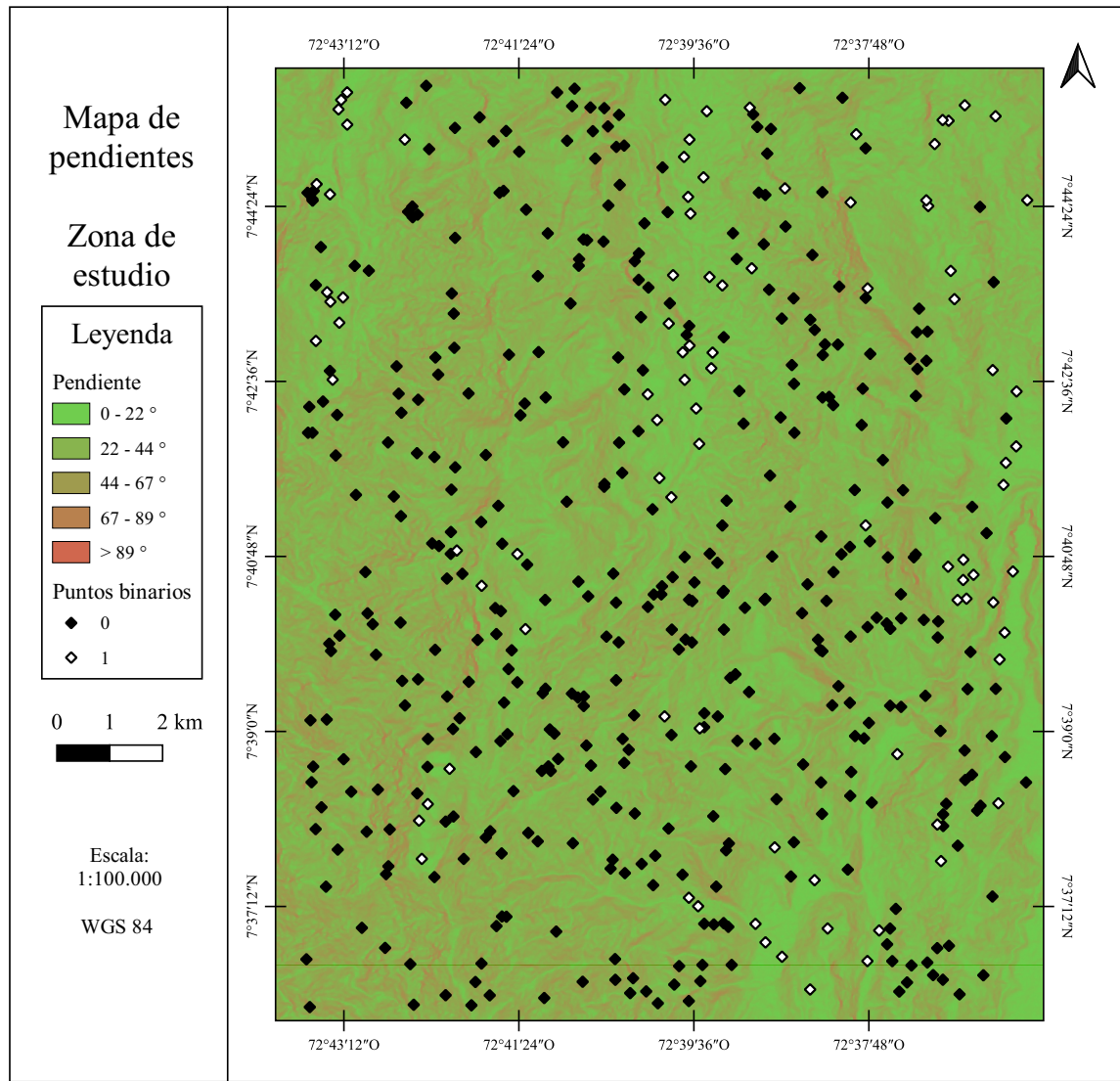


Figura 4: Mapa de pendientes de la zona de estudio.

Los rangos de pendientes fueron obtenidos mediante la herramienta de clasificación por cuantiles disponible en el software QGIS.

Etapla 2: Eliminación del sesgo univariable

Con el fin de eliminar el sesgo que provoca una clasificación binaria guiada a partir de una sola variable, se consideró incluir más variables del terreno que permitieran ampliar el espectro de la clasificación binaria. De esta forma, a través de QGIS y sus algoritmos se calcularon derivadas verticales y horizontales de primer y segundo orden con el fin de analizar el modelo de elevación digital de manera más completa. Así, mediante la herramienta *r.slope.aspect* de GRASS en QGIS

se logró calcular los mapas asociados a cada una de las siguientes variables: rugosidad, orientación, curvatura de perfil y curvatura tangencial. En la Tabla 3 se observa la descripción de cada una de las variables consideradas para el análisis de terreno, en la Figura 5 se observan los valores de estas variables sobre la zona de interés y en el Apéndice A.1 se observa la correlación entre las variables.

Variable	Tipo	Descripción
Pendiente	Derivada vertical de primer orden	Ángulo de inclinación respecto a la horizontal. Los valores están en el rango de 0-90°.
Rugosidad	Medida estadística	Grado de irregularidad de la superficie. A mayor valor, mayor grado de irregularidad.
Orientación	Derivada de primer orden	Dirección de la brújula asociada a un punto sobre la pendiente. Los valores están en el rango de 0-360, donde el punto de partida es el N
Curvatura de perfil	Derivada de segundo orden	Es la forma (cóncava-/convexa+) de la superficie en la vecindad inmediata del punto de muestra contenido dentro del plano vertical
Curvatura tangencial	Derivada de segundo orden	Es la forma (cóncava-/convexa+) asociada a un punto de la superficie respecto al plano entre la curvatura de perfil y la curvatura de plano (horizontal).

Tabla 3: Descripción para la variables de análisis de terreno.

Etapas 3: extracción de datos

Una vez calculados los mapas de variables, a través de la herramienta *Point Sampling tool* se obtuvieron los valores asociados a los puntos binarios distribuidos aleatoriamente en la etapa anterior. Con esto se generó un archivo de texto donde se incluyeron los valores asociados punto a punto para cada una de las variables.

Una vez extraídos los valores de estas variables para los 500 puntos, se agregó un nuevo condicionante para la pendiente con el fin de restringir más la clasificación: zonas de pendiente menor a 10 grados para poder clasificarse con un valor de 1. De igual forma, se eliminaron los datos vacíos que corresponden a los puntos aleatorios que no registraron una o más de una de las variables. Con estos datos es posible entrenar una Red Neuronal Artificial Perceptrón Multicapa (RNAPM) que, a

partir de los 500 puntos aleatorios y los valores de las variables, clasifique toda la zona de estudio.

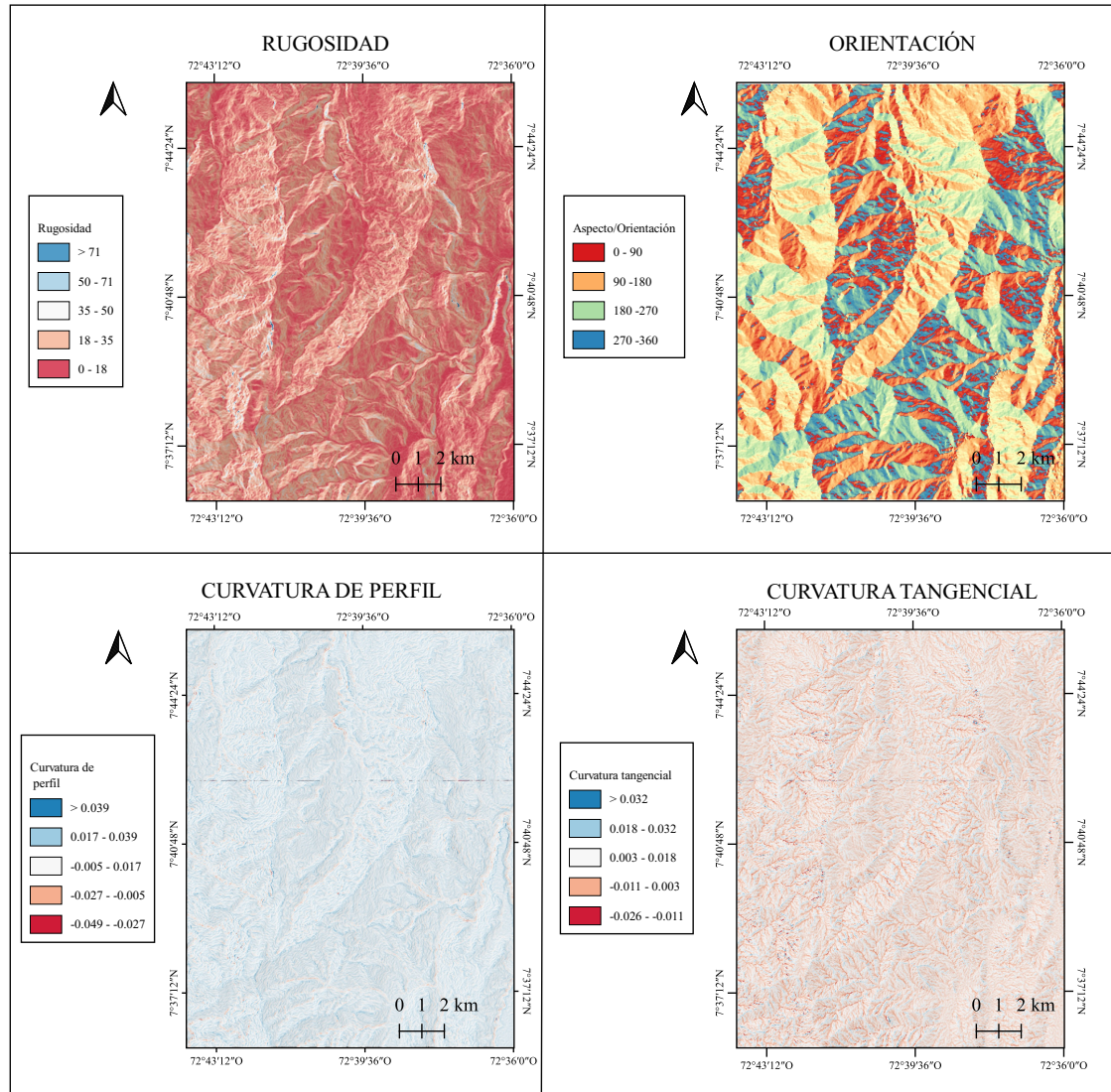


Figura 5: Variables de análisis de terreno.

De la variable de rugosidad se puede afirmar que, en su mayoría, las zonas más rugosas están sobre las elevaciones del terreno y superan el valor de 35, mostrando tonos blancos y azules. Las zonas más rojas representan las zonas menos rugosas. La gráfica de orientación muestra el azimut de sobre los puntos del terreno, las variables de curvatura (perfil y tangencial) muestran las zonas de valor negativo como zonas cóncavas, mientras que los positivos como convexas.

5.1.3. Cuerpos de agua

Considerando que la comparación de etiquetas entre el set de datos EUROSAT y la zona de estudio resulta ineficiente para los cuerpos de agua, este factor se abordó desde la información georreferenciada. La información se obtuvo desde las planchas publicadas en el portal Colombia en Mapas del Instituto Geográfico Agustín Codazzi (IGAC), entidad encargada de la cartografía del país.

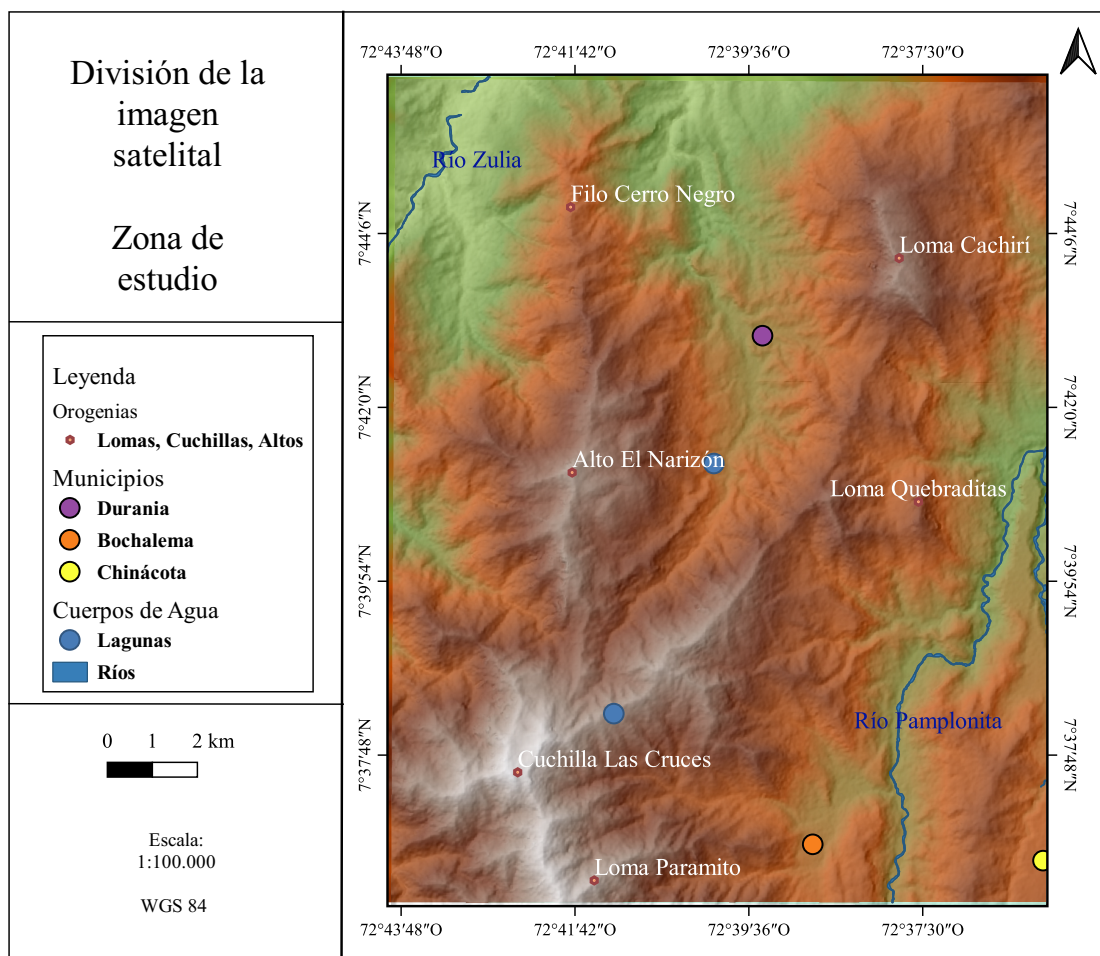


Figura 6: Ubicación de los cuerpos de agua.

Respecto a los cuerpos de agua, se observan las ubicaciones de las lagunas La Barca y El Capote, al igual que los ríos Zulia y Pamplonita.

Así, las planchas (de escala 1:25.000) de las cuales se obtuvo la información fueron: 98IB, 98ID, 98IIA, 98IIC, 98IIIB y 98IVA. De estas planchas se trabajó con la información de drenajes,

orogenias y zonas municipales. De esta forma, en la Figura 6 se observan los principales puntos geográficos de interés para esta investigación. Con la ubicación de estos cuerpo de agua, se considera que es posible rechazar o aceptar una zona para los puntos de adquisición. Por lo tanto, este factor se incluirá, junto a la cobertura vegetal y el análisis de terreno, en el algoritmo que calculará las zonas favorables para la adquisición de datos sobre la zona de estudio.

5.2. Métodos de Interpolación

Con el fin de realizar pruebas remotas de los MIE mencionados en el planteamiento del problema y así determinar el método de mejor desempeño sobre datos puntuales y dispersos, se realizaron pruebas sobre datos de acceso libre. Este planteamiento tiene como fin determinar el funcionamiento de los MIE a través de pruebas sobre una zona diferente a la de estudio. Esto garantizaría tener una ejecución que se repetiría sobre los datos de la zona de estudio y así optimizar los tiempos del modelamiento de la anomalía geofísica.

Por lo anterior, en esta etapa del proyecto se realizaron dos labores: 1) búsqueda de datos geológicos y geofísicos que tuvieran como características la presencia de un cuerpo intrusivo rodeado de rocas sedimentarias/metamórficas y 2) implementación de MIE.

Etapas 1: Búsqueda de datos

De esta forma, se realizó una búsqueda por las bases de datos del Servicio Geológico de Estados Unidos (USGS), en especial en el portal de recursos minerales desarrollado por la entidad en mención. Así, se encontró y se delimitó una zona en el estado de Utah, Estados Unidos, que contaba con un cuerpo ígneo llamativo rodeado de rocas sedimentarias clásticas y carbonatos (Horton et al., 2017) (Figura 7). Adicionalmente, desde el mismo portal se verificó y se descargó el set de datos correspondientes a gravimetría (Bankey et al., 1998) (Figura 8) y magnetometría. (Bankey et al., 2002) (Figura 9). Dentro de los datos generales que se pueden mencionar de los grupos de datos, se resalta que la adquisición de gravimetría es terrestre y cuenta con 61 puntos de adquisición. Por su parte, la adquisición de magnetometría es aérea y cuenta con 3771 puntos.

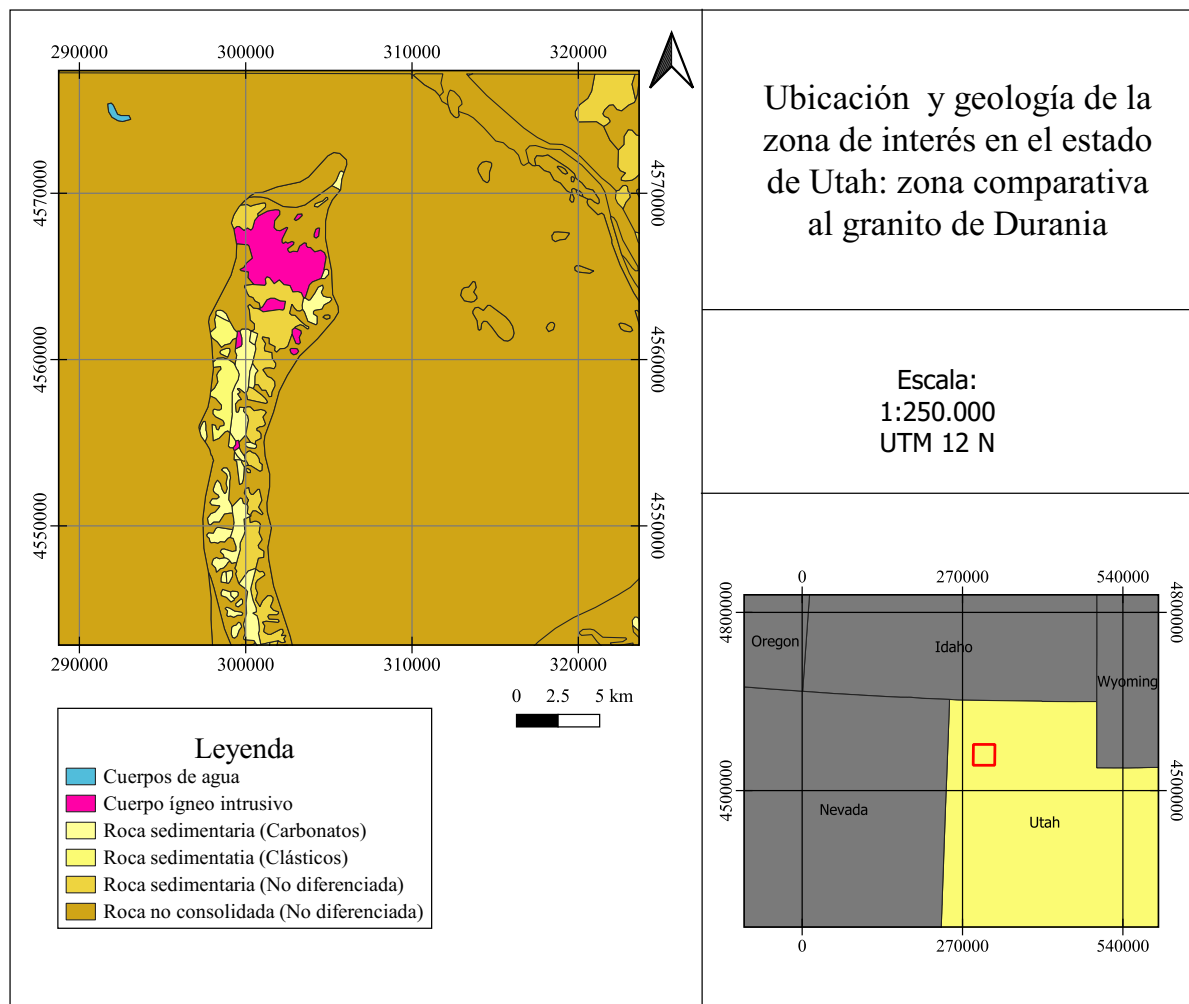


Figura 7: Ubicación y geología de la zona comparativa en Utah, Estados Unidos.

La geología disponible se descargó desde el portal de recursos minerales del Servicio Geológico de Estados Unidos se recortó de acuerdo a la zona de interés sobre el cuerpo intrusivo. Este portal tiene como recurso bibliográfico la publicación de Horton et al. (2017).

A pesar de que los datos de magnetometría cuentan con una cantidad de puntos mayor al de set de datos de gravimetría, se evidencia en la Figura 9 que el cubrimiento de la zona no es el mejor. De lo anterior se puede afirmar que esta zona cuenta con datos puntuales y dispersos sobre el área, lo que convierte a esta zona en un punto de pruebas para implementar los MIE. Por eso, el objetivo de esta metodología consiste en dejar establecido un código que se pueda replicar para el caso de Durania. Es necesario mencionar que se busca una comparación basada en el tipo de puntos y no en la magnitud de los valores geofísicos ya que estos pueden variar entre zonas.

