

UNIVERSIDAD ADOLFO IBAÑEZ

TESIS DE MAGÍSTER

**Estimación de concentraciones de metales pesados
en depósitos de relaves a partir de imágenes
satelitales multiespectrales**

Autor:
Tomás Ignacio Patricio
Contreras Delporte

Profesores guía:
Javier Lopatin
Raúl O’Ryan

Comité de defensa:
Sebastián Moreno
Luis Enrique Olivera

*Tesis realizada acorde a los requerimientos para el grado de
Master of Science in Data Science*

de la

Facultad de Ingeniería y Ciencias

10 de enero de 2025

UAI

FACULTAD DE
INGENIERÍA Y
CIENCIAS

UNIVERSIDAD ADOLFO IBAÑEZ

Resumen

Facultad de Ingeniería y Ciencias

Master of Science in Data Science

Estimación de concentraciones de metales pesados en depósitos de relaves a partir de imágenes satelitales multiespectrales

por Tomás Ignacio Patricio Contreras Delporte

Los depósitos de relaves con metales pesados representan un peligro latente para el ecosistema circundante, especialmente en aquellos inactivos o abandonados. Los metales de estos depósitos pueden movilizarse y afectar el entorno natural y poblaciones cercanas, lo que ha generado creciente preocupación en organismos como Sernageomin. Por lo tanto, monitorear las características fisicoquímicas de los depósitos relaves de forma constante es fundamental, pero implica grandes costos y esfuerzos en terreno. Debido a esto, el uso de datos satelitales resulta prometedor para estimar concentraciones de metales. Sin embargo, la interdependencia entre la reflectancia medida por satélites y las características de los relaves es compleja, por lo que debe evaluarse su viabilidad en Chile con datos históricos de Sernageomin. Esta tesis tiene como objetivo estimar las concentraciones de Cu, Ni, Zn, Co y As en depósitos de relaves de la región de Coquimbo utilizando imágenes satelitales multiespectrales. Se analizan 228 depósitos mediante algoritmos de regresión lineales y no lineales, incorporando variables categóricas y aspectos físicos para explorar su impacto.

Los resultados indican que los valores extremos generan distribuciones sesgadas que reducen la correlación entre bandas espectrales y concentraciones metálicas. Generalized Linear Models (GLM) muestran mayor precisión en Ni, Cu y As, logrando R^2 de 0.386, 0.076 y 0.222, respectivamente, debido a su capacidad para ajustar distribuciones de errores no Gaussianos. Partial Least Square Regression (PLSR), tradicionalmente el método más utilizado en estas aplicaciones, resulta más efectivo para Zn, con un R^2 de 0.109. Los modelos no lineales, como Generalized Additive Models (GAM), mejoran la estimación del Co, alcanzando un R^2 de 0.214. Se concluye que las concentraciones metálicas no son determinantes en la precisión de las estimaciones, siendo más relevantes las correlaciones y los valores extremos. Estos hallazgos sugieren un potencial para monitorear depósitos abandonados con propiedades homogéneas, donde la uniformidad reduce el sesgo y mejora la precisión. Sin embargo, los bajos ajustes limitan la aplicabilidad de esta metodología en contextos reales, principalmente debido a la baja correlación entre los espectros de absorción de los metales y las bandas disponibles.

Palabras clave: depósitos de relaves, teledetección, metales pesados, PLSR, GLM, GAM, imágenes multiespectrales.

UNIVERSIDAD ADOLFO IBAÑEZ

Abstract

Faculty of Engineering and Science

Master of Science in Data Science

Estimation of heavy metals concentrations in tailings dams using multispectral satellite imagery

by Tomás Ignacio Patricio Contreras Delporte

Tailings dams containing heavy metals pose a latent threat to the surrounding ecosystem, especially those that are inactive or abandoned. Metals from these deposits can mobilize and impact the natural environment and nearby populations, raising increasing concerns among organizations such as Sernageomin. Therefore, continuously monitoring the physicochemical characteristics of tailings dams is crucial, although it entails significant costs and field efforts. Consequently, the use of satellite data appears promising for estimating metal concentrations. However, the interdependence between satellite-measured reflectance and the characteristics of tailings is complex, necessitating an evaluation of its feasibility in Chile using historical data from Sernageomin. This thesis aims to estimate the concentrations of Cu, Ni, Zn, Co, and As in tailings deposits in the Coquimbo region using multispectral satellite images. A total of 228 deposits are analyzed through linear and nonlinear regression algorithms, incorporating categorical variables and physical aspects to explore their impact.

The results indicate that extreme values generate skewed distributions, reducing the correlation between spectral bands and metal concentrations. Generalized Linear Models (GLM) demonstrate higher precision for Ni, Cu, and As, achieving R^2 values of 0.386, 0.076, and 0.222, respectively, due to their ability to handle non-Gaussian error distributions. Partial Least Squares Regression (PLSR), traditionally the most widely used method in these applications, is more effective for Zn, achieving an R^2 of 0.109. Nonlinear models, such as Generalized Additive Models (GAM), improve the estimation of Co, reaching an R^2 of 0.214. It is concluded that metal concentrations are not determinant factors in estimation accuracy, with correlations and extreme values playing more significant roles. These findings suggest potential for monitoring abandoned deposits with homogeneous properties, where uniformity reduces bias and improves precision. However, the low adjustments limit the applicability of this methodology in real-world contexts, primarily due to the low correlation between the absorption spectra of metals and the available spectral bands.

Keywords: tailings dams, remote sensing, heavy metals, PLSR, GLM, GAM multispectral images.

Índice general

Resumen	I
Abstract	II
1. Introducción	1
2. Estado del arte	4
2.1. Programa Tranque y medidas en Chile	4
2.2. Teledetección y monitoreo de metales en depósitos de relaves	5
2.2.1. Estimación de metales con imágenes hiperespectrales	5
2.2.2. Estimación de metales pesados con imágenes multiespectrales .	6
2.2.3. Factores adicionales utilizados en teledetección de metales pe- sados	8
3. Hipótesis y objetivos	10
4. Metodología	11
4.1. Área de estudio	12
4.2. Obtención y selección de metales pesados	13
4.3. Adquisición y preprocesamiento de imágenes satelitales	14
4.4. Modelamiento predictivo	15
4.5. Impacto de variables ficticias y físicas en la estimación de metales . . .	18
4.6. Evaluación de modelos	18
5. Resultados	20
5.1. Distribución de metales y pre-selección de variables	20
5.2. Resultados para cada modelo	25
5.3. Influencia del tamaño y estado de los depósitos en la precisión	31
6. Discusión	34
6.1. Distribución de los metales pesados en depósitos de relaves	34
6.2. Selección e importancia de variables	35
6.3. Análisis de la estimación de los modelos	35
6.4. Análisis de las variables categóricas	37
7. Conclusiones	38
Referencias	40

Capítulo 1

Introducción

La minería en Chile se reconoce como una de las principales actividades económicas del país, y posiciona a Chile como el quinto país del mundo en dependencia de su Producto Interno Bruto (Zerené y Muñoz, 2017). Sin embargo, en los últimos años, la demanda por estos minerales ha estado en alza, lo que impulsa la explotación de yacimientos con menor ley (Vela-Almeida *et al.*, 2021). Estos depósitos contienen bajas concentraciones del mineral de interés en relación con el material extraído. Además, explotar yacimientos de menor calidad incrementa la cantidad de residuos sólidos producidos, lo que podría sobrepasar la capacidad de recuperación del ecosistema si no se regula adecuadamente. Por ello, existe una preocupación en el sector minero debido a la falta de control y monitoreo continuo de residuos sin dueño y de gran envergadura en el país. Entre los residuos generados se incluyen roca estéril, ácido sulfúrico, relaves sólidos y desechos de metales traza, que pueden dispersarse a través de derrames de lixiviación subterránea (Vela-Almeida *et al.*, 2021).

Dentro de los residuos mineros, los relaves destacan por su abundancia y longevidad, almacenándose en estructuras denominadas *depósitos de relaves*. Un ejemplo representativo es el depósito “Las Tórtolas” de Anglo American Chile, con una capacidad de 2.000 millones de toneladas y una vida útil proyectada hasta 2042 (Ada Contreras, 2022). Estos depósitos están regulados por el Servicio Nacional de Geología y Minería (Sernageomin) y el D.S. 248. Según Sernageomin, un depósito de relaves se define como: “una obra de ingeniería diseñada para satisfacer exigencias legales nacionales, de modo que se aisle completamente los sólidos (relaves) depositados del ecosistema circundante”. Por su parte, se define “relave” como: “un sólido finamente molido, que se descarta en operaciones mineras”, representando entre el 70 y 80 % de los descartes en dichas operaciones (Sernageomin, 2024a).