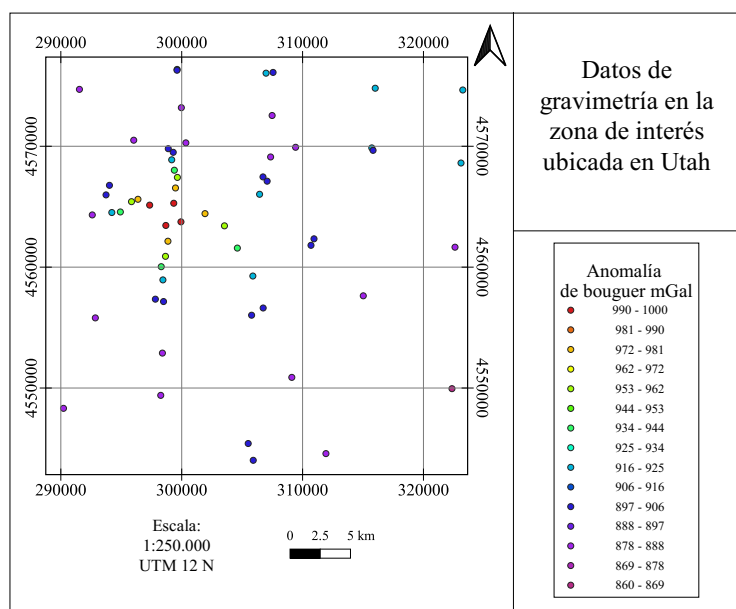


Figura 8: Gravimetría sobre la zona comparativa en Utah, Estados Unidos.

La gravimetría disponible se descargó desde el portal de datos del Servicio Geológico de Estados Unidos y se recortó de acuerdo a la zona de interés sobre el cuerpo intrusivo. Este set de datos tiene como recurso bibliográfico la publicación de Bankey et al. (1998).

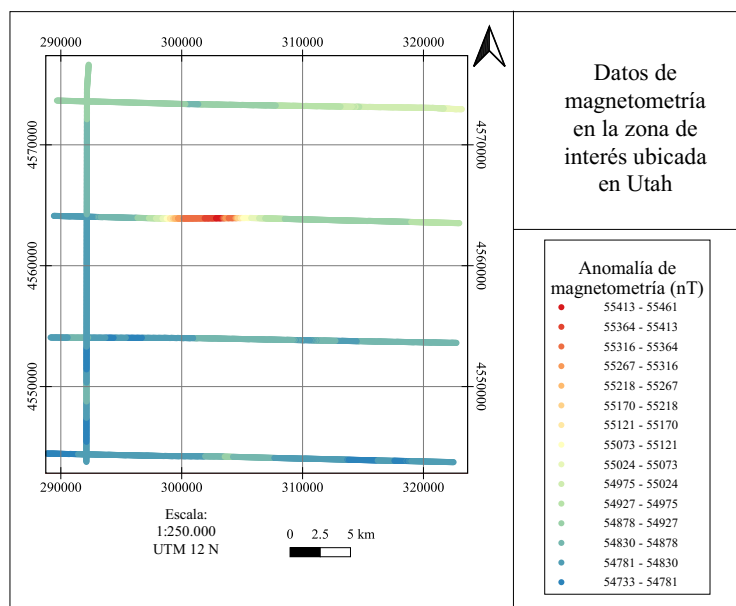


Figura 9: Magnetometría sobre la zona comparativa en Utah, Estados Unidos.

La magnetometría disponible se descargó desde el portal de recursos minerales del Servicio Geológico de Estados Unidos y se recortó de acuerdo a la zona de interés sobre el cuerpo intrusivo. Este set de datos tiene como recurso bibliográfico la publicación de Bankey et al. (2002). Los datos corresponden a la anomalía de campo total (datos de vuelo menos el campo geomagnético de referencia IGRF).

Etapla 2: implementación de MIE

En esta sección se hará énfasis a los parámetros y funcionamientos de los códigos o algoritmos implementados en la ejecución de los MIE. Los resultados se presentan en la sección correspondiente.

Kriging Ordinario (OK)

Para la implementación de este método, se utilizó la librería PrKrige y su clase OrdinaryKriging en el lenguaje de programación Python. Los parámetros utilizados fueron las coordenadas de cada una de las adquisiciones (gravimetría y magnetometría) con sus respectivos valores, un modelo de semivariograma esférico para ambos y un número de intervalos de promedio de 30. Los otros parámetros considerados en la librería fueron los que se establecen por defecto: escalamiento de la anisotropía igual a 1, ángulo de anisotropía en 0 y el tipo de inversión igual a la solución por mínimos cuadrados. Además, esta interpolación se desarrolló en una grilla definida previamente que contó con una redistribución de los límites observados en las figuras 8 y 9, y se definió un tamaño de celda de 350 píxeles. Los ajustes del modelo de semivariograma para gravimetría y magnetometría se observan en las figuras 10 y 11, respectivamente.

En el caso del semivariograma de gravimetría se observa un comportamiento de tendencia negativa cíclica sugiriendo una variación geométrica y zonal, mientras que en el caso de la magnetometría se observa una tendencia cíclica estable sugiriendo una variación zonal en los datos. Los resultados de las anomalías geofísicas generadas a partir de esta interpolación se detallarán en la sección de resultados.

Ponderación de distancia inversa (IDW)

Para la implementación de este método se utilizó el componente de interpolación IDW disponible de manera libre en el software QGIS. Los parámetros que se definieron fue un coeficiente de distancia igual a 2 y un tamaño de celda de 350 píxeles dentro de los límites definidos en las figuras 8 y 9. Al igual que el método OK, los resultados de la interpolación se presentarán en la sección correspondiente.

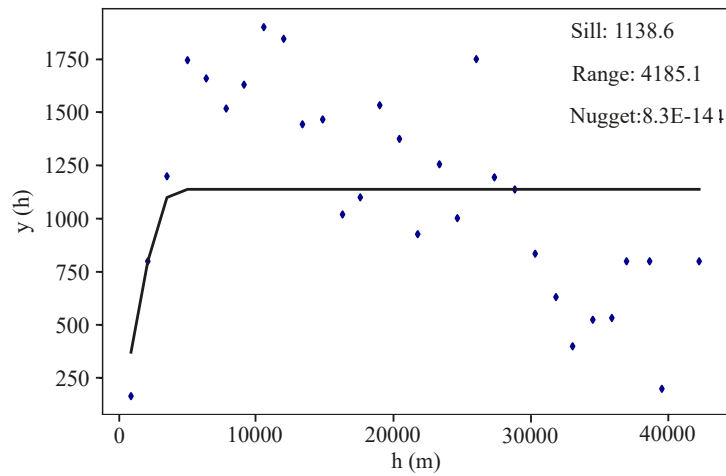


Figura 10: Semivariograma para datos de gravimetría sobre la zona de interés en Utah.

El semivariograma esférico fue el que presentó un mejor desempeño. Los resultados fueron 1138.6 para el sill, 4185.1 para el rango y $8.3E-14$ para el efecto pepita o nugget.

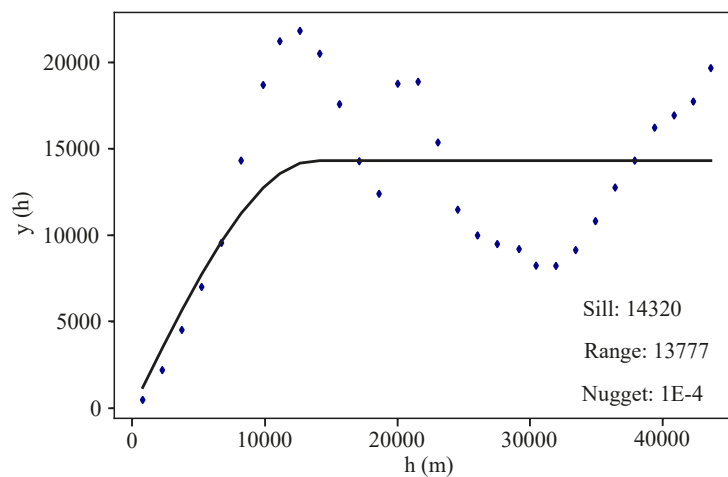


Figura 11: Semivariograma para datos de magnetometría sobre la zona de interés en Utah.

El semivariograma esférico fue el que presentó un mejor desempeño. Los resultados fueron 14320 para el sill, 13777 para el rango y $1E-4$ para el efecto pepita o nugget.

Random Forest (RF)

Para la implementación de este método, se utilizó la librería de Machine Learning Sklearn y su regresor de Random Forest para el lenguaje de programación Python. Para implementar este MIE se diseñó una grilla en función de las coordenadas UTM zona 12 N y se definió un tamaño de píxel de 350. Los parámetros asociados a la librería fueron los que se establecen por defecto: 100 árboles, profundidad máxima ninguna, número mínimo de muestras para ramificarse 2, número de muestras para ser un nodo de hoja 1, fracción de la suma total de pesos para ser un nodo de hoja 0, número de características a considerar para ramificarse automático, máximo de nodos de hoja ninguno, la impureza ponderada 0, trabajos paralelos 0 y estado aleatorio ninguno. Los resultados se presentarán en la sección correspondiente.

5.3. Modelamiento Geofísico

Durante la contemplación de esta etapa se plantearon diferentes escenarios con el fin de ejecutar las geometrías diseñadas. Se planteó una toma de datos en el sitio con los equipos adecuados (problema inverso) y, de igual forma, se planteó un escenario alternativo en el cual la adquisición en la zona no se pudiera realizar por problemas sociales (problema directo).

Una vez diseñadas las geometrías y planteados los MIE, la etapa de modelamiento inverso a través de la toma de datos en el sitio de estudio tuvo que ser desestimada debido a los eventos llevados a cabo por grupos armados durante los días 23, 24, 25 y 26 de febrero de 2022 en el territorio nacional. Sin embargo, se ejecutó un proceso remoto que diera continuidad a la investigación a través del modelamiento del problema directo: datos simulados modelados a partir de métodos numéricos. Para esto, se realizó un trabajo de tres etapas para modelar las anomalías geofísicas: 1) búsqueda de un modelo geológico, 2) reconstrucción computacional del modelo geológico y 3) predicción de los datos. Los resultados de esta simulación de datos se presentarán en la sección de resultados donde, además, se implementarán las interpolaciones espaciales de estos datos.

Etapla 1: Modelo Geológico

Con el fin de simular las anomalías geofísicas, es necesario abordar el problema directo desde un modelo geológico del subsuelo. Con este modelo geológico, no solamente se conocen las capas del subsuelo sino que también se permite asignar valores a las propiedades geofísicas de interés para la investigación. Así, a partir de la literatura es posible conocer un modelo geológico de la zona. Para esto, la plancha 98 (escala 1:100.000) de la cartografía geológica elaborada por el Servicio Geológico Colombiano (SGC) (Fúquen et al., 2011) fue la guía del modelo geológico. .

De acuerdo a Fúquen et al. (2011), se considera que en la zona, el granito de Durania se presenta buzando hacia el E y sus intrusión se da en medio de formaciones sedimentarias. En la Figura 12 se observa un corte de interés de la geología de la zona de estudio.

Etapla 2: Reconstrucción computacional del modelo

Basados en el modelo de la Figura 12, se realizó una reconstrucción del modelo desde un punto de vista computacional utilizando la librería SimPEG disponible en Python (Cockett et al., 2015).

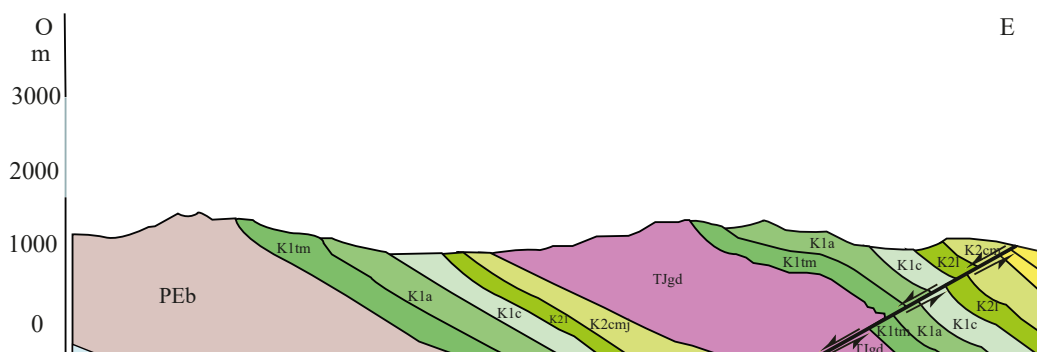


Figura 12: Corte geológico del granito de Durania.

Modificado de la plancha 98 del Servicio Geológico Colombiano. El corte de ubica en 853532 N. La coordenada más hacia el Oeste es 750233 E, mientras que la que está al Este es 764962 E. El granito de Durania (TJgd por la notación dada en el SGC) se encuentra intruyendo formaciones sedimentarias.

Para esta reconstrucción, se distribuyeron cortes en el área de estudio que permitieran, a través de la herramienta computacional mencionada, hacer una reconstrucción por zonas. Así, con base

al mapa geológico de la zona de estudio (Figura 1) y de la información de la plancha 98, se distribuyeron nueve cortes (Figura 13). Una vez definidas las ubicaciones de los cortes, se realizó la construcción del modelo en profundidad mediante bloques regulares que definieran, de la manera más cercana posible, el comportamiento del cuerpo intrusivo (Figuras 14 y 15).

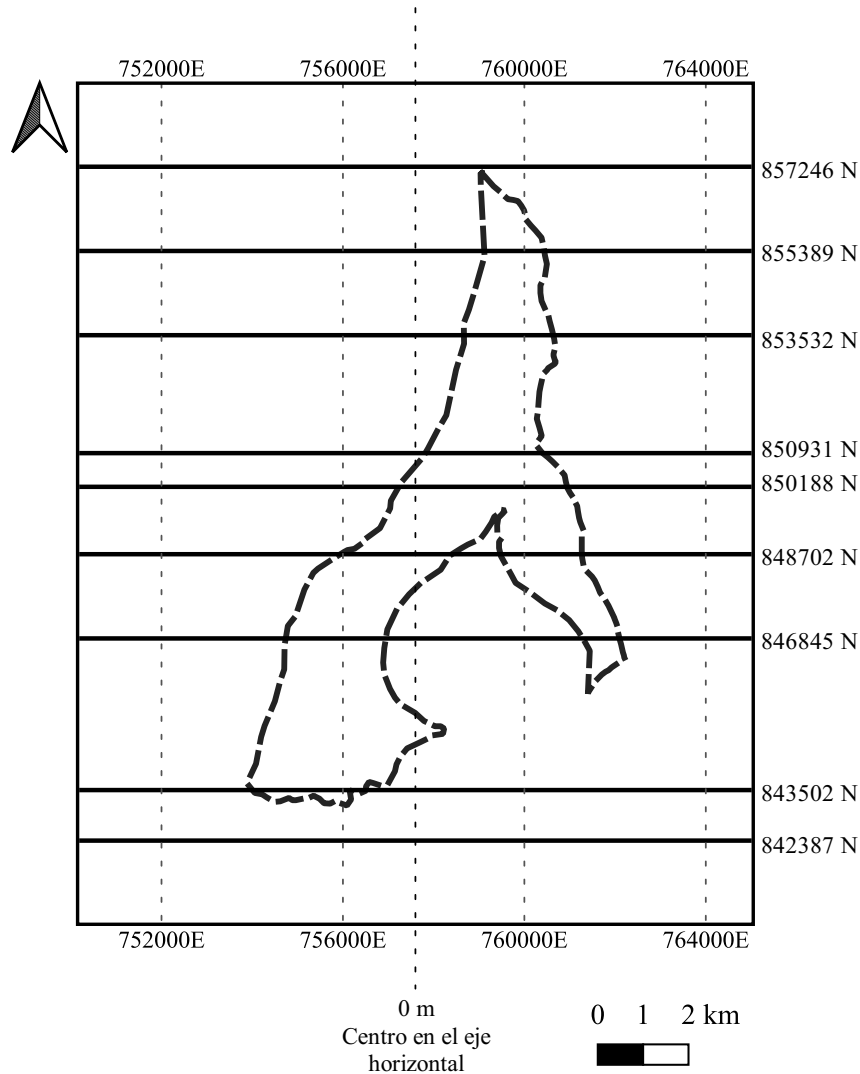


Figura 13: Vista en planta de la ubicación de los cortes geológicos.

Ubicación de los cortes mostrados en los apéndices del modelo. Además, se señala la ubicación central en el eje horizontal para representar el punto 0 m observado en los modelos computacionales presentados en el anexo.

Cabe destacar que, por motivos de simplicidad, la reconstrucción del modelo solo tuvo en cuenta dos litologías: el cuerpo intrusivo y las formaciones sedimentarias alrededor. Por esta razón,

y considerando que el cuerpo de interés es el Granito de Durania, los esfuerzos del modelamiento se centraron en reconstruir este cuerpo intrusivo de la manera más próxima a la realidad. Teniendo en cuenta lo anterior, para completar el modelo geológico se asignaron los valores correspondientes a las propiedades físicas a analizar: densidad y susceptibilidad magnética. En la Tabla 4 se presentan valores promedios de estas dos propiedades para las dos litologías de interés.

	g/cm ³	Susceptibilidad magnética (SI)
Granito de Durania	2.61	0.01
Formaciones Sedimentarias	2.42	0.0001

Tabla 4: Valores referencia para las propiedades físicas.

En la tabla se muestra los valores de referencia considerados para el modelamiento computacional. Los valores asociados al granito son los valores promedio reportados por Lowrie & Fichtner (2020). Por su parte, los valores de las formaciones sedimentarias se obtuvieron a partir de un promedio del reporte de Lowrie & Fichtner (2020) para tres rocas sedimentarias: esquistos, areniscas y calizas. Los valores promedio de estos tipos de roca se observan en la tabla del apéndice B.2.

Además, en el caso de la magnetometría, se tuvo en cuenta los valores asociados a la inclinación, declinación e intensidad del campo magnético. Esto último con el fin de incluir estos valores dentro de los parámetros de la herramienta computacional SimPEG. Así, se utilizó la calculadora geomagnética disponible desde el portal del NOAA (*National Oceanic and Atmospheric Administration*), la cual se basa en el campo de referencia geomagnético internacional (IGRF). Se obtuvieron los siguientes valores para las coordenadas del centro de la zona de estudio: inclinación = 31.0666 °, declinación = -9.2479 ° e intensidad horizontal = 27117 nT.

Etapas 3: Predicción de datos

Para esta etapa, se presentan en esta sección la información computacional relacionada al funcionamiento de la herramienta y los parámetros. Los resultados de la ejecución se presentarán en la sección correspondiente a los resultados de la adquisición. La herramienta SimPEG es un paquete de Python de código abierto diseñado para la simulación y estimación de parámetros basados en gradientes. Es una herramienta amplia que permite la inversión y la simulación de datos geofísicos. Es de esta última a la que se hará énfasis ya que será la base del modelamiento por problema directo.

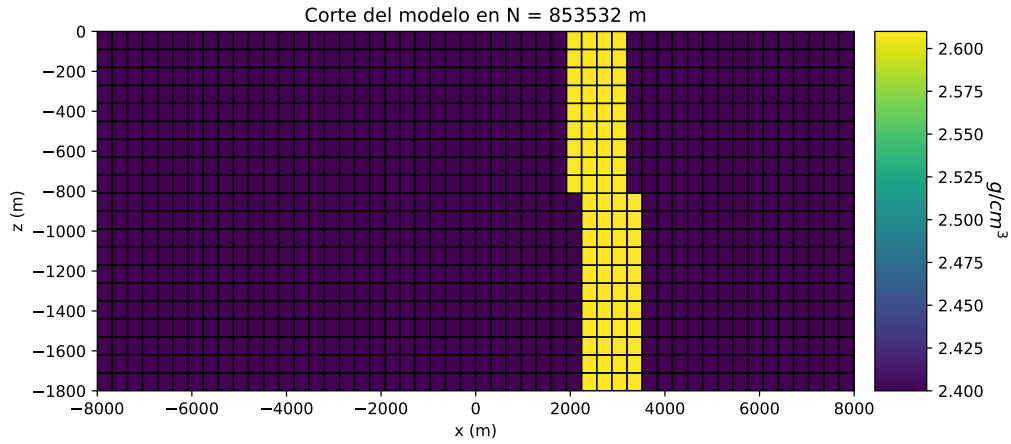


Figura 14: Modelo de densidades en 853532 N.

En la imagen se observa el modelamiento a partir de bloques del sector ubicado a 853532 N. El punto 0 m corresponde al centro del eje horizontal mostrado en la figura 13.

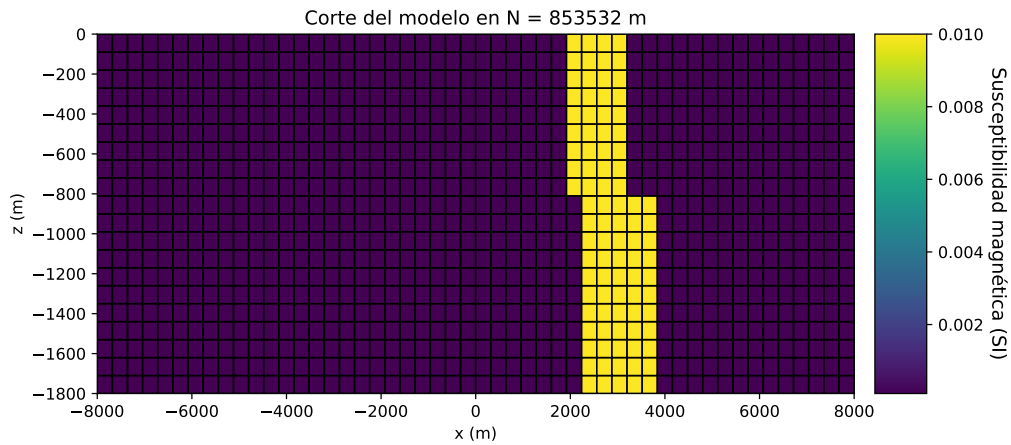


Figura 15: Modelo de susceptibilidad magnética en 853532 N.

En la imagen se observa el modelamiento a partir de bloques del sector ubicado a 853532 N. El punto 0 m corresponde al centro del eje horizontal mostrado en la figura 13.

Así, de acuerdo a Cockett et al. (2015), la simulación de datos por el modelo directo se basa en la computación de la siguiente ecuación:

$$F[m] = d_{pred} \quad (2)$$

donde F simula las mediciones específicas tomadas en la adquisición geofísica y $[m]$ corresponde al modelo físico de propiedades creado. Este modelo matemático inicial se aborda a través de un planteamiento diferencial que está dado por la siguiente ecuación:

$$C(m, u) = 0 \quad (3)$$

donde m son las propiedades físicas modeladas, u son los campos o flujos y C describe la ecuación diferencial parcial o el conjunto de ecuaciones diferenciales parciales. Así, el sistema de ecuaciones se soluciona para u y el dato a predecir se obtiene mediante una proyección de u :

$$d_{pred} = P[u] \quad (4)$$

6. Resultados

6.1. Análisis de los factores que afectan las adquisiciones de gravimetría y magnetometría

6.1.1. Coberturas vegetales

En este apartado se desarrollará la etapa 4 mencionada en la sección de metodología para coberturas vegetales donde se busca diseñar, entrenar y ejecutar la arquitectura de RNC. Para esto, antes de realizar una clasificación sobre la zona de estudio, es necesario entrenar la RNC a partir del set de datos ya clasificados (EUROSAT). Así, para entrenar la RNC se realizó una división de las imágenes de EUROSAT de la siguiente manera: 75 % de las imágenes fueron consideradas como grupo de entrenamiento y EL 25 % fue considerado como grupo de prueba. De igual forma, el grupo de imágenes de entrenamiento fue dividido en dos grupos: 75 % de entrenamiento y el restante de validación. De esta forma, el grupo de entrenamiento contó con 12.375 imágenes, el de validación con 4.125 y el de prueba con 5.500.

El diseño de RNC se evidencia en la Figura 16. El modelo es de tipo secuencial, lo que implica que recibe un solo *input* y termina en un solo resultado *output*.¹ La función de activación empleada en las capas ocultas es de tipo ReLu, mientras que la de salida es la softmax. Las capas *Conv2D*

¹Los términos relacionados al Aprendizaje Profundo se describen en la sección de Glosario.

convolucionan el *input* recibido de la etapa inmediatamente anterior y produce un tensor de salida que es reducido espacialmente en la etapa de *MaxPooling* mediante la selección de una ventana de análisis y su valor máximo (*poolsize*). Esta ventana se desplaza por pasos a lo largo de las dimensiones y dirige el resultado a una nueva etapa convolucional. Antes de llegar a la etapa final, el proceso pasa por un aplanado (*flatten*) y un abandono (*dropout*) de la mitad de datos. Esto último es útil para evitar *overfitting*.

La compilación del modelo contó con los siguientes parámetros: *RMSprop* como optimizador, *categorical cross – entropy* como la función de pérdida y *accuracy* como la métrica. Cabe destacar que esta configuración permite que se desarrolle la clasificación multiclase de forma óptima. Los resultados del entrenamiento se obtuvieron luego de ejecutar 20 épocas donde se pudo controlar el *overfitting*. Así, los resultados luego del entrenamiento se establecieron en una exactitud de prueba (*test – accuracy*) de 0.887 (Figura 17) y una pérdida de prueba de 0.322 (Figura 18).

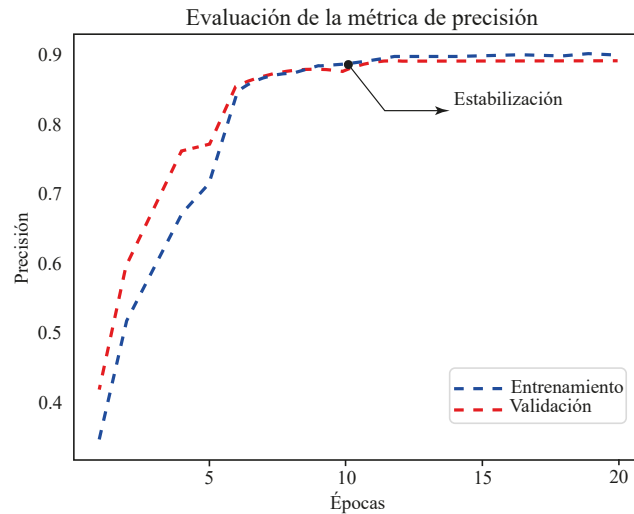


Figura 17: Evolución de la precisión para una RNC de clasificación de cobertura.

Gráfica de la precisión contra las épocas de entrenamiento. La RNC implementada contó con la función *RMSprop* como optimizador, *categorical cross – entropy* como la función de pérdida y *accuracy* como la métrica.

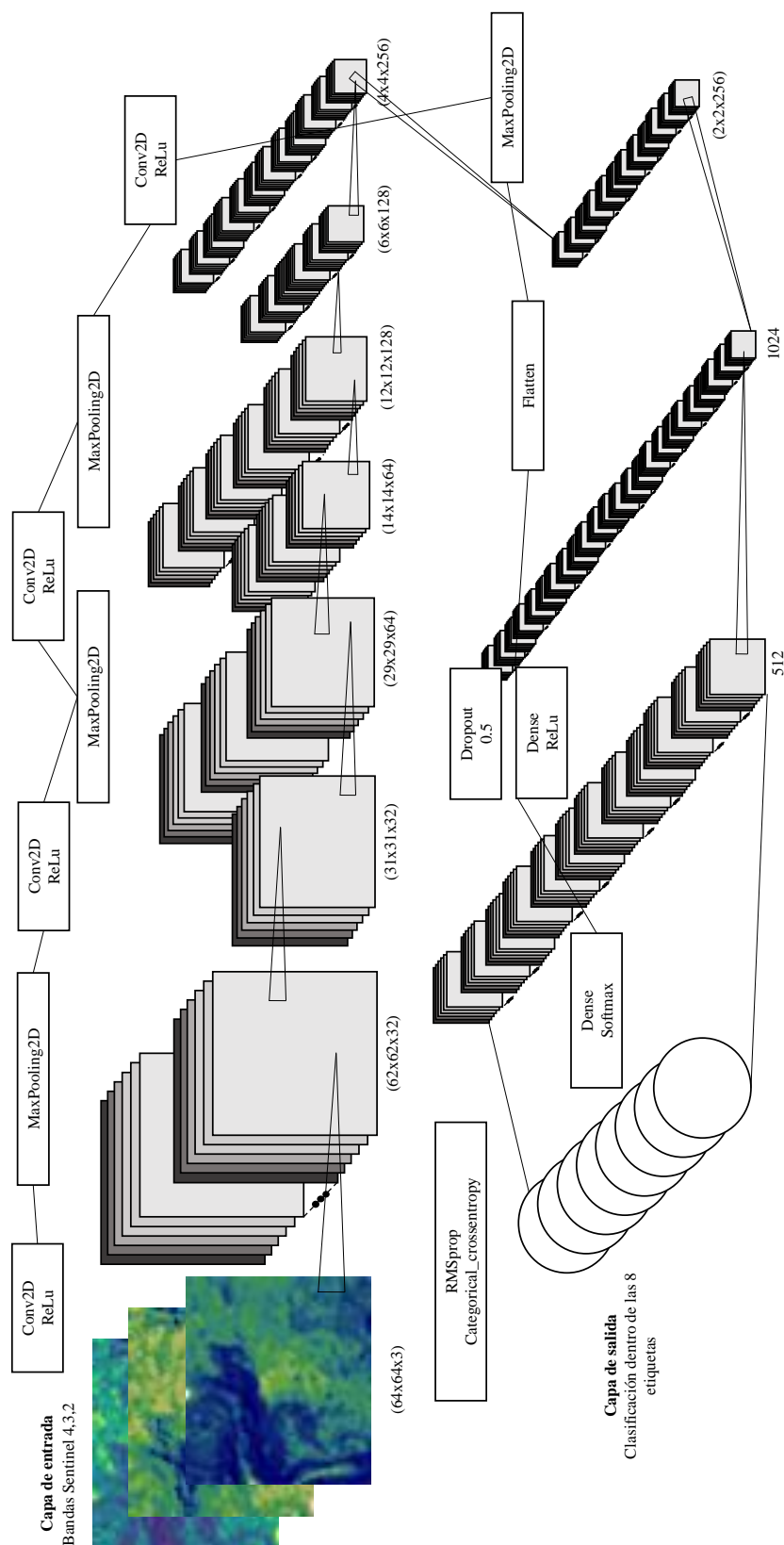


Figura 16: Modelo de Red Neuronal Convocional.
Modelo de red neuronal convocional (RNC) diseñado para el entrenamiento y posterior clasificación de la cobertura de suelo.

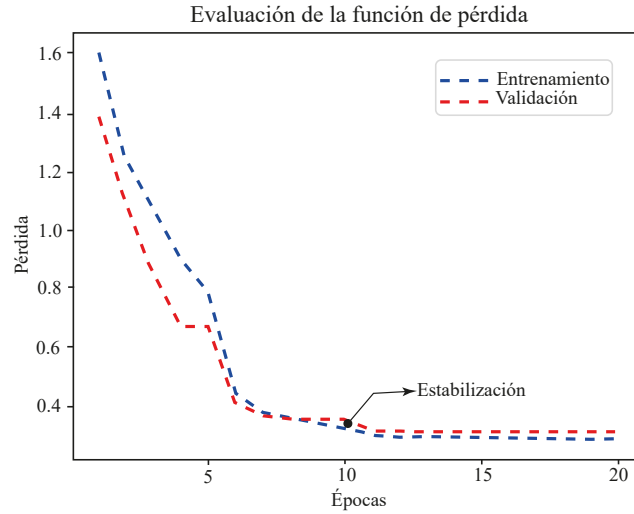


Figura 18: Evolución de la pérdida para una RNC de clasificación de cobertura.

Gráfica de la función pérdida contra las épocas de entrenamiento. La RNC implementada contó con la función *RMSprop* como optimizador, *categorical cross – entropy* como la función de pérdida y *accuracy* como la métrica.

El comportamiento de ambos grupos de datos es de tipo convergente antes de la época 10 y tiende a ser paralelo después de la época 10. Este comportamiento apunta a un buen resultado de entrenamiento con un *overfitting* reducido, lo que sugiere una garantía de buenos resultados para la etapa de clasificación. Luego de ejecutar el algoritmo de clasificación, los resultados de las ventanas de análisis fueron: bosque, pasto, vegetación herbácea, residencial/urbano, cultivos anuales y cultivos permanentes (Figura 19).

Las observaciones iniciales, luego de ejecutar la clasificación, apuntan a un resultado aceptable y que necesitaría, en principio, una verificación de campo para evidenciar que las diferencias entre el set de datos de entrenamiento EUROSAT y la zona de estudio no reflejan un efecto negativo mayor en la clasificación. La clasificación se considera aceptable ya que lo único que no está clasificando, y que se conoce presencia en el sector, son los cuerpos de agua. La etiqueta de río se considera inexistente en la clasificación, principalmente debido a la diferencia en tamaños y visibilidad entre los ríos de EUROSAT y los que se encuentran en la zona de estudio (ríos Pamplonita y Zulia). Por su parte, la etiqueta de mar-lago se considera inexistente en la clasificación ya que, de manera similar a la etiqueta de ríos, la RNC no resulta suficientemente sensible a las diferencias

entre el EUROSAT y la zona de estudio (donde se conoce la ubicación de lagunas La Barca y El Capote).

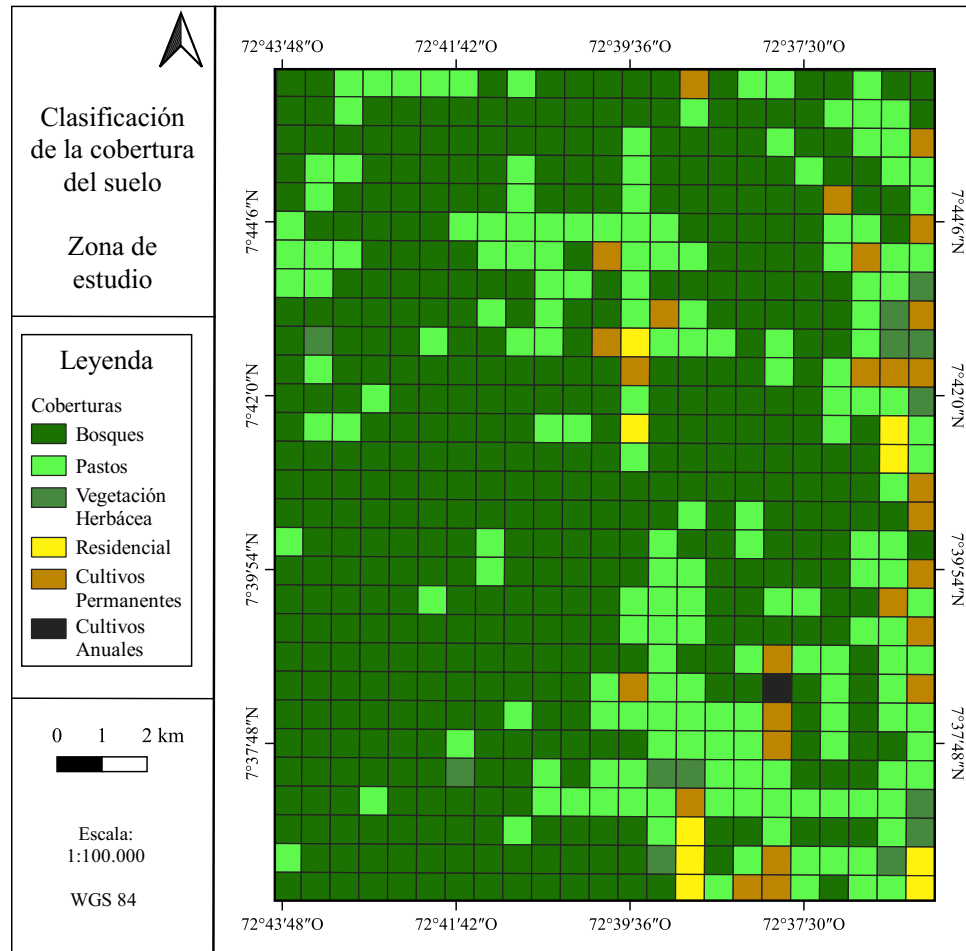


Figura 19: Clasificación de la cobertura del suelo.

Las ventanas de clasificación se agruparon en 6 de las 10 clases que hacen parte del grupo de datos de EUROSAT: Bosques, Pastos, Vegetación Herbácea, Residencial, Cultivos Permanentes y Cultivos Anuales.

6.1.2. *Análisis de terreno*

Teniendo la distribución de puntos aleatoria (Figura 4) y las variables generadas a partir de un modelo de elevación digital (Figura 5), se extrajeron los valores de datos sobre las coordenadas de los puntos aleatorios con el fin de tener un set de datos. Una vez lo anterior, el objetivo se centró en diseñar, entrenar y ejecutar una red neuronal. A diferencia de la clasificación de la cobertura

del suelo, en este factor se implementó la red neuronal artificial de tipo perceptrón multicapa (RNAPM). Esta red da un valor de salida a partir de varias capas de entrada.

Este tipo de red, contó con un costo computacional mucho menor, lo que permitió realizar múltiples pruebas antes de seleccionar la arquitectura idónea para predecir la favorabilidad de la adquisición. En la tabla 5 se observan las 10 pruebas realizadas para decidir la mejor arquitectura de RNC para el caso análisis de terreno. Para las pruebas se dividió el set de datos en un 40 % para datos de validación y un 60 % para datos de entrenamiento. Las pruebas tuvieron variaciones en la cantidad de capas ocultas (cuya función de activación siempre fue de tipo ReLu), neuronas, tasas de aprendizaje y épocas. El *input* siempre fue las 5 variables de terreno y la capa de salida siempre contó con la función de activación Sigmoide. El optimizador fue siempre *Adam*, la función de pérdida *Binary Crossentropy*, el tamaño del lote 16 y la métrica *accuracy*.

Prueba	Compilado			Épocas	Tamaño del lote	Ajuste		
	Optimizador	Taza de aprendizaje	Función de pérdida			Overfitting (época)	Entrena/Valida	
1	Adam	1.00E-03	Binary Crossentropy	250	16	NO ()	0.88/0.89	0.26/0.25
2		1.00E-03		500		SI (120)	0.89/0.89	0.22/0.28
3		1.00E-03		250		SI (70)	0.90/0.87	0.22/0.30
4		1.00E-03		500		SI (70)	0.95/0.87	0.13/0.39
5		1.00E-03		250		SI (50)	0.90/0.87	0.19/0.28
6		1.00E-03		500		SI (250)	0.97/0.86	0.11/0.36
7		1.00E-04		500		NO ()	0.87/0.85	0.28/0.28
8		1.00E-05		1000		NO ()	0.81/0.81	0.4/0.42
9		1.00E-05		3000		NO ()	0.87/0.88	0.29/0.28
10		1.00E-05		5000		NO ()	0.86/0.89	0.28/0.25

Tabla 5: Resultados de arquitecturas RNC para análisis del terreno.

Las arquitecturas de RNC asociadas a cada una de las pruebas se ven en el Apéndice A.1. De los resultados, el de mejor desempeño por su rendimiento en el ajuste (overfitting, tamaño del lote, épocas, precisión y pérdida) y por sus características de compilación (taza de aprendizaje, optimizador y función de pérdida) es la prueba 1.

Las arquitecturas de cada una de las pruebas se presentan en detalle en la Tabla del Apéndice A.1. De las pruebas realizadas se observó *overfitting* en 5 de las 10 pruebas realizadas. Estas últimas pruebas fueron descartadas y se tomó como referencia la primera prueba (que no presentó *overfitting*) para ejecutar nuevas pruebas con distintos valores de tasa de aprendizaje y épocas de entrenamiento, pero con el mismo diseño de arquitectura. Luego de comparar los resultados de precisión y pérdida, se observó que el mejor desempeño se logró con la arquitectura de la primera prueba. De esta forma, se escogió la arquitectura que se muestra en la Figura 20. De esta

arquitectura se lograron los resultados observados en las figuras 21 y 22. El comportamiento de las gráficas de precisión y pérdida muestra un comportamiento paralelo después de la época 200 y alcanza resultados buenos para un entrenamiento de 250 épocas.

Con la RNAPM entrenada, el siguiente paso consistió en predecir las zonas no muestreadas. Para esto, se creó un ráster virtual en QGIS que constara cinco bandas correspondientes a las 5 variables de análisis de terreno. Sobre este ráster se ejecutó la predicción, específicamente sobre los píxeles del ráster. En total, se clasificaron 1.791.305 píxeles extraídos del ráster virtual. La distribución de los valores resultado para estos píxeles se presentan en un histograma en el Anexo A.2.

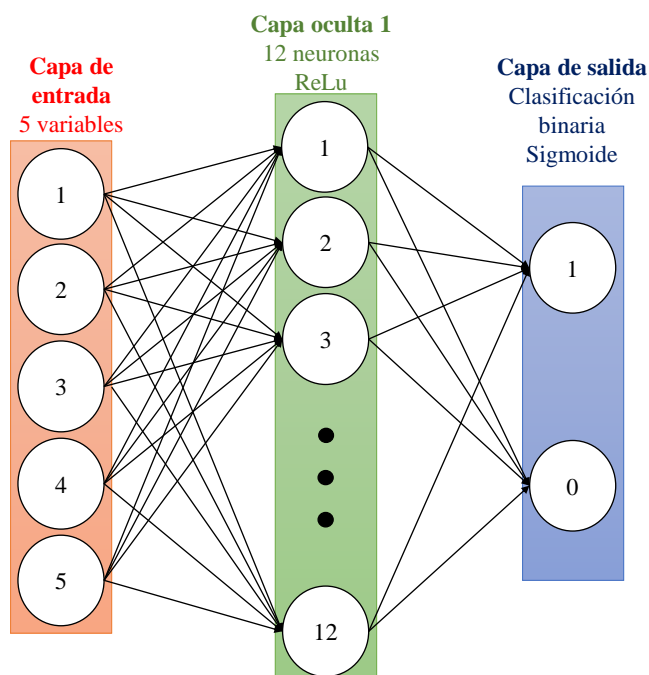


Figura 20: Modelo de Red Neuronal Artificial Perceptrón Multicapa para análisis de terreno.

Modelo de red neuronal artificial de tipo perceptrón multicapa (RNAPM) diseñado para el entrenamiento de datos de análisis de terreno.

La Figura 23 muestra el resultado final del análisis del terreno. Se muestra que las zonas de color verde (clasificación mayor a 0.1507) como áreas óptimas para la toma de datos. Las zonas de color amarillo, cuyos valores están en el rango de 0.0276 a 0.1507, se consideran zonas donde es probable la toma de datos y, finalmente, las zonas rojas, donde los valores son menores o iguales a

0.0276, son las zonas que se descartan para ser parte de la geometría de adquisición.

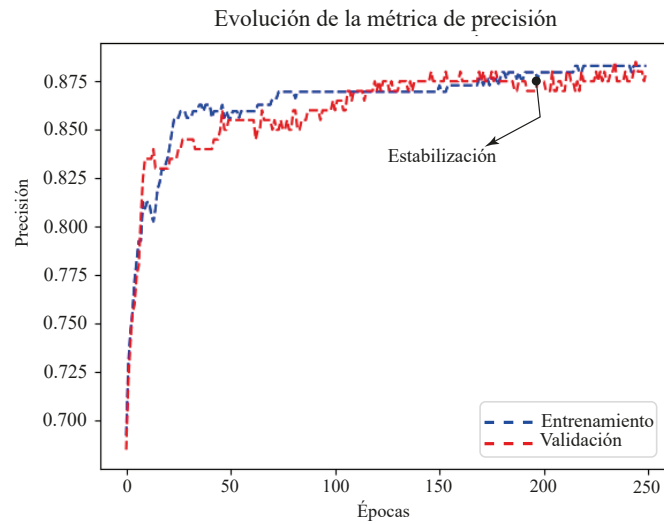


Figura 21: Exactitud de prueba para el análisis de terreno.
Gráfica de la exactitud de prueba en función de las épocas de entrenamiento.

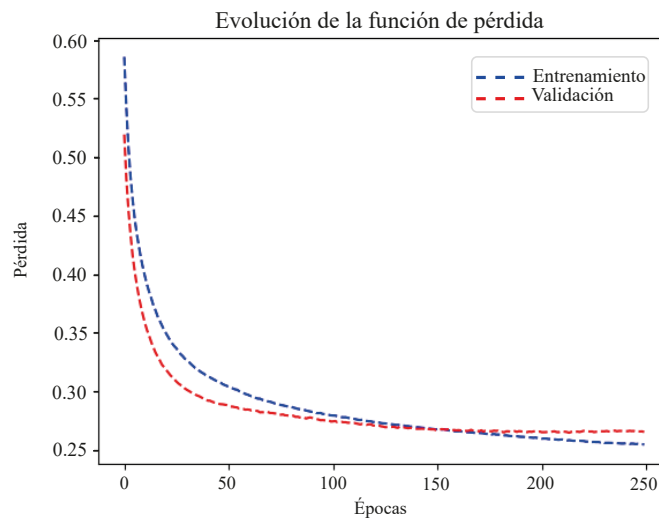


Figura 22: Pérdida de prueba para el análisis de terreno.
Gráfica de la exactitud de prueba en función de las épocas de entrenamiento.