Aunque se consideran desechos, los relaves presentan un potencial económico, principalmente si se reutilizan o reprocesan desechos antiguos que contienen trazas significativas de metales de valor económico. Por esta razón, Sernageomin ha desarrollado y actualizado desde 2017 el informe “Geoquímica de Superficie de Depósitos de Relaves de Chile 2023” (en adelante, Reporte Geoquímico 2023). Este informe detalla que los depósitos están compuestos mayoritariamente por 12 elementos inócuos (99,06 % de la masa total), 30 elementos traza comunes (0,82 %) y 14 tierras raras elementales (0,01 %) (Sernageomin, 2023). Algunos elementos traza, clasificados como *Elementos Estratégicos* por la Unión Europea, poseen una creciente relevancia económica por su utilidad en tecnologías avanzadas y disruptivas. Entre los Elementos Estratégicos identificados en el Reporte Geoquímico 2023 se destacan el Co, Sb y Hf, siendo el cobalto un ejemplo relevante, con un precio de 6,3 US/US(Cu) en 2019 (JRI Ingeniería S.A., 2020).

En este contexto, se reporta que Chile es el tercer país con mayor cantidad de depósitos de relaves a nivel mundial, acumulando 537 millones de toneladas anuales de relaves depositados. Actualmente, se identifican 764 depósitos distribuidos en 10 regiones, con la región de Coquimbo concentrando el 49 % del total. Este predominio en Coquimbo se atribuye a la elevada actividad de pequeños productores mineros, quienes generan relaves de pequeña escala ([División de Desarrollo Sostenible del Ministerio de Minería, 2019](#)). Del total de depósitos en la región, un 11 % se encuentra en estado operativo de activo, un 84 % inactivo y un 5 % abandonado ([Equipo Programa Tranque, 2018](#)). Los depósitos inactivos y abandonados se caracterizan por la finalización de sus actividades y la ausencia de un plan de cierre legal, así como de un responsable para su monitoreo y seguridad.

De acuerdo al D.S. 148 del Ministerio de Salud, los depósitos de relaves fuera de proyectos mineros se clasifican como residuos industriales sólidos peligrosos ([Sernageomin, 2023](#)). Estos depósitos son susceptibles a inestabilidades físicas, como rebalses, fallas de taludes y erosiones internas. Ejemplo de ello ocurrieron entre 2018 y 2022, donde se registraron 24 accidentes relacionados con depósitos de relaves, principalmente en Brasil, China e India. En Chile, el incidente más reciente ocurrió en 2010, causando la muerte de cuatro personas ([Hernández Fernández, 2021](#)). Por ello, se considera fundamental realizar monitoreos continuos en depósitos inactivos y abandonados para prevenir riesgos y mitigar los impactos en el entorno ([Silva Rotta et al., 2020](#)).

Además de las fallas físicas, se identifica que los depósitos de relaves liberan compuestos químicos de connotación ambiental (CCA) mediante la acción del viento o drenajes ácidos que se infiltran en aguas circundantes. Estos compuestos se caracterizan por ser persistentes, bioacumulables, tóxicos y capaces de transportarse a largas distancias ([Sernageomin, 2023](#)). Según el Reporte Geoquímico 2023, los CCA presentes en los relaves incluyen Cu, Ni, Zn y As, con el Cu como el más abundante, registrando una concentración promedio de 2.460,38 g/t en la región de Coquimbo. Para ejemplificar el impacto, en concentraciones elevadas (i.e., > 91 g/t), el Cu puede ser tóxico para plantas y animales, afectando procesos esenciales como la fotosíntesis y la salud humana de comunidades vecinas ([Canadian Council of Ministers of the Environment, 1999](#); [Equipo Programa Tranque, 2018](#)). Adicionalmente, se observa que el cambio climático incrementa estos riesgos debido al aumento de precipitaciones proyectadas en zonas áridas, comprometiendo las toneladas de relave autorizadas por Sernageomin y la hidrología circundante ([A. Pica-Téllez, 2020](#)).