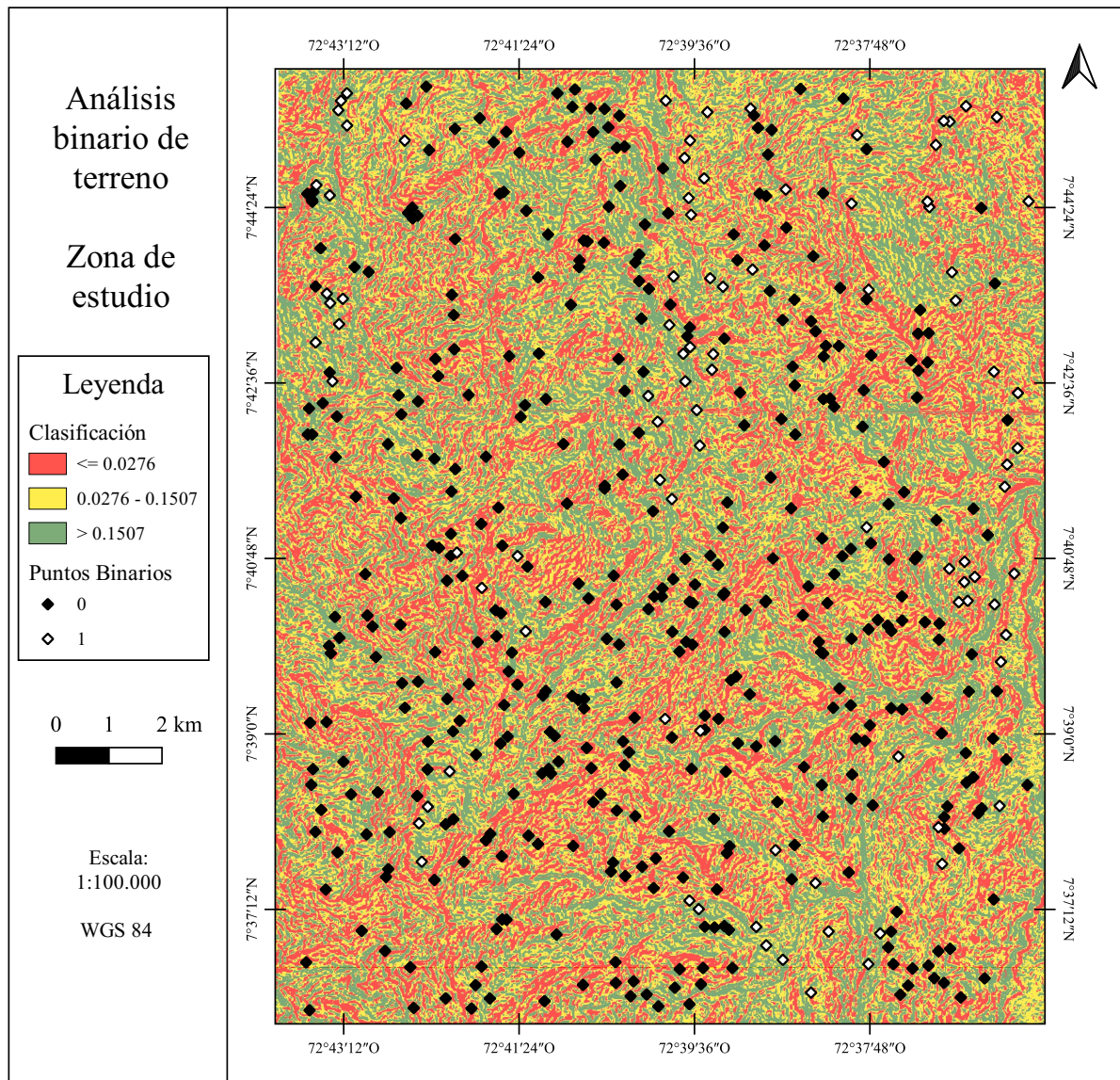


Figura 23: Mapa binario de análisis de terreno.

La clasificación de tonos verdes corresponde a los sectores verdes que serían óptimos para la toma de datos, mientras que las zonas rojas no se consideran óptimas. Por su parte, las zonas amarillas corresponden a zonas donde sería probable tomar el dato.

6.1.3. *Cuerpos de agua*

Para incluir los cuerpos de agua dentro del análisis de factores que pueden afectar la adquisición, se tomó en cuenta las ubicaciones publicadas por el IGAC (Figura 6). Con las ubicaciones de los cuerpos de agua, es posible incluir el resultado de las geometrías un factor de rechazo para las

zonas donde están presentes los ríos y lagunas.

6.1.4. *Puntos de adquisición*

Para diseñar las geometrías de adquisición se tuvo en cuenta los análisis realizados en las secciones anteriores para cada uno de los factores que afectan la toma de datos. De esta forma se trabajó de manera matemática con los archivos tipo ráster correspondientes a la cobertura del suelo, la favorabilidad del análisis del terreno y la ubicación de los cuerpos de agua. Para esto se estipularon con un valor de cero las zonas donde la cobertura de terreno no es la ideal, las zonas de laguna y ríos, y se multiplicaron por el ráster de análisis de terreno (Figura 23) y se obtuvo un nuevo ráster de favorabilidad de toma de datos donde se consideraron todos los factores. Con este cálculo se garantizó que todos los factores que pudieran ejercer un efecto negativo sobre la adquisición fueran delimitados a través de zonificaciones. Con estas zonificaciones de probabilidad, fue posible diseñar una geometría de adquisición.

Para el diseño, se trazaron 31 transectas sobre zonas donde surgiera una continuidad entre probabilidades medias y altas (amarillo y verde, respectivamente) de tal manera que se lograra una extensión de terreno aceptable para realizar las adquisiciones de los métodos de gravimetría y magnetometría. Cabe destacar que estas transectas suponen recorridos para ambos métodos ya que, con el fin de optimizar la logística de adquisición, se considera que estos recorridos son suficientes para el propósito de la adquisición. Sobre las transectas trazadas, se puntualizaron zonas de adquisición cada 100 m generando un total de 234 puntos sugeridos para las adquisiciones. En la figura 24 se observa la distribución de las transectas y un acercamiento a una zona donde hay 3 transectas con puntos cada 100 m.

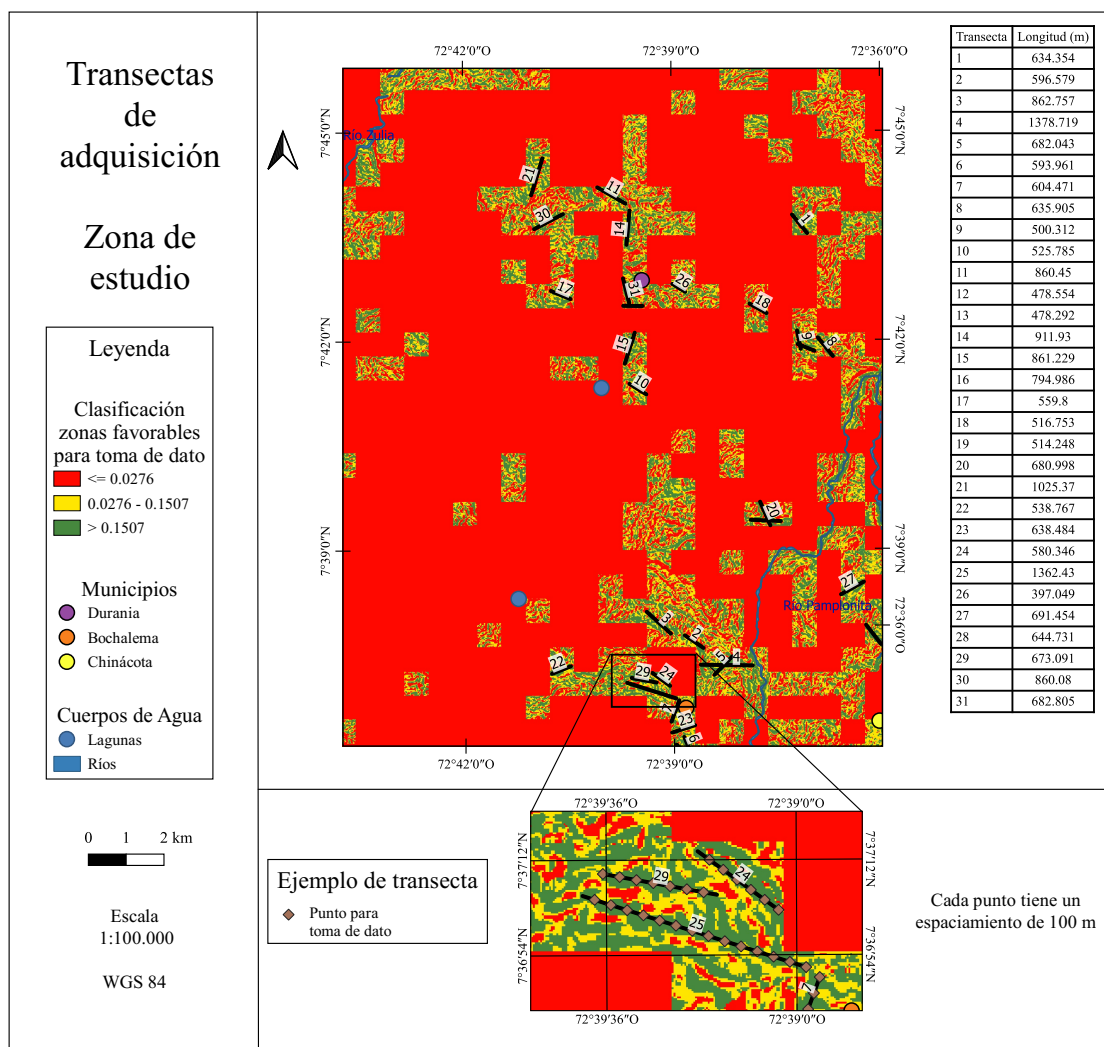


Figura 24: Zonas favorables para toma de datos.

Las zonas de color rojo (probabilidad ≤ 0.0276) corresponden a las zonas donde la combinación de factores conllevarían a una adquisición poco favorable. Las zonas de color amarillo (0.0276 - 0.1507) muestran las zonas donde la adquisición puede resultar medianamente favorable. Las zonas de color verde (probabilidad > 0.1507) muestran las zonas más favorables para la adquisición de datos. En la parte inferior se observa un acercamiento a una de las zonas y se muestra un ejemplo del resultado de dividir la transecta en puntos separados cada 100 m.

6.2. Implementación de métodos de interpolación

Las pruebas remotas de los métodos de interpolación se desarrollaron sobre los datos de norte-américa mencionados en la sección de metodología y datos. Cabe destacar que esto es un proceso realizado con el fin de verificar la eficacia o rendimiento de la interpolación espacial sobre un set de datos de acceso libre y que tengan características de tipo puntual y disperso.

Con los datos disponibles y ajustados a la zona de interés sobre el cuerpo intrusivo señalado, se desarrollaron los códigos de Python donde se implementaron librerías de libre acceso con el fin de lograr la implementación de los métodos de interpolación de Kriging ordinario (OK) y Random Forest (RF). Además, se utilizó el complemento disponible en la paquete de interpolación de QGIS para implementar la Ponderación de Distancia Inversa (IDW).

Adicionalmente, se realizó una proyección a coordenadas planas (WGS 84 - UTM 12N) con el fin de que los tres métodos de interpolación propuestos se ejecutaran en función de un set de datos planos y que la inversión se desarrollara de la mejor manera ya que las coordenadas geográficas darían un error adicional. Así, los resultados de los MIE considerados se presentan en las figuras 25 y 26.

En el caso de los resultados de Ordinary Kriging (OK), los resultados se observan en las figuras 25A. y 26A. En ambos resultados se destaca una zona de altos valores (> 980 mGal en gravimetría y > 55300 nT en magnetometría, siendo el valor alto de magnetometría un polo de unidad magnética de tipo dipolo) hacia el sector noroeste. Cabe mencionar que, en el caso de magnetometría, la anomalía es de campo total y la zona de alto valor hace parte de un dipolo. Este dipolo está como una unidad magnética que se extiende en la zona oeste con un sentido NE-SO. Este comportamiento de la unidad magnética hace que la unidad gravimétrica o anómalamente densa esté contenida dentro de la magnética. Así, se puede afirmar que la totalidad de una zona magnetizada no necesariamente tiene correlación a una zona anómalamente densa. Las métricas de error sobre este MIE se evidencian en la Tabla 6, y se puede afirmar que, de acuerdo al error absoluto medio (MAE) y el coeficiente de determinación (r^2), es un método que tiene un buen desempeño sobre los grupos de datos de la zona en Utah.

Por su parte, los resultados de la Ponderación de Distancia Inversa (IDW) se observan en las figuras 25B. y 26B. De manera similar a la interpolación OK, en esta se observan valores altos para las anomalías (> 980 mGal en gravimetría y > 55300 nT en magnetometría, siendo el valor alto de magnetometría un polo de unidad magnética de tipo dipolo) y se destaca la presencia de un dipolo o unidad magnética hacia el oeste con sentido NE-SO que contiene a la zona anómalamente densa.

Respecto a la ubicación del afloramiento, la anomalía gravimétrica se presenta desplazada hacia el SO, mientras que uno de los polos de la unidad magnética se encuentra desplazado ligeramente al S. De acuerdo a las métricas implementadas (MAE, MSE y r^2 ; Tabla 6), este método es el que tiene un mejor desempeño sobre los datos de la zona comparativa.

Finalmente los resultados de la interpolación por Random Forest se presentan en las figuras 25C. y 26C. arrojaron, al igual de los otros MIE, una zona hacia el sector noroeste que presenta valores altos de gravimetría (> 980 mGal) y de magnetometría (> 55300 nT, que corresponde a un polo de unidad magnética de tipo dipolo). Similar a lo que ocurre en los anteriores MIE, se observa un dipolo o unidad magnética hacia el O con sentido NE-SO que contiene a la zona anómalamente densa. En cuanto al afloramiento del cuerpo intrusivo, se observa que la unidad gravimétrica se encuentra desplazada hacia el O, mientras que el polo de la unidad magnética corresponde directamente al afloramiento del cuerpo intrusivo. Sin embargo, según las métricas de error (Tabla 6), este método es el de menor rendimiento y por lo tanto sus resultados se consideran rechazables.

En general, se puede afirmar que hay un comportamiento definido para la zona donde se destaca una unidad magnética de tipo dipolo que se extiende en sentido NE-SO hacia el costado oeste de la zona de estudio. Además, se observa que en los tres casos, esta unidad magnética contiene una unidad o zona anómalamente densa que se presenta, en los casos de gravimetría, ubicada ligeramente hacia el SO del afloramiento del cuerpo intrusivo. En adición, los resultados sugieren que el método IDW es, por cuantificación del error, el de mejor desempeño. Sin embargo, los contornos generados por el método OK para las unidades gravimétricas y magnetométricas se observan mejor definidos. Esto último sugiere que es posible considerar, desde un punto de vista cualitativo, el resultado del OK para interpretaciones generales de las anomalías.

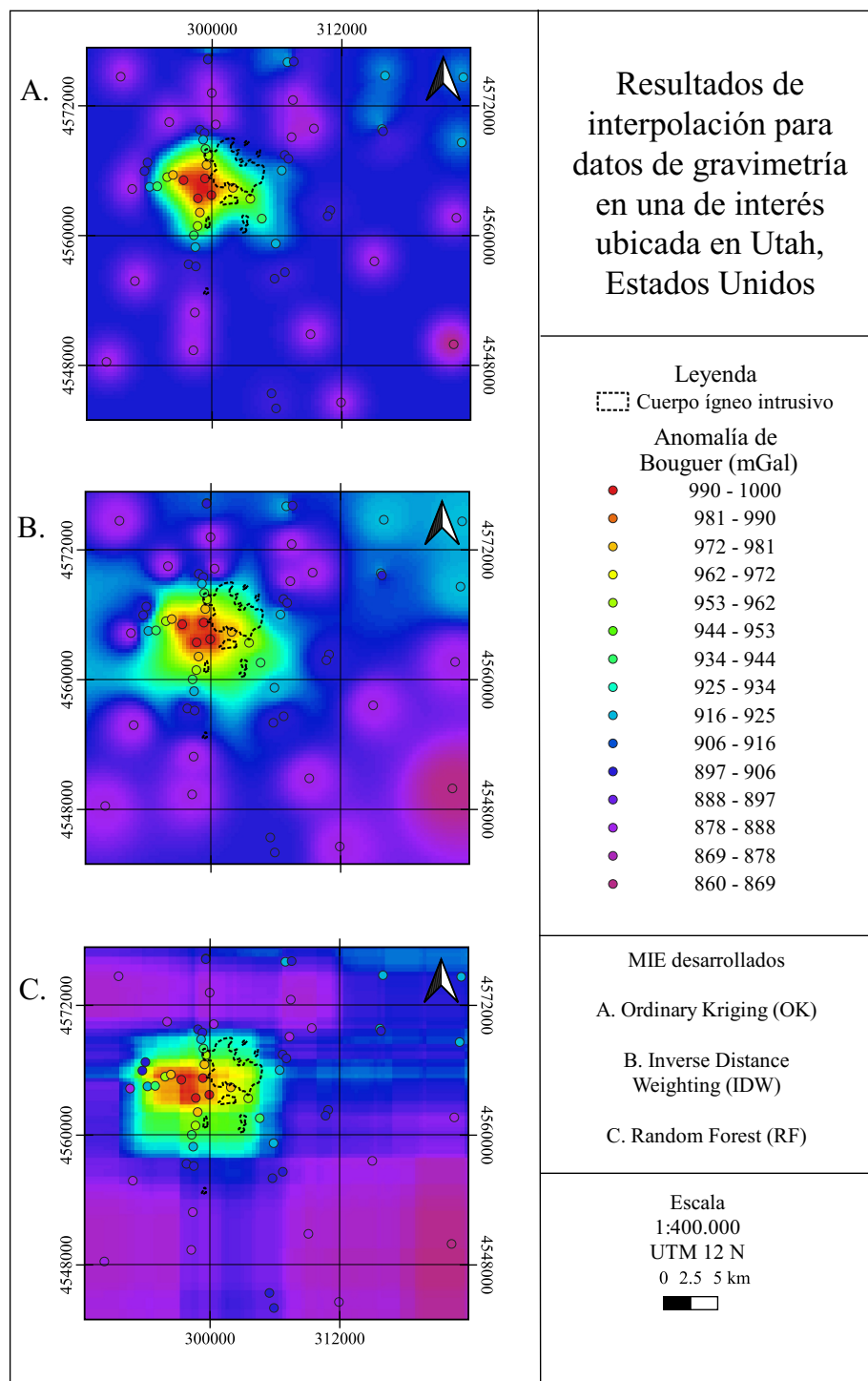


Figura 25: Métodos de interpolación en datos de gravimetría sobre la zona comparativa.

Los métodos presentados son: A.) Ordinary Kriging (OK), B.) Inverse distance weighting (IDW) y C.) Random Forest (RF). Las gráficas muestran el resultado de interpolación de datos de gravimetría. Los tres MIE muestran una zona de alta densidad (colores rojos) hacia el sector noroeste, con una geometría circular vista en plano. Este sector se interpreta como un efecto del granito que aflora en la zona. Los puntos muestran los datos originales.

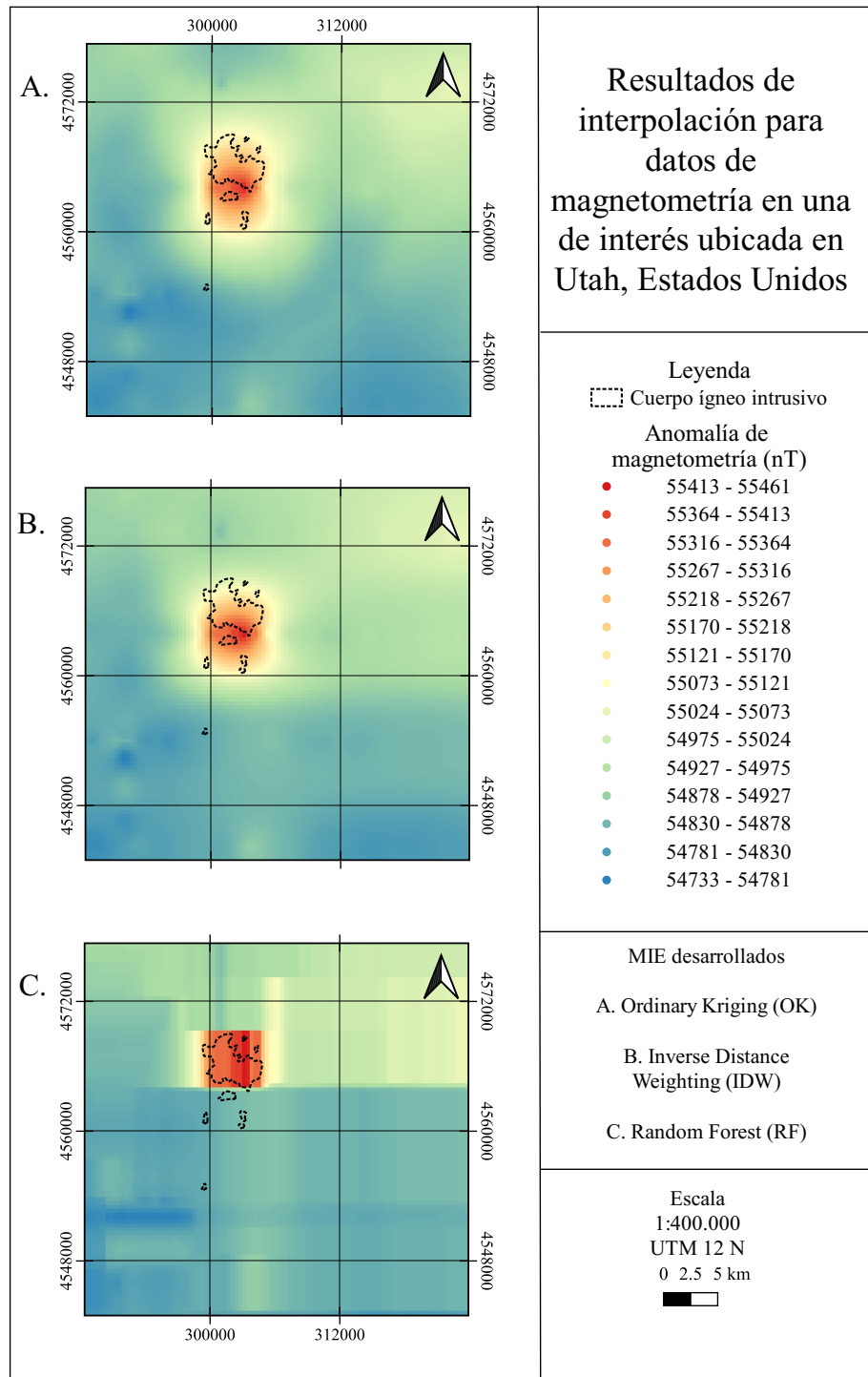


Figura 26: Métodos de interpolación en datos de magnetometría sobre la zona comparativa.

Los métodos presentados son: A.) Ordinary Kriging (OK), B.) Inverse distance weighting (IDW) y C.) Random Forest (RF). Las gráficas muestran el resultado de interpolación de datos de magnetometría. Los tres MIE muestran una zona de alta magnetización (colores rojos) hacia el sector noroeste, con una geometría circular vista en plano. Este sector se interpreta como un efecto del granito que aflora en la zona.

Método de interpolación	Métrica					
	MAE		MSE		r ²	
	Grav.	Mag.	Grav.	Mag.	Grav.	Mag.
Ordinary Kriging	3.503	4.293	24.19	57.92	0.982	0.995
Ponderación de distancia inversa	0.471	1.810	5.174	22.98	0.996	0.998
Random Forest	12.59	41.59	289.9	10084	0.787	0.299

Tabla 6: Métricas de error de los MIE.

En la tabla se muestra los resultados para tres métricas de error: Error Absoluto Medio (MAE), del Error Cuadrático Medio (MSE) y el coeficiente de determinación (r^2). Las métricas de error fueron calculadas a partir de los datos originales (Figuras 9 y 8) y datos extraídos, según las coordenadas de los originales, de los resultados de interpolación. De acuerdo a los resultados, se puede afirmar que el método de interpolación espacial IDW fue el de mejor desempeño para los dos grupos de datos geofísicos de la zona comparativa.

6.2.1. Comparativa de los MIE

Para hacer la comparativa se utilizaron tres métricas de error: Error Absoluto Medio, Error Cuadrático Medio (MSE) y coeficiente de determinación (r^2). Estas métricas de error fueron calculadas a partir de los datos de adquisición u originales (Figuras 8 y 9), y de los datos resultantes de las interpolaciones (Figuras 25 y 26). Para esto, se utilizó el complemento *Point Sampling Tool* de QGIS sobre los ráster de interpolación para extraer los datos en las mismas coordenadas que los originales y ejecutar los algoritmos de la librería *Sklearn.metrics*.

6.3. Adquisición detallada de los métodos de gravimetría y magnetometría

Una vez considerados los parámetros para el desarrollo del modelo directo para cada uno de los métodos geofísicos y sus respectivas propiedades físicas, se realizó el modelamiento completo a lo largo de la zona de estudio (en las figuras de los Apéndices B.1 y B.2 se observan el resultado del modelamiento en los cortes considerados). Un ejemplo directo del modelamiento realizado se muestra en las figuras 14 y 15, donde el corte está ubicado en la coordenada vertical 853532 N. Estos modelo corresponden a la ubicación del corte mostrado en la plancha publicada por el SGC (Figura 12).

Predicción de datos

Una vez desarrollado el modelamiento computacional, se realizó la predicción de datos a través

de la librería SimPEG (Cockett et al., 2015). Con esta librería se busca generar la anomalía de campo total para magnetometría y la anomalía de Bouguer en el caso de la gravedad. Para emplear esta librería, fue necesario adicionar una grilla de puntos (29 x 36) correspondientes a los valores de elevación en la zona de estudio. Además, se incluyó los puntos de adquisición determinados en la Figura 24 y sus respectivos valores de elevación para poder obtener los valores donde se quería obtener el dato.

Con la ejecución del código en Python, se logró obtener un modelo de datos de gravimetría y magnetometría (Figuras 27 y 28, respectivamente). Los valores de estos datos para gravimetría estuvieron en el rango de -148 a -82 mGal (con un rango específico para el Granito de Durania de -148 a -137), mientras que para magnetometría estuvieron en un rango de -23.3 a 16.9 nT (con un rango para el intrusivo que abarca todo el espectro).

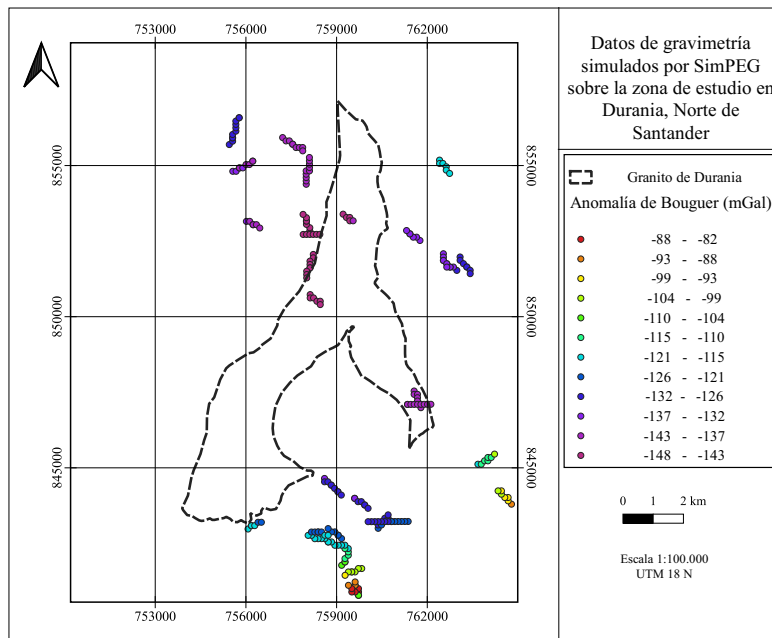


Figura 27: Datos de gravimetría simulados mediante un modelo computacional.

Los datos de gravimetría fueron modelados a partir de la geología y valores de referencia para la densidad de las litologías.

Cabe recordar que estos valores fueron simulados y focalizados respecto a la zona de zona de estudio por lo que incluyen datos de elevación del terreno, de coordenadas de adquisición y, en

el caso de magnetometría, datos puntuales para la inclinación, declinación e intensidad del campo magnético. Así, estos datos se pueden considerar como indirectos, implicando que no son datos extraídos directamente de la roca. Esto puede suponer cambios en las propiedades de las rocas, incluso las que hacen parte de un mismo grupo de rocas pueden verse afectadas. Esto es parte de las limitaciones asociadas al problema directo. De hecho, en el caso de magnetometría, se evidencia que dentro de la zona donde aflora el Granito de Durania hay variación de los datos.

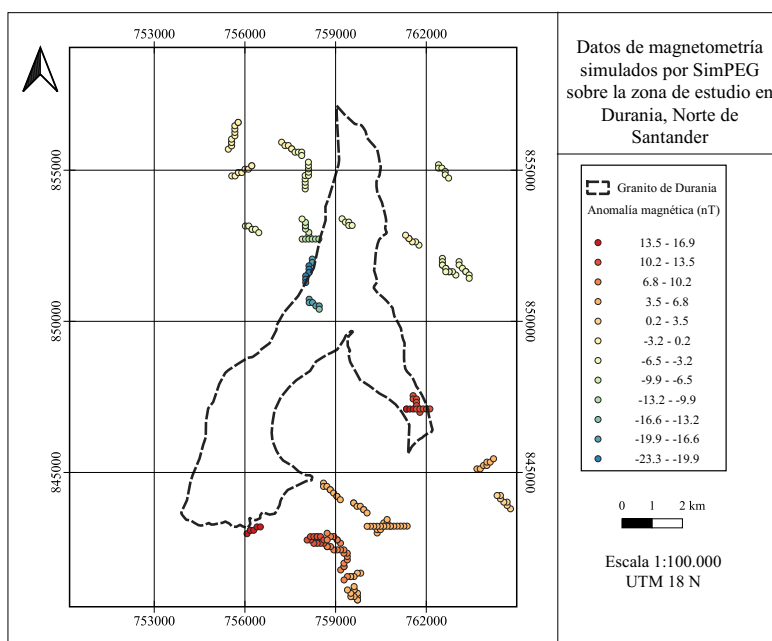


Figura 28: Datos de magnetometría simulados mediante un modelo computacional.

Los datos de magnetometría fueron modelados a partir de la geología y valores de referencia para la susceptibilidad magnética de las litologías.

Implementación de los MIE

Con los datos generados sobre las coordenadas diseñadas en la geometría, se realizó la implementación de los MIE para ambos métodos que contó con parámetros similares a los utilizados en la zona de Utah: 1) para el OK se empleó modelo de semivariograma esférico para ambos métodos y un número de intervalos de promedio de 30 (Figuras B.19 y B.20); 2) para IDW un coeficiente de distancia de 2 y un número de píxeles de 250; 3) para RF los parámetros por defecto de la librería de Python. Los resultados de los MIE se observan en las figuras 29 y 30. De estos resultados de

interpolación se puede comentar que, en ambos casos, se observa un sector de valores más bajos para cada una de las anomalías en la zona central del área de estudio.

En general, hacia la zona sur se presenta un aumento en los valores de estos métodos geofísicos. Respecto a la influencia del cuerpo intrusivo, se puede afirmar que esas zonas centrales coinciden con una sección de afloramiento del cuerpo intrusivo. Sin embargo, no se observa una influencia definida totalmente sobre el afloramiento del granito. Esto se debe a las características de los datos.

Los resultados de las anomalías se consideran una buena aproximación, teniendo en cuenta que el desarrollo de la investigación se llevó a cabo de manera remota y se considera prudente realizar, en investigaciones futuras, una adquisición real en la zona de estudio con el fin de tener datos reales de elevación, gravimetría y magnetometría. Así, se podría implementar el modelamiento inverso y generar un estudio comparativo con los resultados presentados en las anomalías finales de este estudio. Por otro lado, las métricas del cálculo de error (Tabla 7) señalan, una vez más, al método IDW como el de mejor desempeño, mientras que el método de RF resulta ser el de menor desempeño.

Método de interpolación	Métrica					
	MAE		MSE		r^2	
	Grav.	Mag.	Grav.	Mag.	Grav.	Mag.
Ordinary Kriging	0.75	0.198	3.32	0.08	0.98	0.998
Ponderación de distancia inversa	0.08	0.014	0.02	6E-4	0.999	0.999
Random Forest	4.33	3.57	32.77	25.12	0.86	0.66

Tabla 7: Métricas de error de los MIE para un modelamiento directo.

En la tabla se muestra los resultados para tres métricas de error: Error Absoluto Medio (MAE), del Error Cuadrático Medio (MSE) y el coeficiente de determinación (r^2). Las métricas de error fueron calculadas a partir de los datos modelados por el problema directo (Figuras 29 y 30) y datos extraídos, según las coordenadas de los originales, de los resultados de interpolación.

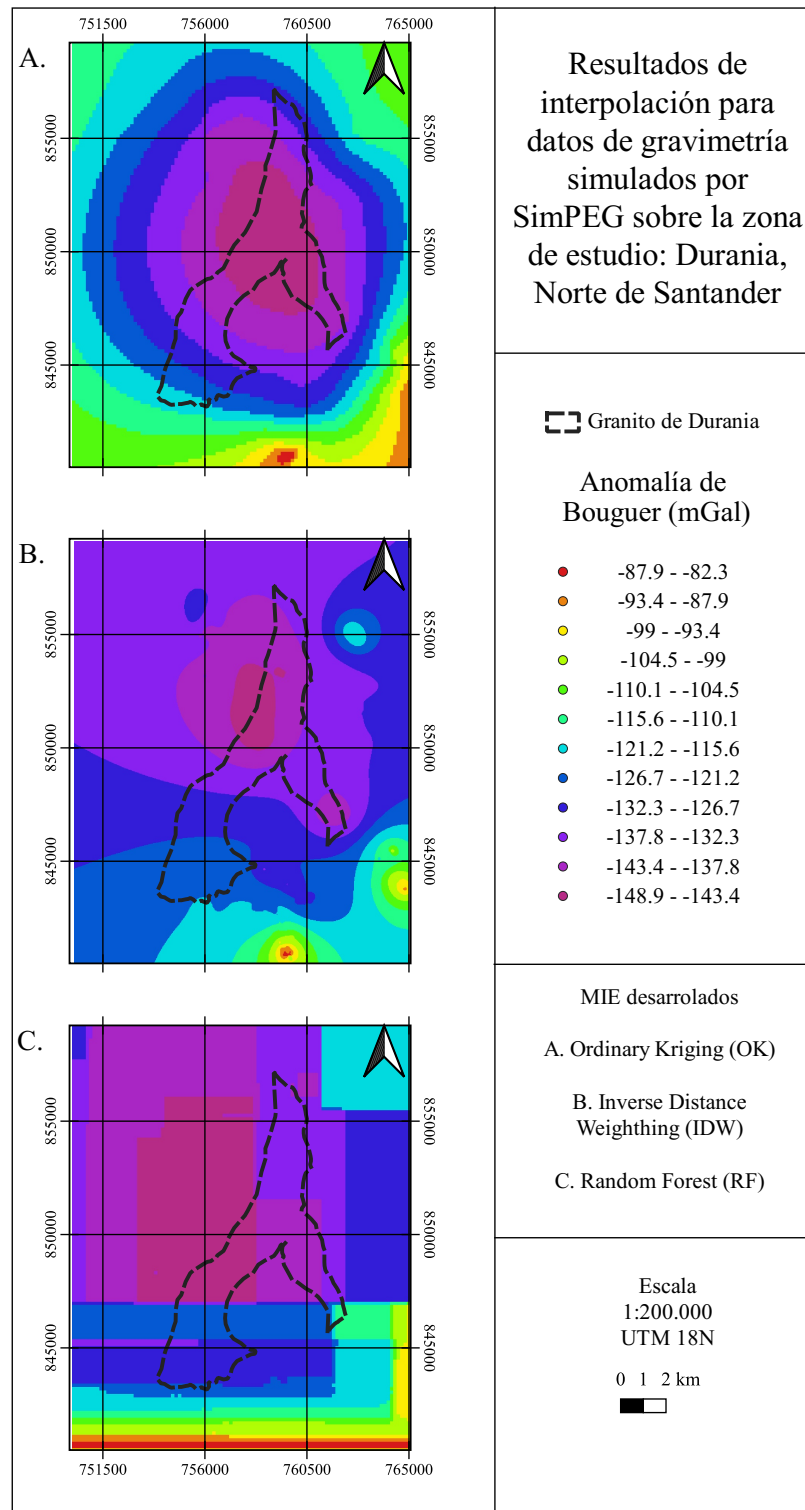


Figura 29: Resultados de interpolación sobre datos de gravimetría simulados por modelamiento computacional sobre la zona de estudio en Durania.

Interpolación de datos calculados a partir del modelo directo realizado en SimPEG. Los métodos presentados son: A.) Ordinary Kriging (OK), B.) Inverse Distance Weighthing (IDW) y C.) Random Forest (RF). Sobre la zona se presenta, con líneas punteadas, el afloramiento del granito de Durania.

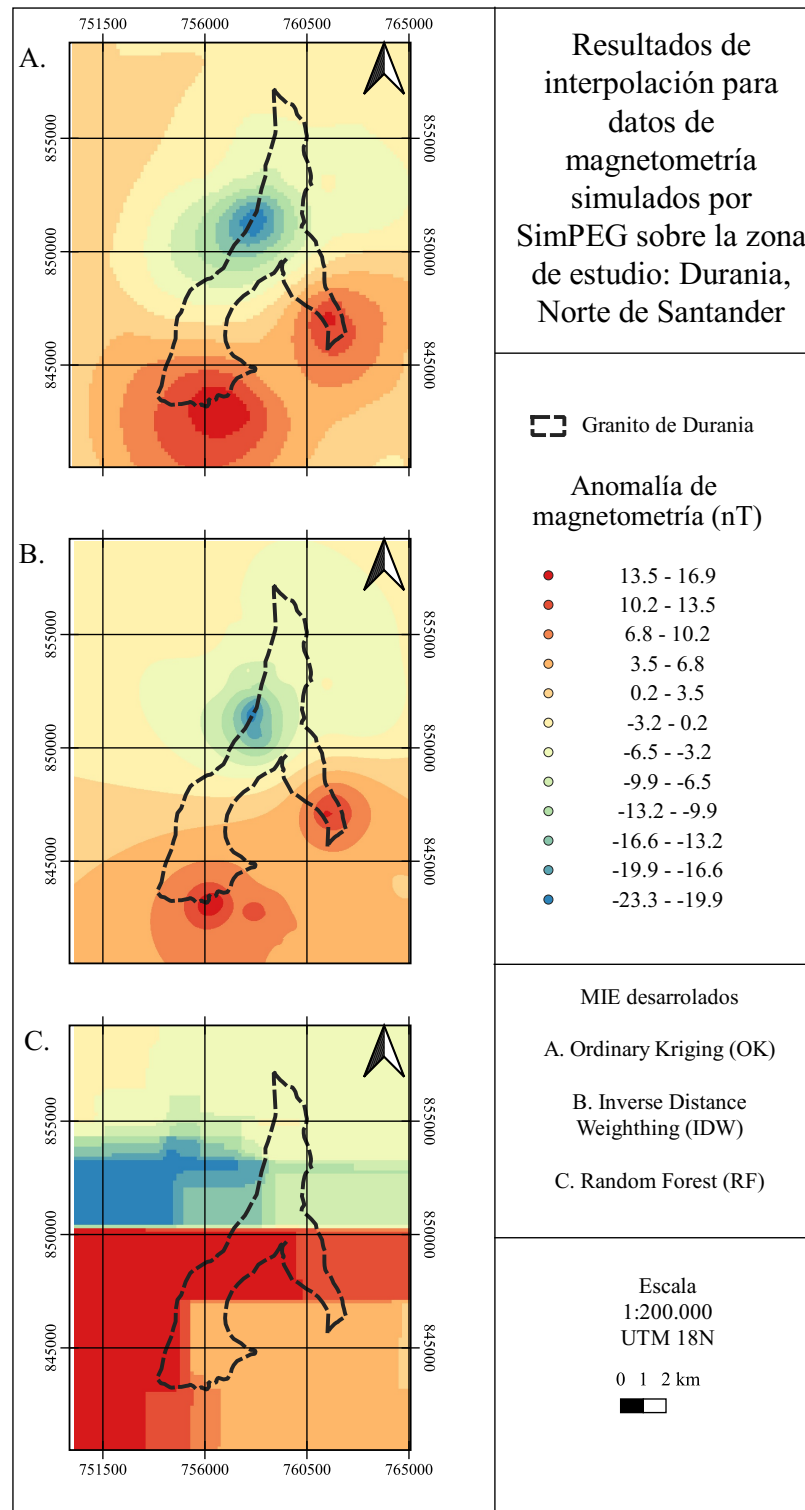


Figura 30: Resultados de interpolación sobre datos de magnetometría simulados por modelamiento computacional sobre la zona de estudio en Durania.

Interpolación de datos calculados a partir del modelo directo realizado en SimPEG. Los métodos presentados son: A.) Ordinary Kriging (OK), B.) Inverse Distance Weighthing (IDW) y C.) Random Forest (RF). Sobre la zona se presenta, con líneas punteadas, el afloramiento del granito de Durania.

7. Discusión

A partir de la información presentada en esta investigación es posible presentar: 1) un método de análisis de factores que afectan la adquisición, 2) un método de interpolación espacial que permita predecir datos en las zonas no muestreadas, 3) un diseño o geometría de adquisición y, como generalidad, 4) un flujo de trabajo reproducible para los métodos de gravimetría y magnetometría en aplicaciones que tengan como objetivo de exploración un cuerpo intrusivo similar al del granito de Durania.

7.1. Análisis de factores

A partir del análisis de imágenes satelitales fue posible delimitar las zonas donde la influencia de uno o más factores podrían generar un efecto negativo en la adquisición e datos. La zonificación se llevó a cabo mediante tres métodos computacionales: 1) Redes Neuronales Convolucionales (RNC), 2) Redes Neuronales Artificiales Perceptrón Multicapa (RNAPM) y 3) creación de un ráster virtual.

La RNC (Figura 16) fue entrenada a partir de una clasificación de uso de suelo publicada por el proyecto EUROSAT (Helber et al., 2019). Este set de datos contó con etiquetas de interés para los objetivos de la investigación ya que se tomaron en cuenta las siguientes etiquetas: vegetación herbácea, bosques, zonas urbanas, ríos, etc. En adición a las imágenes de entrenamiento, se obtuvo una imagen satelital Sentinel-2 que abarcara la zona de estudio. Esta imagen contó con las bandas del espectro RGB y fue recortada de manera que la clasificación no tuviera inconvenientes dimensionales entre las imágenes de entrenamiento y las de predicción.

A pesar que los resultados de entrenamiento fueron buenos y se evitó el *overfitting* (Figuras 18 y 17), es necesario mencionar que las diferencias entre el set de datos de entrenamiento y la zona de estudio resultan en un factor a destacar. Las diferencias en ubicación pueden suponer cambios en los tipos de cultivos, la distribución de las zonas urbanas y la presencia de los cuerpos de agua. Estas diferencias pueden generar una inexistencia de algunas etiquetas o, en ciertos casos, las etiquetas terminan siendo la de mayor cobertura en la ventana de análisis. Por ejemplo, la cla-

sificación para los cuerpos de agua (ríos y lagunas), resultó inexistente luego de la clasificación a pesar de que se conoce la presencia de los ríos Pamplonita y Zulia, pero sobre la ubicación de estos cuerpos de agua también se conoce vegetaciones espesas que, al ocupar gran espacio de la ventana de análisis, generan una clasificación preferente la para las etiquetas de bosques, principalmente.

Para el análisis de terreno, la RNAPM diseñada (Figura 20) fue entrenada con una clasificación binaria aleatoria en la zona. Esta clasificación fue realizada sobre el modelo de pendientes generado a partir del modelo de elevación digital. Se consideró la pendiente como el factor principal dentro del análisis de terreno teniendo en cuenta que puede ser el de mayor impacto en el desplazamiento en campo y, a su vez, un factor negativo para la adquisición y procesamiento de datos de gravimetría. Así, se clasificó en función del valor de pendiente asignando ceros a valores de pendiente mayor a 10° . En adición a la pendiente, se incluyeron 4 valores más asociados a otros factores del terreno que permitieran a la RNAPM entrenarse de manera imparcial ante la clasificación binaria asignada por el investigador. Así, se lograron resultados buenos y sin *over fitting* como se presenta en las Figuras 21 y 22. Teniendo en cuenta que el rango principal fue la pendiente, cabe destacar que el valor de corte fue modificado de 22° a 10° para considerarse como zona válida. Esto se hizo según criterio del investigador y se considera como un valor modificable de acuerdo a los criterios que se deseen tener en cuenta.

El ráster virtual contó con los resultados de los métodos computacionales anteriores y además, contó con la localización de los cuerpos de agua presentes en la zona de estudio (debido a la inexistencia en la clasificación de la RNC). Con los tres factores considerados, la matemática condicional aplicada dio como producto una zonificación a partir de la probabilidad (Figura 24). Sobre esta zonificación, se trazaron transectas de manera manual que permitieran la mayor conexión de zonas favorables. Además, en esta etapa de trazado de líneas de adquisición, se tuvo en cuenta los desplazamientos y puntos a adquirir ya que, por temas de logística, es importante optimizar los tiempos empleados en cualquier adquisición geofísica. Es necesario destacar que, la longitud de las transectas estuvo en función de las zonas de favorabilidad conectadas de la manera más recta posible. Esto hace que la cantidad de puntos de adquisición varíe de una transecta a otra.

A pesar de que se consideran suficientes las transectas propuestas, es posible sugerir o trazar más partiendo del mismo concepto. Además, al ser este el resultado final del análisis de factores, se considera que es un producto remoto y temporal ya que, debido a eventos naturales o antropogénicos, se pueden dar alteraciones en la cobertura del suelo. Esto puede implicar que durante la toma de datos se presenten inconsistencias por el paso del tiempo entre la fecha de la imagen satelital y la fecha de adquisición.

7.2. Determinación del MIE de mejor desempeño

Partiendo de que el resultado de la Figura 24 es considerado remoto y temporal, se consideró importante realizar un trabajo remoto para la interpolación de datos. De esta forma es posible garantizar que las transectas de adquisición pueden complementarse de manera óptima y la planeación geofísica resulte exitosa. El objetivo será generar una imagen geofísica de calidad a partir de la interpolación de los datos.

Considerando que para esta etapa no se cuenta con datos de libre acceso sobre la zona, resultó necesario hacer un proceso reproducible a partir de un sector que si contara con datos geofísicos de acceso libre y que, además, contara con litologías similares a las la zona de estudio. Con esto en consideración, se implementaron MIE sobre una zona en Utah, Estados Unidos. Cabe destacar que los MIE implementados se limitaron a dos métodos clásicos (Ordinary Kriging y el método IDW) y uno moderno (Random Forest). Sin embargo, ninguno de estos métodos está planteado específicamente para el caso de datos puntuales y dispersos (resultado del análisis de factores), y aunque la literatura presenta algunos métodos enfocados a este tipo de datos (Chen et al., 2019), estos métodos son desarrollados y aplicados sobre geofísica 1D, por lo que no serían métodos reproducibles sobre esta investigación. Esto, en principio, supone una limitante al resultado de las anomalías geofísicas. Con los MIE aplicados sobre la zona de Utah, se calcularon métricas de error (MAE, MSE y r^2) que tuvieran como referencia los datos originales y los datos resultantes de la interpolación. Estas métricas fueron calculadas de manera computacional y los resultados señalaron, tanto en gravimetría como magnetometría, al método IDW como el de mejor desempeño.

A pesar de que en esta etapa se define un MIE de buen desempeño, se considera importante destacar que: 1) es necesario replicar los MIE escogidos y los cálculos de métrica para cualquier caso, 2) se puede considerar variar los parámetros de los MIE con el fin de lograr mejores desempeños y 3) se considera necesario desarrollar e implementar, en investigaciones futuras, una aplicación directa a datos puntuales y dispersos en la geofísica 2D. Además, esta observación parte desde la cuantificación del error, pero a nivel interpretativo los contornos generados para el Ordinary Kriging se observan más definidos y más acordes al cuerpo intrusivo, en el caso de gravimetría, y al dipolo generado por la anomalía de cuerpo total en el caso de magnetometría. Esto último puede asociarse a los resultados presentados por la literatura donde señalan, en casos generalizados sin énfasis a a datos puntuales y dispersos, que el método de mejor desempeño es el Ordinary Kriging (de Amorim Borges et al., 2016).

7.3. Diseño de la geometría de adquisición

Con los dos resultados previos, el siguiente paso consistía en dirigirse a la zona de adquisición para realizar la toma de datos correspondiente. Sin embargo, los problemas sociales del país fueron causal de desestimación. Así, y con el fin de dar un resultado final al diseño de la geometría, se realizó un trabajo remoto basado en el modelamiento directo de las anomalías geofísicas mediante métodos numéricos.

De lo anterior, el proceso partió de la construcción computacional del modelo geológico de la zona donde, además de señalar la litología relevante, se asignaron valores aproximados para las características físicas en cuestión: densidad y susceptibilidad magnética. El modelamiento se llevó a cabo en la librería SimPEG (Cockett et al., 2015), donde, además del modelamiento, se obtienen datos puntuales para las coordenadas definidas. Cabe destacar que este modelamiento se considera una aproximación a la realidad por las siguientes razones: 1) el modelamiento geológico al interior del subsuelo puede variar en forma, distribución y tamaño; 2) los parámetros asociados a la elevación del terreno en los puntos de adquisición pueden variar con la magnitud real; 3) los valores de magnetometría se pueden ver afectados por la variación del campo magnético en el

tiempo en el momento de la adquisición; y 4) los valores de las propiedades físicas pueden variar o presentar rangos más amplios para cada una de las litologías dependiendo del investigador.

De esta forma, los resultados del modelamiento (Figuras 27 y 28) se consideran valores aproximados a la realidad. Una vez obtenidos estos datos modelados para las coordenadas planteadas, se ejecutaron los MIE considerados en la discusión anterior. Los resultados (Figuras 29 y 30) y las métricas de error (Tabla 7) señalan al método IDW como el de mejor desempeño. Esta afirmación parte de la cuantificación del error y sobre el resultado del IDW se puede afirmar que se observa un efecto subdimensionado de la anomalía con respecto al afloramiento del granito. Sin embargo, los rangos que se observan sobre la zona central del cuerpo intrusivo señalan valores medios bajos (-126 a -143 mGal) para la anomalía en el caso de gravimetría y, además, se observa una unidad gravimétrica llamativa por ser la de valores más bajos. Por su parte, la anomalía de campo total en el caso de magnetometría muestra un comportamiento de tipo dipolo que sugiere una unidad magnética amplia sobre el afloramiento del granito. Así, se puede afirmar que, en general, la unidad gravimétrica está contenida dentro de la magnética ya que se estima que la totalidad de la zona magnetizada no está relacionada restrictivamente a una zona anómalamente densa.

Al señalar al IDW como el de mejor desempeño, se tiene en cuenta el resultado de la cuantificación del error y de las observaciones sobre la anomalía. Sin embargo, es posible considerar el OK como otro resultado de interés ya que la anomalía tiene contornos más definidos sobre el área de interés. Así, sobre este resultado, la interpretación apuntaría también a una unidad gravimétrica de valores bajos contenida dentro de una unidad magnética de tipo dipolo. En este caso, las unidades son más grandes que las que se generan en IDW por lo que abarcarían, especialmente en el caso de la zona anómalamente densa, una gran extensión del afloramiento del granito de Durania.

El resultado de los MIE se ve claramente afectado por la distribución de los puntos de adquisición ya que, por ejemplo, los puntos distribuidos hacia el sector Sur-Oeste del granito son más escasos que los de la zona Norte y la zona Sur-Este. Esto, en parte, se debe a las restricciones de adquisición de esta zona (Figura 24). De igual forma, es necesario destacar que se desestimaron algunas zonas de favorabilidad por un tema de distribución de tiempo y logística para un escenario

real de toma de datos. Por ende, una recomendación a investigaciones futuras es considerar aumentar y/o extender las transectas en función del tiempo disponible para la campaña de adquisición (p. ej. se estima que para la geometría y los métodos planteados en esta investigación, la logística requeriría de aproximadamente 4 días). Así, no solo se lograría tener más datos en el momento de la interpolación, sino que los valores reales de elevación de terreno y de propiedades geofísicas, permitirían realizar un modelamiento inverso más acorde y preciso con la realidad.

7.4. Flujo de trabajo

En la Figura 31 se muestra el flujo de trabajo implementado y propuesto a futuras investigaciones.

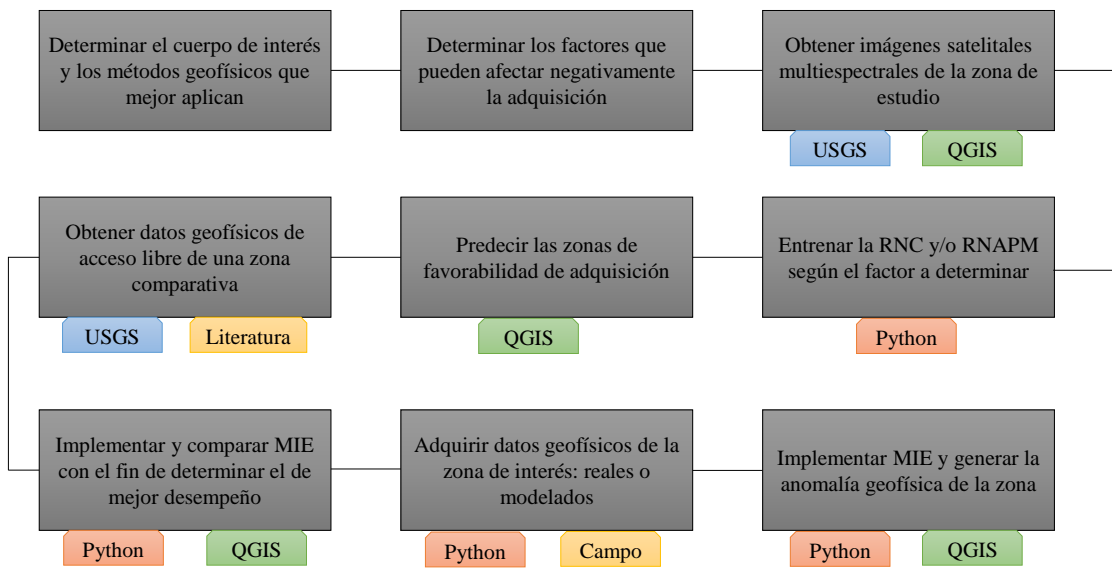


Figura 31: Flujo de trabajo

Flujo de trabajo implementado y propuesto a futuras investigaciones donde se detalla las herramientas utilizadas en cada caso. La notación USGS hace referencia a los datos de los portales del Servicio Geológico de Estados Unidos.

Se propone: 1) que la planeación geofísica parta del cuerpo de interés y los métodos geofísicos que mejor lo pueden explorar; 2) determinar los factores asociados a los métodos geofísicos; 3) descargar imágenes satelitales multiespectrales de los portales del USGS (p. ej. Sentinel-2 y ALOS

PALSAR) con el fin de visualizar y procesar la información en QGIS; 4) entrenar las redes de Aprendizaje profundo a implementar de acuerdo al factor negativo que se desea evitar en el diseño de la geometría; 5) predecir las zonas de favorabilidad en QGIS a través de las redes del paso anterior; 6) obtener datos geofísicos de acceso libre sobre una zona comparativa (se busca que la zona comparta características geológicas con la zona de estudio que no cuenta con datos libres) ; 7) determinar el MIE de mejor desempeño sobre los datos de la zona comparativa; 8) adquirir datos geofísicos sobre la zona de interés (reales u obtenidos por modelamiento directo); y 9) implementar el MIE de mejor desempeño con el fin de obtener imágenes geofísicas de calidad.

8. Conclusiones

El análisis de factores que afectan la adquisición geofísica de los métodos de gravimetría y magnetometría se considera una etapa fundamental en la planeación geofísica. Así, respecto al primer objetivo específico, se puede afirmar que mediante la aplicación de Aprendizaje Profundo aplicado sobre imágenes satelitales, es posible determinar zonas de favorabilidad para el diseño de geometrías de adquisición. Sin embargo, es necesario considerar que los resultados obtenidos de manera remota se consideran temporales, lo que implica que en el momento de realizar una adquisición en campo es posible encontrar variación en cuanto a la cobertura del suelo. Esto último se debe a la diferencia de tiempo entre la fecha de la imagen satelital y la fecha real de adquisición.

Dentro de los resultados enfocados al segundo objetivo específico, se encontró que el método de Ponderación de Distancia Inversa (IDW) es el de mejor desempeño dentro de los métodos de interpolación espacial (MIE) implementados. Esto último se determinó a partir del análisis cuantitativo del error, pero se deja una consideración especial a los resultados de Ordinary Kriging (OK) debido a que, desde un análisis cualitativo, los contornos de las anomalías se observan bien definidos. En general, se considera que el problema de datos puntuales y dispersos no se logra solucionar del todo con los MIE sugeridos ya no corresponden directamente a una solución de este caso. A pesar de esto, y teniendo en cuenta que existe una carencia de un método que solucione este problema para geofísica 2D, se puede señalar al método IDW como el de mejor rendimiento

en la generación de imágenes o anomalías geofísicas.

En relación al tercer objetivo específico, proponer una adquisición detallada sugiere una muestra del flujo de trabajo (Figura 31) propuesto y un resultado conciso (Figura 24). De esta manera, esta investigación aporta un avance en la planeación geofísica donde se logra, de manera remota, diseñar y evaluar geometrías de adquisición para los métodos de gravimetría y magnetometría con datos de libre acceso aplicados sobre un cuerpo intrusivo. Las geometrías diseñadas a partir de la zonificación de favorabilidad, pueden aumentar en puntos de adquisición de acuerdo a los tiempos logísticos que se consideren para el trabajo de campo.

Referencias

- Alarcón, C. M., Clavijo-Torres, J., Mantilla-Figueroa, L. C., & Rodríguez, J. G. (2020). Nueva propuesta de edades para el registro sedimentario de las formaciones bocas y jordán y su relación con el desarrollo de la actividad magmática del grupo plutónico de santander (cordillera oriental, colombia). *Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales*, 44(173), 1137–1151.
- Bankey, V., Cuevas, A., Daniels, D., Finn, C., Hernandez, I., Hill, P., Kucks, R., Miles, W., Pilkinton, M., Roberts, C., Roest, W., Rystrom, V., Shearer, S., Snyder, S., Sweeny, R., Velez, J., Phillips, J., & Ravat, D. (2002). Digital data grids for the magnetic anomaly map of north america.
- Bankey, V., Grauch, V., & Kucks, R. (1998). Utah aeromagnetic and gravity maps and data: a web site for distribution of data.
- Botello-Becerra, F. M., Mantilla-Figueroa, L. C., & Colegial-Gutiérrez, J. D. (2014). Aportes al conocimiento geológico del granito de durania (silúrico temprano): Macizo de santander (colombia).
- Chen, Y., Chen, X., Wang, Y., & Zu, S. (2019). The interpolation of sparse geophysical data. *Surveys in Geophysics*, 40(1), 73–105.
- Cockett, R., Kang, S., Heagy, L. J., Pidlisecky, A., & Oldenburg, D. W. (2015). Simpeg: An open source framework for simulation and gradient based parameter estimation in geophysical applications. *Computers & Geosciences*, 85, 142–154.
- de Amorim Borges, P., Franke, J., da Anunciação, Y. M. T., Weiss, H., & Bernhofer, C. (2016). Comparison of spatial interpolation methods for the estimation of precipitation distribution in distrito federal, brazil. *Theoretical and applied climatology*, 123(1-2), 335–348.

- El-Sayed Ewis, O. (2012). Improving the prediction accuracy of soil mapping through geostatistics. *International Journal of Geosciences*, 2012.
- Elumalai, V., Brindha, K., Sithole, B., & Lakshmanan, E. (2017). Spatial interpolation methods and geostatistics for mapping groundwater contamination in a coastal area. *Environmental Science and Pollution Research*, 24(12), 11601–11617.
- Fúquen, J., Ceballos, L., Marín, E., Patiño, A., & Pedraza, A. (2011). Geología de las planchas 98-durania y 99-villa del rosario norte de santander–colombia.
- Gadiraju, K. K., Ramachandra, B., Chen, Z., & Vatsavai, R. R. (2020). Multimodal deep learning based crop classification using multispectral and multitemporal satellite imagery. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, (pp. 3234–3242).
- Gardner, D. & Nichols, D. (2017). Multi-label classification of satellite images with deep learning.
- Gunarathna, M., Kumari, M., & Nirmanee, K. (2016). Evaluation of interpolation methods for mapping ph of groundwater. *International journal of latest technology in engineering, management & applied science*, 3, 1–5.
- Helber, P., Bischke, B., Dengel, A., & Borth, D. (2018). Introducing eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, (pp. 204–207). IEEE.
- Helber, P., Bischke, B., Dengel, A., & Borth, D. (2019). Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(7), 2217–2226.
- Hodam, S., Sarkar, S., Marak, A. G., Bandyopadhyay, A., & Bhadra, A. (2017). Spatial interpolation of reference evapotranspiration in india: comparison of idw and kriging methods. *Journal of The Institution of Engineers (India): Series A*, 98(4), 511–524.

- Horton, J. D., San-Juan, C. A., & Stoesser, D. B. (2017). The state geologic map compilation geodatabase of the conterminous united states.
- Houborg, R. & McCabe, M. F. (2018). A hybrid training approach for leaf area index estimation via cubist and random forests machine-learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 135, 173–188.
- Ikechukwu, M. N., Ebinne, E., Idorenyin, U., Raphael, N. I., et al. (2017). Accuracy assessment and comparative analysis of idw, spline and kriging in spatial interpolation of landform (topography): an experimental study. *Journal of Geographic Information System*, 9(03), 354.
- Jang, C.-S., Chen, S.-K., & Cheng, Y.-T. (2016). Spatial estimation of the thickness of low permeability topsoil materials by using a combined ordinary-indicator kriging approach with multiple thresholds. *Engineering Geology*, 207, 56–65.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 1097–1105.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90.
- Lee, W., Kim, D., Chae, Y., & Ryu, D. (2011). Probabilistic evaluation of spatial distribution of secondary compression by using kriging estimates of geo-layers. *Engineering geology*, 122(3-4), 239–248.
- Li, J. & Heap, A. D. (2014). Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling & Software*, 53, 173–189.
- Li, J., Heap, A. D., Potter, A., & Daniell, J. J. (2011). Application of machine learning methods to spatial interpolation of environmental variables. *Environmental Modelling & Software*, 26(12), 1647–1659.
- Lowrie, W. & Fichtner, A. (2020). *Fundamentals of geophysics*. Cambridge university press.

- Mantilla, L. C., Mendoza, H., Bissig, T., & Craig, H. (2011). Nuevas evidencias sobre el magmatismo miocénico en el distrito minero de vetas-california (macizo de santander, cordillera oriental, colombia). *Boletín de Geología*, 33(1).
- Mantilla, L. C., Valencia, V. A., Barra, F., Pinto, J., & Colegial, J. (2009). Geocronología u-pb de los cuerpos porfiríticos del distrito aurífero de vetas-california (dpto de santander, colombia). *Boletín de Geología*, 31(1), 31–43.
- Marinoni, O. (2003). Improving geological models using a combined ordinary–indicator kriging approach. *Engineering geology*, 69(1-2), 37–45.
- Pokhrel, R. M., Kuwano, J., & Tachibana, S. (2013). A kriging method of interpolation used to map liquefaction potential over alluvial ground. *Engineering geology*, 152(1), 26–37.
- Pouladi, N., Møller, A. B., Tabatabai, S., & Greve, M. H. (2019). Mapping soil organic matter contents at field level with cubist, random forest and kriging. *Geoderma*, 342, 85–92.
- Prathap, G. & Afanasyev, I. (2018). Deep learning approach for building detection in satellite multispectral imagery. In *2018 International Conference on Intelligent Systems (IS)*, (pp. 461–465). IEEE.
- Pritt, M. & Chern, G. (2017). Satellite image classification with deep learning. In *2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, (pp. 1–7). IEEE.
- Rakshit, S., Debnath, S., & Mondal, D. (2018). Identifying land patterns from satellite imagery in amazon rainforest using deep learning.
- Rojas, S. & Cramer, T. (2013). Metalogenia de las mineralizaciones auríferas en la zona de vetas, santander.
- Sanabria, L., Qin, X., Li, J., Cechet, R., & Lucas, C. (2013). Spatial interpolation of mcarthur’s forest fire danger index across australia: observational study. *Environmental modelling & software*, 50, 37–50.

- Sekulić, A., Kilibarda, M., Heuvelink, G., Nikolić, M., & Bajat, B. (2020). Random forest spatial interpolation. *Remote Sensing*, 12(10), 1687.
- Setianto, A. & Triandini, T. (2013). Comparison of kriging and inverse distance weighted (idw) interpolation methods in lineament extraction and analysis. *Journal of Applied Geology*, 5(1).
- Shi, C. & Wang, Y. (2021). Non-parametric machine learning methods for interpolation of spatially varying non-stationary and non-gaussian geotechnical properties. *Geoscience Frontiers*, 12(1), 339–350.
- Simonyan, K. & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition.
- Wang, Y., Akeju, O. V., & Zhao, T. (2017). Interpolation of spatially varying but sparsely measured geo-data: a comparative study. *Engineering Geology*, 231, 200–217.
- Ward, D., Goldsmith, R., Cruz, J., & Restrepo, H. (1973). Geología de los cuadrángulos h-12 bucaramanga y h-13 pamplona, departamento de santander. *Boletín Geológico*, 21, 34.
- Zeiler, M. D. & Fergus, R. (2014). Visualizing and understanding convolutional networks. In *European conference on computer vision*, (pp. 818–833). Springer.
- Zhao, T., Montoya-Noguera, S., Phoon, K.-K., & Wang, Y. (2018). Interpolating spatially varying soil property values from sparse data for facilitating characteristic value selection. *Canadian Geotechnical Journal*, 55(2), 171–181.

A. Análisis de terreno

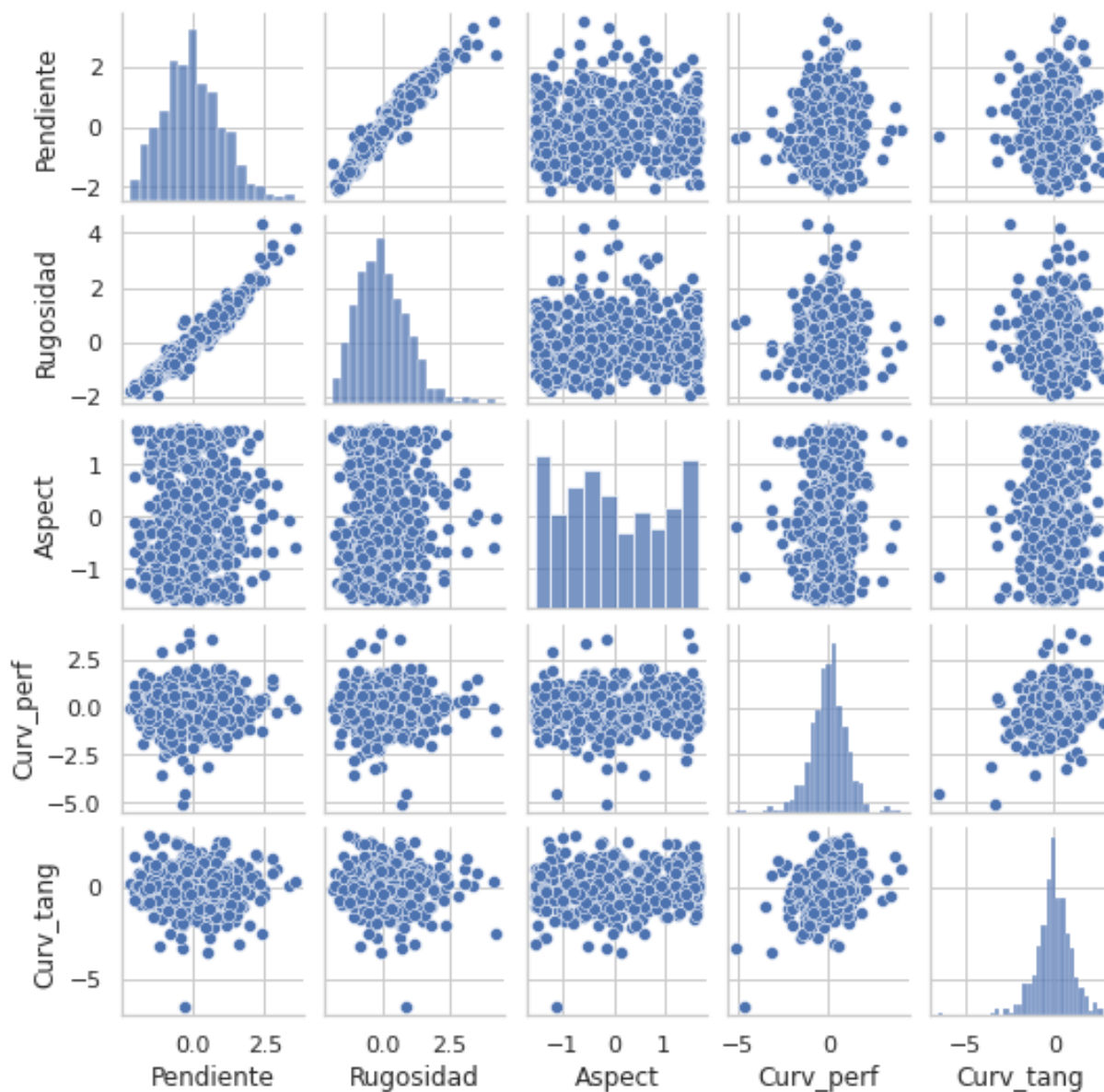


Figura A.1: Correlación de las variables para análisis de terreno.

Se observa que, en su mayoría, los datos tienen un valor menor a 0.1 que, de acuerdo a la clasificación de la RNC de análisis de terreno, corresponde a zonas donde las condiciones del terreno no serían favorables para la toma de datos.

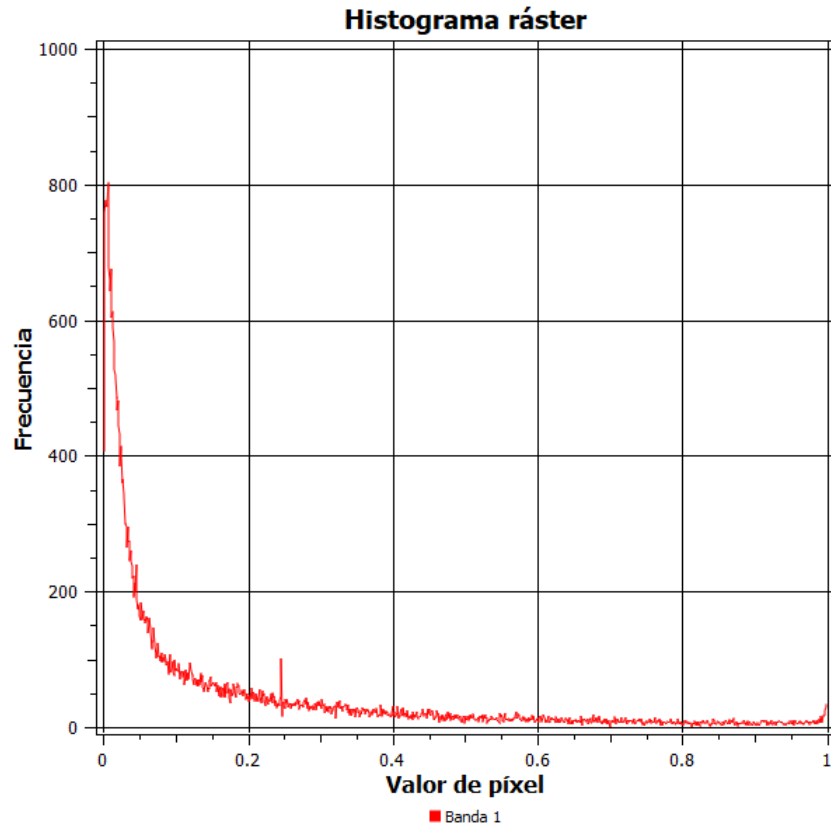


Figura A.2: Histograma de la clasificación de píxeles según la RNC de análisis de terreno.

Se observa que, en su mayoría, los datos tienen un valor menor a 0.1 que, de acuerdo a la clasificación de la RNC de análisis de terreno, corresponde a zonas donde las condiciones del terreno no serían favorables para la toma de datos.

Arquitectura/ Prueba	Entrada	Capa oculta 1			Capa oculta 2		Capa oculta 3		Salida	
	Forma	Neuronas	Activación	Kernel	Neuronas	Activación	Neuronas	Activación	Neuronas	Activación
1	(5,)	12	ReLU	he_normal	-	-	-	-	1	Sigmoide
2					-	-	-	-		
3					8	ReLU	-	-		
4					8	ReLU	-	-		
5					8	ReLU	4	ReLU		
6					8	ReLU	4	ReLU		
7					-	-	-	-		
8					-	-	-	-		
9					-	-	-	-		
10					-	-	-	-		

Tabla A.1: Arquitecturas probadas para el análisis de terreno

En la tabla se presentan las arquitecturas de RNAPM asociadas a cada una de las pruebas desarrolladas durante la consideración a los efectos del terreno.

B. Modelamiento computacional para modelo directo

Tipo de roca	g/cm ³		Susceptibilidad magnética (SI)	
	Valor promedio	Rango	Valor promedio	Rango
Granito	2.61	2.3 - 3.1	1E-1.15	1E-1.2 - 1E-2.2
Esquisto	2.42	1.75 - 2.48	1E-3.3	1E-2 - 1E-4.2
Arenisca	2.32	1.6 - 2.7	1E-3.4	1E-1.8 - 1E-5
Caliza	2.54	1.9 - 2.8	1E-3.5	1E-2.5 - 1E-4.7

Tabla B.2: Valores de referencia para propiedades físicas según tipo de roca. Los valores presentados en esta tabla fueron obtenidos de Lowrie & Fichtner (2020).

B.1. Modelamiento en función de la densidad

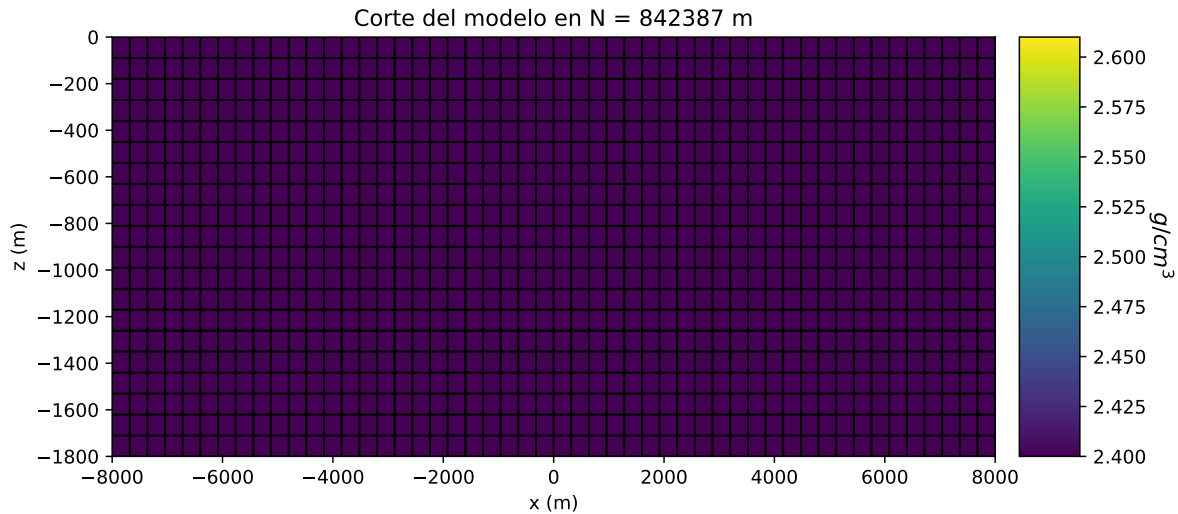


Figura B.3: Modelo de densidades en 842387 N

Modelo computacional en función de los valores de densidad para un corte geológico ubicado en la zona de estudio: Durania. La ubicación en planta de este corte se observa en la Figura 13.

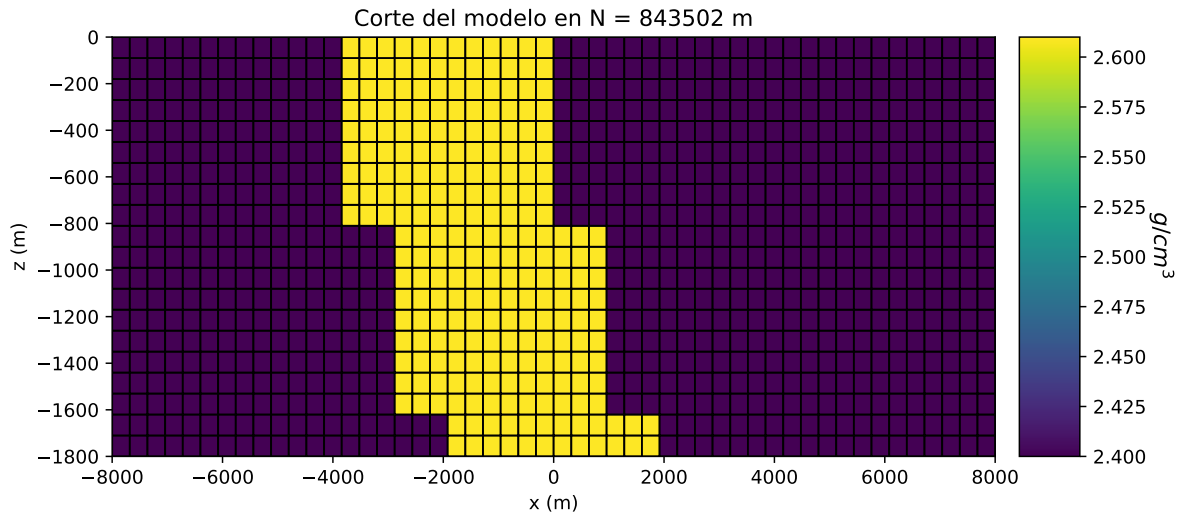


Figura B.4: Modelo de densidades en 843502 N

Modelo computacional en función de los valores de densidad para un corte geológico ubicado en la zona de estudio: Durania. La ubicación en planta de este corte se observa en la Figura 13.

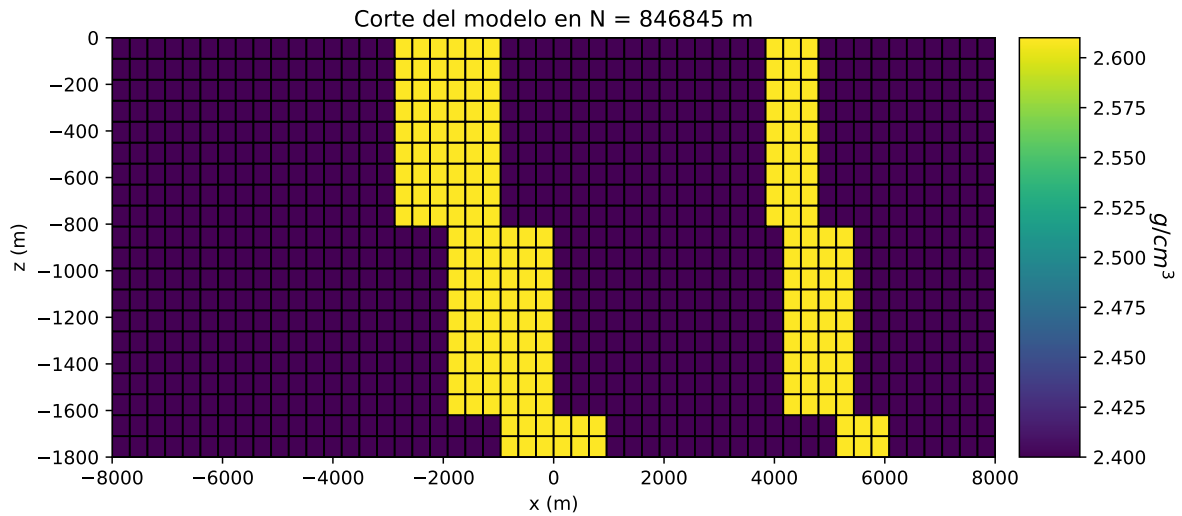


Figura B.5: Modelo de densidades en 846845 N

Modelo computacional en función de los valores de densidad para un corte geológico ubicado en la zona de estudio: Durania. La ubicación en planta de este corte se observa en la Figura 13.

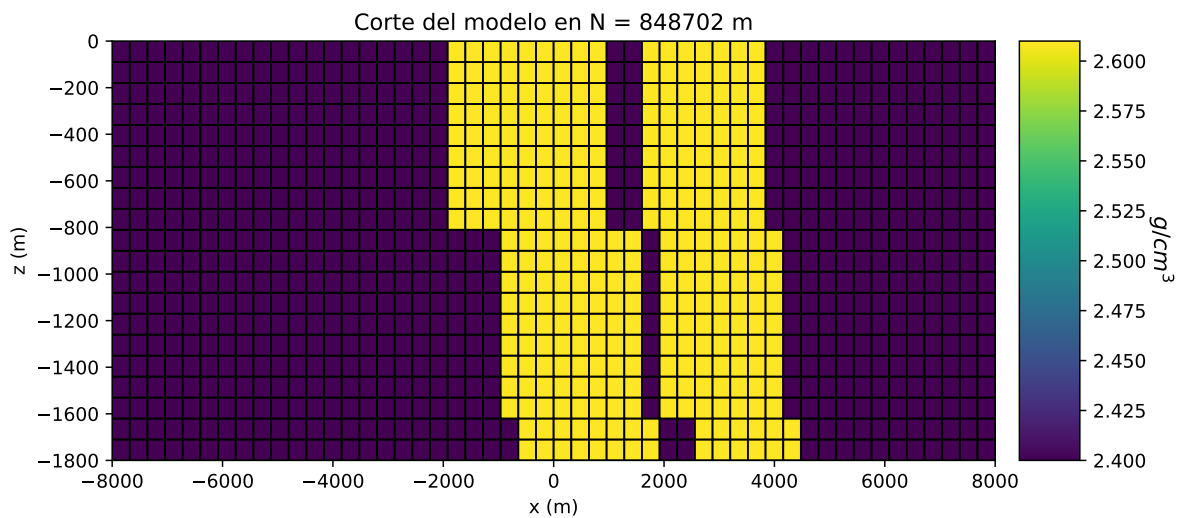


Figura B.6: Modelo de densidades en 848702 N

Modelo computacional en función de los valores de densidad para un corte geológico ubicado en la zona de estudio: Durania. La ubicación en planta de este corte se observa en la Figura 13.

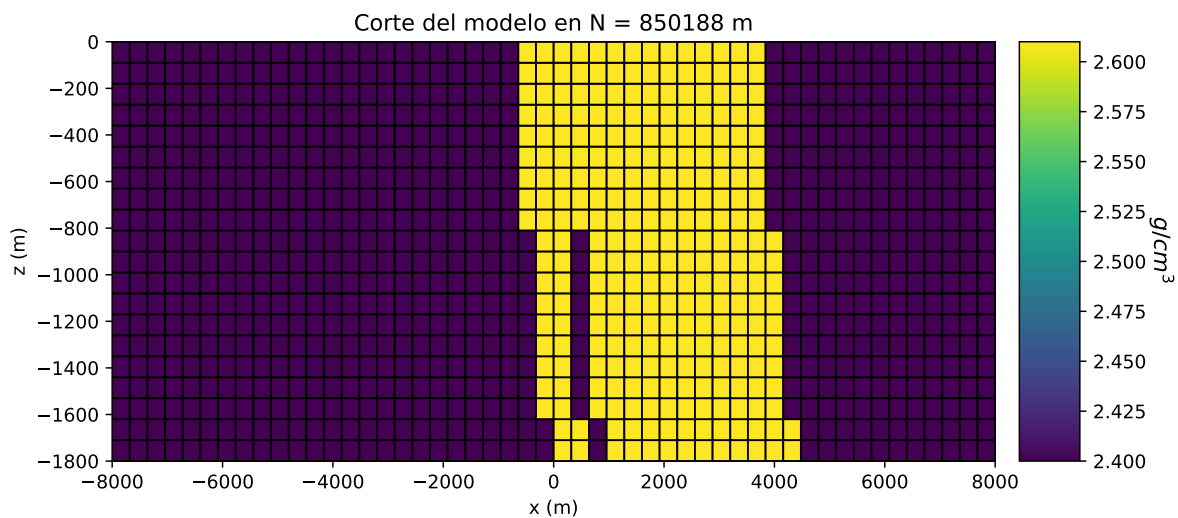


Figura B.7: Modelo de densidades en 850188 N

Modelo computacional en función de los valores de densidad para un corte geológico ubicado en la zona de estudio: Durania. La ubicación en planta de este corte se observa en la Figura 13.

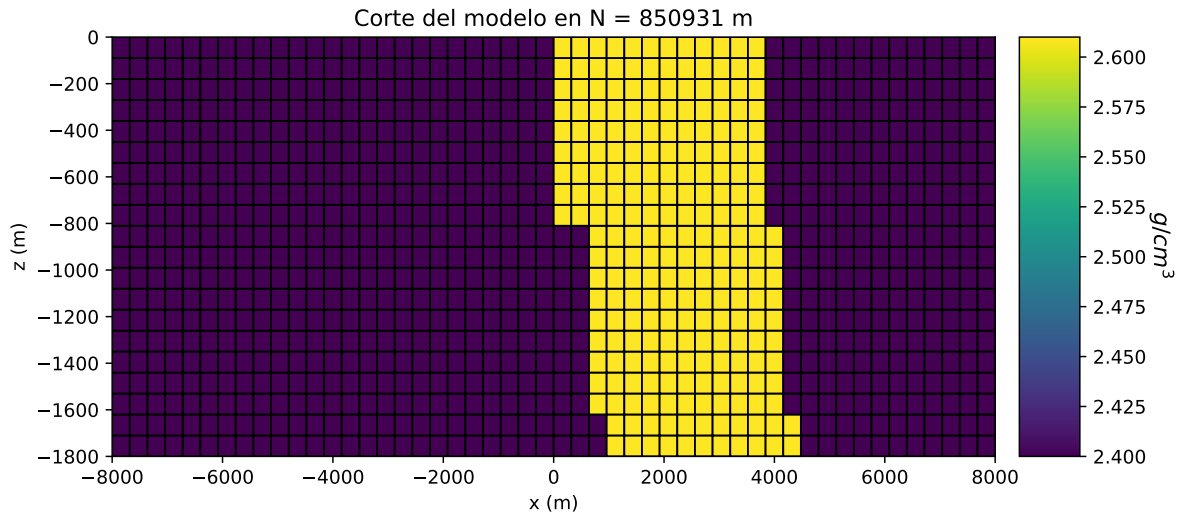


Figura B.8: Modelo de densidades en 850188 N

Modelo computacional en función de los valores de densidad para un corte geológico ubicado en la zona de estudio: Durania. La ubicación en planta de este corte se observa en la Figura 13.

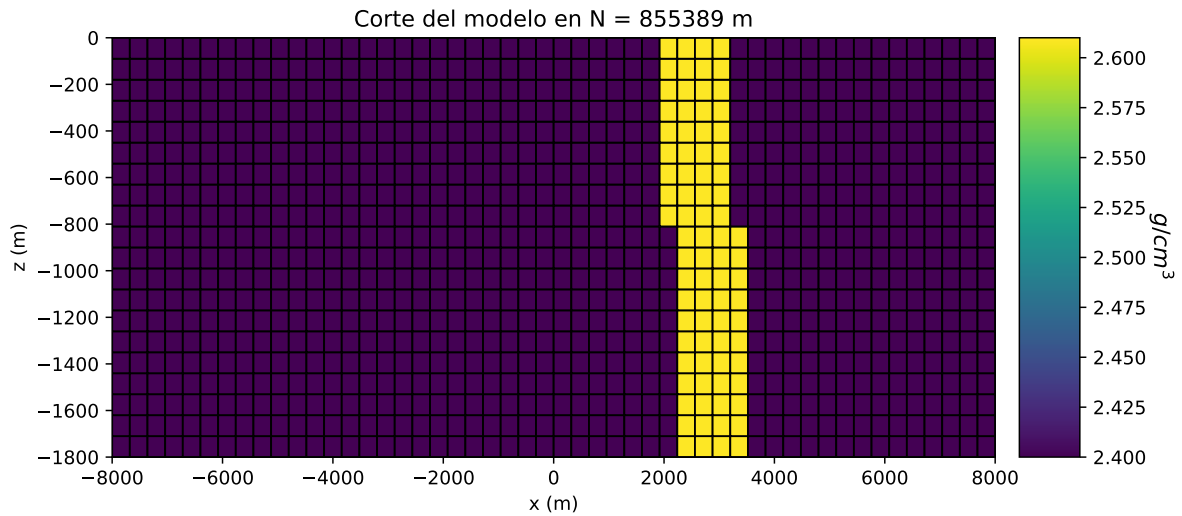


Figura B.9: Modelo de densidades en 850931 N

Modelo computacional en función de los valores de densidad para un corte geológico ubicado en la zona de estudio: Durania. La ubicación en planta de este corte se observa en la Figura 13.

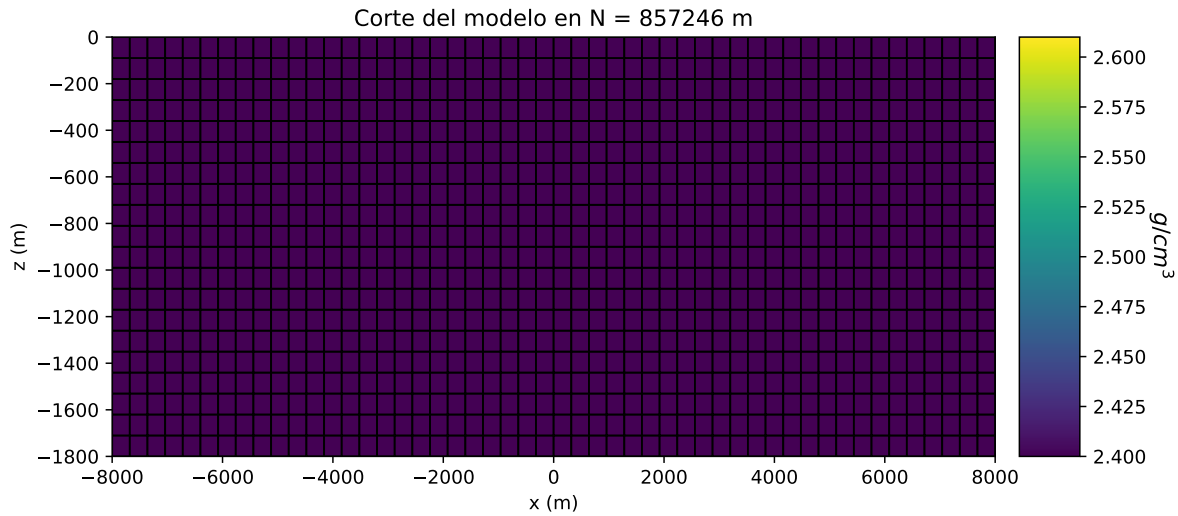


Figura B.10: Modelo de densidades en 857246 N
Modelo computacional en función de los valores de densidad para un corte geológico ubicado en la zona de estudio: Durania. La ubicación en planta de este corte se observa en la Figura 13.

B.2. Modelamiento en función de la susceptibilidad magnética

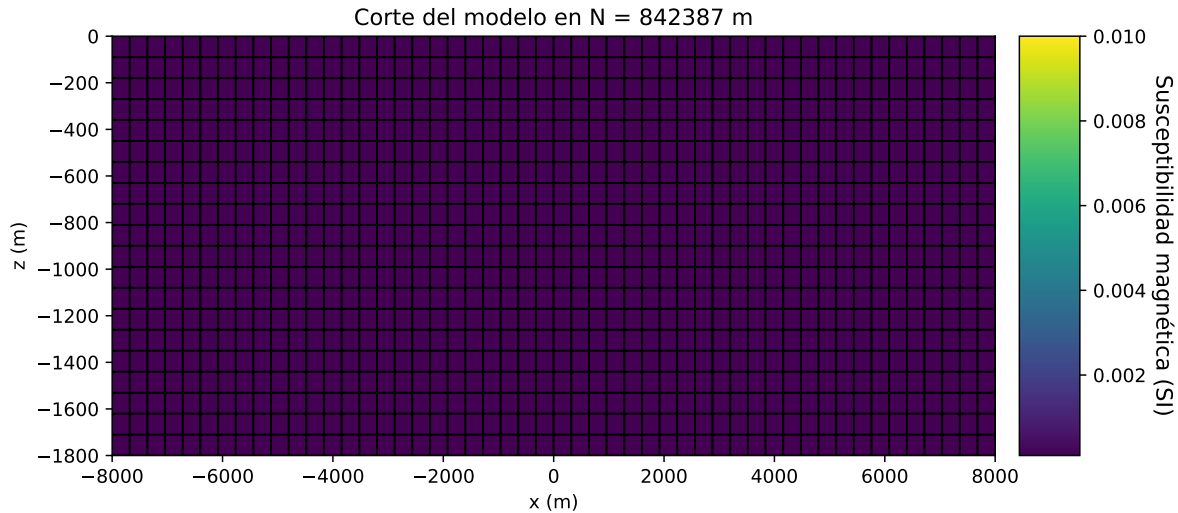


Figura B.11: Modelo de susceptibilidad magnética en 842387 N
Modelo computacional en función de los valores de densidad para un corte geológico ubicado en la zona de estudio: Durania. La ubicación en planta de este corte se observa en la Figura 13.

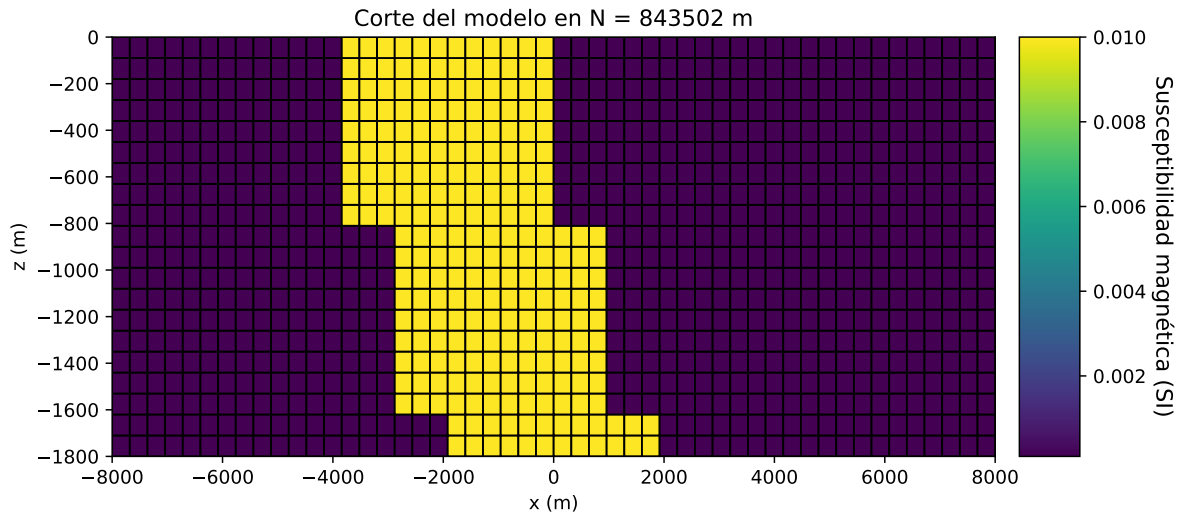


Figura B.12: Modelo de susceptibilidad magnética en 843502 N

Modelo computacional en función de los valores de densidad para un corte geológico ubicado en la zona de estudio: Durania. La ubicación en planta de este corte se observa en la Figura 13.

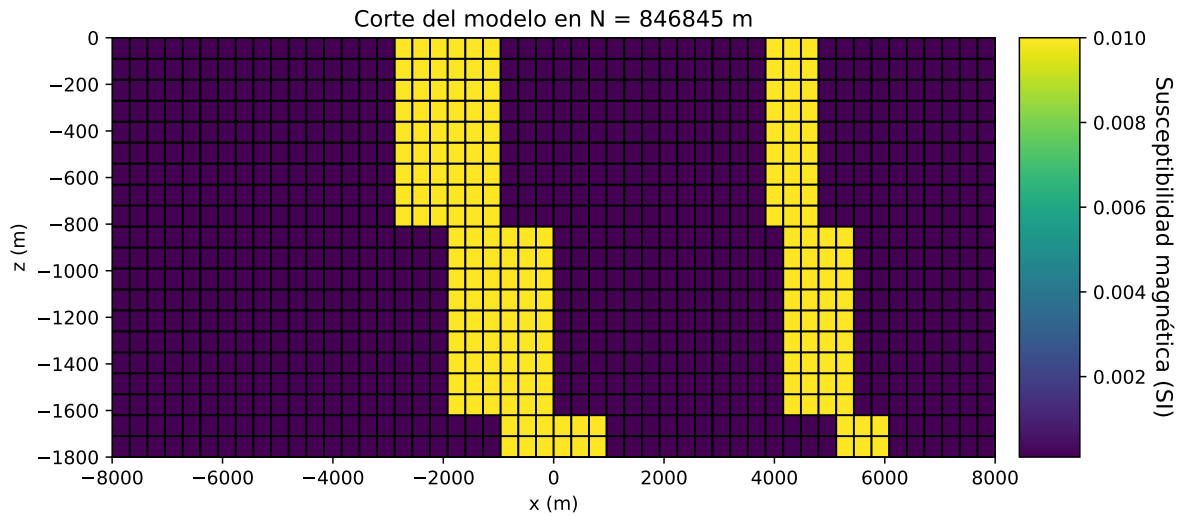


Figura B.13: Modelo de susceptibilidad magnética en 846845 N

Modelo computacional en función de los valores de densidad para un corte geológico ubicado en la zona de estudio: Durania. La ubicación en planta de este corte se observa en la Figura 13.

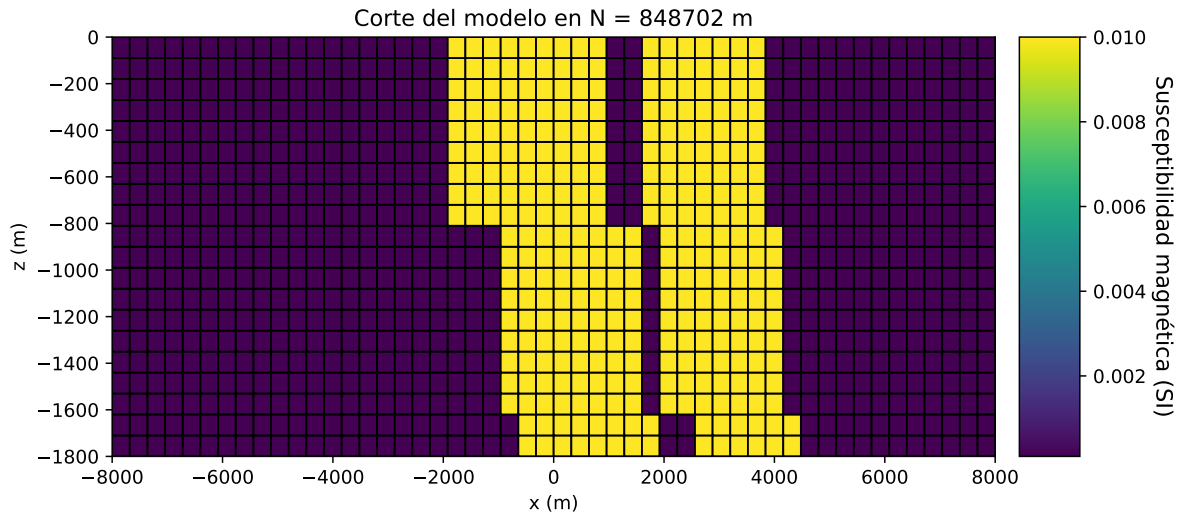


Figura B.14: Modelo de susceptibilidad en 848702 N

Modelo computacional en función de los valores de densidad para un corte geológico ubicado en la zona de estudio: Durania. La ubicación en planta de este corte se observa en la Figura 13.

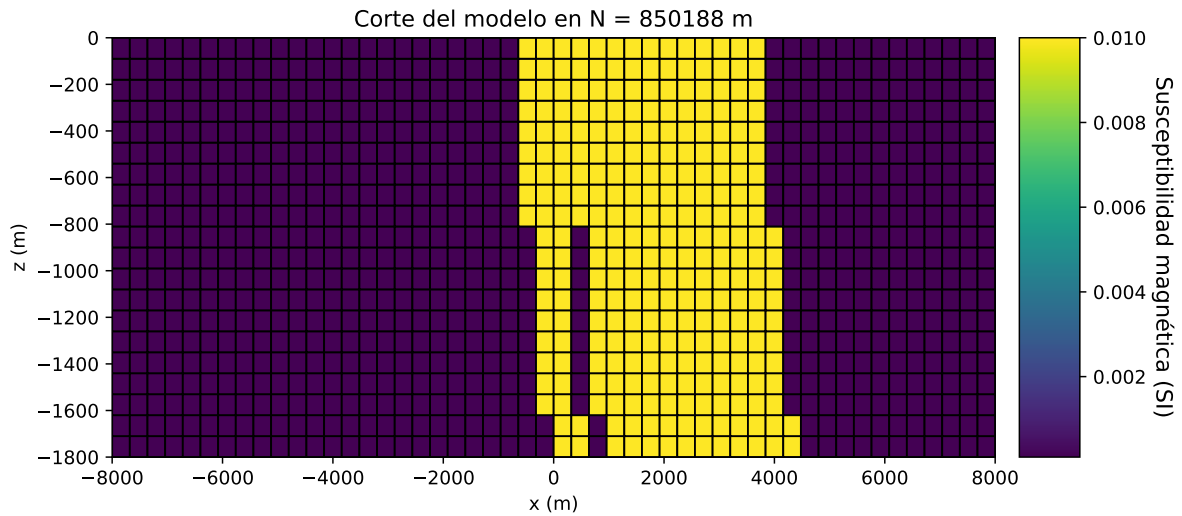


Figura B.15: Modelo de susceptibilidad magnética en 850188 N

Modelo computacional en función de los valores de densidad para un corte geológico ubicado en la zona de estudio: Durania. La ubicación en planta de este corte se observa en la Figura 13.

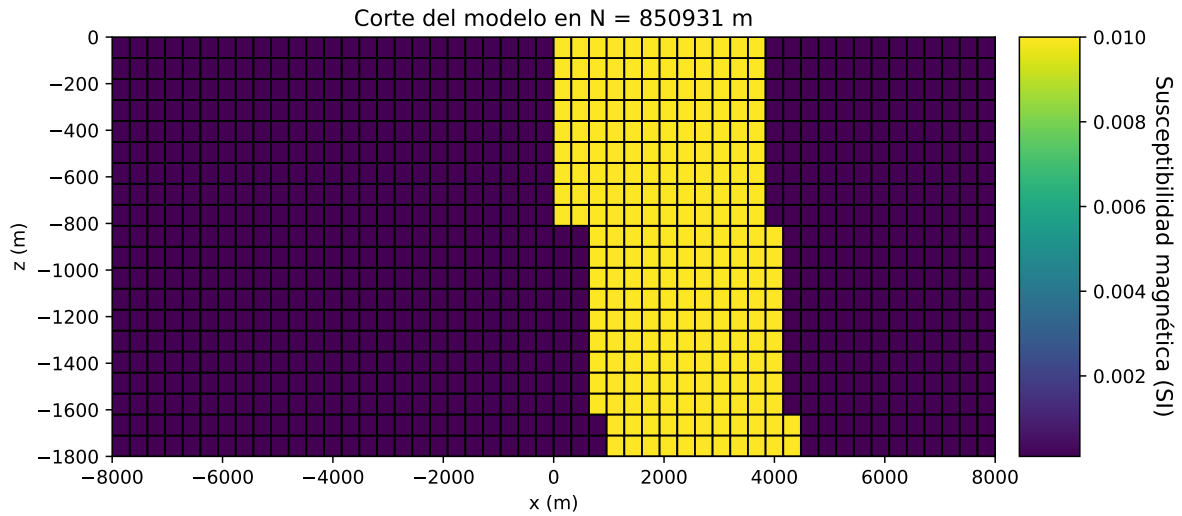


Figura B.16: Modelo de susceptibilidad magnética en 850188 N

Modelo computacional en función de los valores de densidad para un corte geológico ubicado en la zona de estudio: Durania. La ubicación en planta de este corte se observa en la Figura 13.

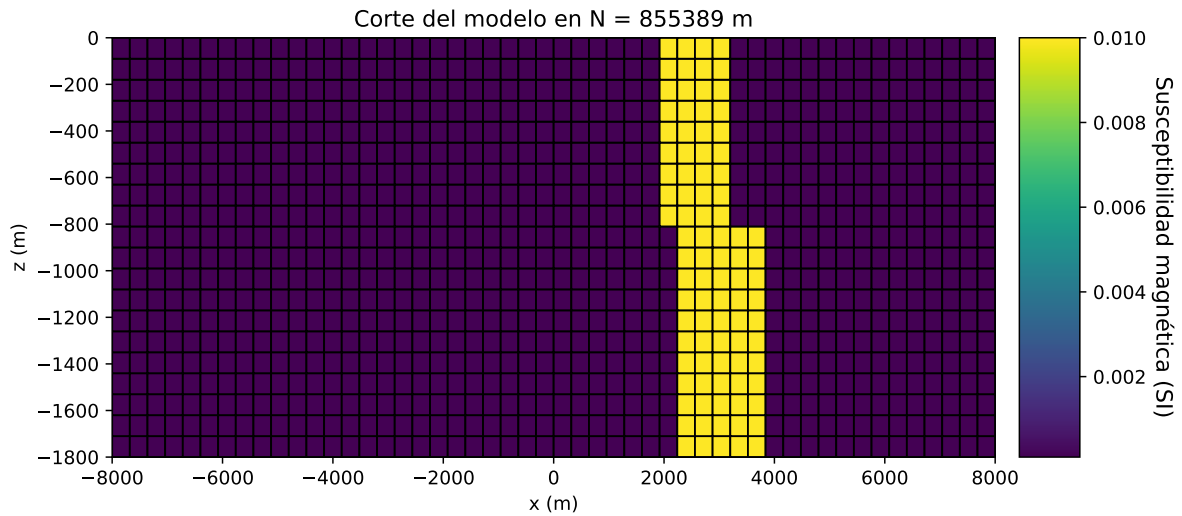


Figura B.17: Modelo de susceptibilidad magnética en 850931 N

Modelo computacional en función de los valores de densidad para un corte geológico ubicado en la zona de estudio: Durania. La ubicación en planta de este corte se observa en la Figura 13.

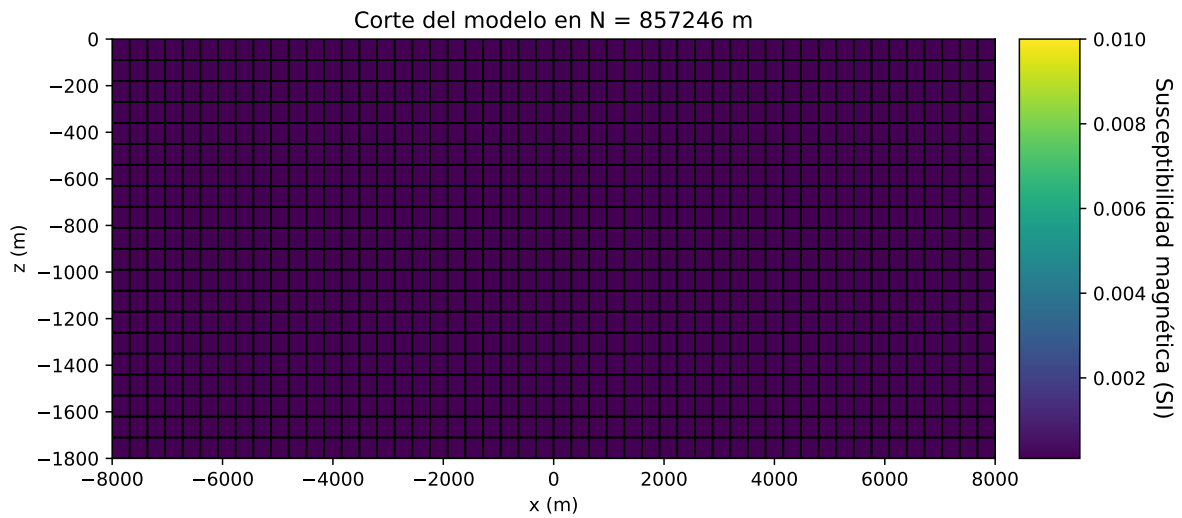


Figura B.18: Modelo de susceptibilidad magnética en 857246 N
Modelo computacional en función de los valores de densidad para un corte geológico ubicado en la zona de estudio: Durania. La ubicación en planta de este corte se observa en la Figura 13.

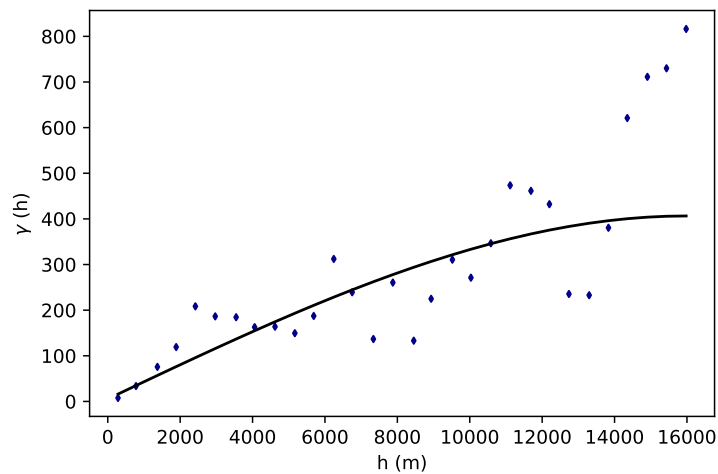


Figura B.19: Semivariograma para datos gravimétricos modelados en Durania
Modelo de ajuste esférico para datos de gravimetría modelados a partir del problema directo.

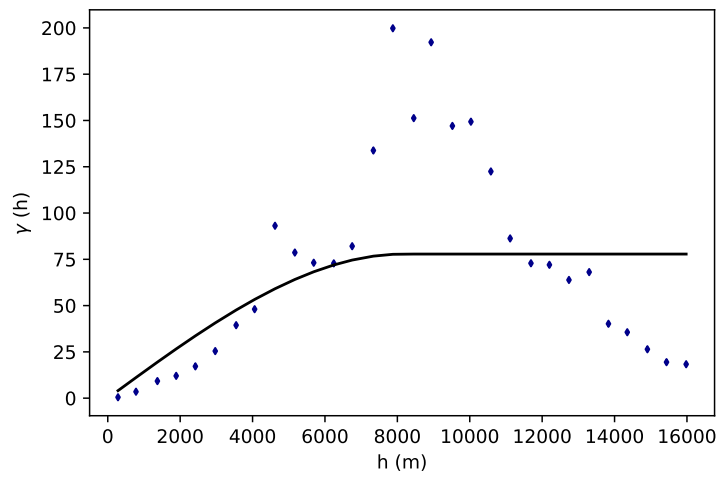


Figura B.20: Semivariograma para datos magnetométricos modelados en Durania
Modelo de ajuste esférico para datos de magnetometría modelados a partir del problema directo.