El monitoreo actual de depósitos se realiza mediante herramientas como el Estudio de Impacto Ambiental y el formulario E-700. Estas herramientas se aplican previo a la construcción del depósito y trimestralmente, respectivamente ([Sernageomin, Agencia de la Cooperación Internacional de Japón, 2007](#)). Dichos controles se enfocan principalmente en depósitos activos, los cuales suelen estar en buenas condiciones, con un monitoreo robusto, heterogéneo en sensorización y con un responsable designado. En contraste, los depósitos abandonados y gran parte de los inactivos carecen de normativas específicas para alertas tempranas de concentraciones de metales pesados ([División de Desarrollo Sostenible del Ministerio de Minería, 2019](#)). Por ello, el Plan Nacional de Depósitos de Relaves propone fortalecer el monitoreo y reportabilidad, implementar programas de remediación y fomentar el reprocesamiento, generando oportunidades económicas.

En este contexto, se reconoce la teledetección como una herramienta clave para el monitoreo de depósitos. Este método no invasivo, basado en la medición de ondas electromagnéticas, permite obtener información espectral, espacial y temporal

de manera remota, utilizando datos satelitales (Chuvieco, 2017; Sacristán Romero, 2005). Cada píxel en las imágenes obtenidas representa una superficie específica, definiendo el nivel de detalle de la captura, conocido como resolución espacial. Además, al analizar distintas longitudes de onda del espectro electromagnético, definidas como resolución espectral, se facilita la identificación y digitalización de características físicas y químicas de los depósitos (Rigol, 2000). Según el número de bandas espectrales capturadas, las imágenes se clasifican en multiespectrales e hiperspectrales. Las imágenes multiespectrales abarcan entre 4 y 10 bandas, desde el espectro visible hasta el infrarrojo cercano (NIR) e incluso algunas bandas del infrarrojo de onda corta (SWIR). Por otro lado, las imágenes hiperspectrales capturan entre 100 y 400 bandas, proporcionando un espectro casi continuo por píxel, que abarca desde el espectro visible hasta el infrarrojo medio, incluyendo el NIR y SWIR.

De esta manera, se considera que los datos obtenidos de estas estimaciones pueden servir como base de información económica inicial para la posible recuperación de elementos valiosos presentes en los relaves. Esta idea se revisa y promueve por actores como la Corporación de Fomento de la Producción. Este enfoque facilita una evaluación preliminar de las características del terreno o la presencia de metales en amplias áreas geográficas, priorizando el uso de métodos convencionales, más precisos pero también más costosos, en zonas de mayor interés (Sernageomin, 2023).

Sin embargo, se observa que el uso de datos satelitales en la teledetección de metales pesados enfrenta limitaciones técnicas que afectan la precisión de los modelos predictivos. Dependiendo de la resolución espectral y espacial de los sensores, se dificulta la detección de concentraciones bajas de metales en suelos y la identificación de áreas contaminadas pequeñas (Y. Wang *et al.*, 2022). Factores como la humedad, la cobertura vegetal y las condiciones atmosféricas alteran las reflectancias espectrales, complicando la relación entre las bandas predictoras y las concentraciones reales. Además, la heterogeneidad del suelo y su composición mineralógica enmascaran las señales asociadas a metales específicos, especialmente cuando las muestras de campo no son representativas (Yang *et al.*, 2021). Se observa también que los valores de reflectancia capturados por los sensores solo representan la superficie del suelo, lo que impide la detección de metales en capas profundas (Lee *et al.*, 2016).

Estas limitaciones generan incertidumbre en las estimaciones, especialmente en depósitos con condiciones geológicas complejas. Por ello, se identifica la necesidad de desarrollar investigaciones específicas que evalúen la viabilidad de los datos satelitales según las particularidades del área y los objetivos de cada estudio.

Capítulo 2

Estado del arte

En los últimos años, la teledetección ha emergido como una herramienta clave para el monitoreo y análisis de depósitos de relaves, especialmente mediante el uso de tecnologías multiespectrales e hiperespectrales. A continuación, se presenta un análisis del estado del arte relacionado con estas técnicas, estructurado en dos secciones principales: las iniciativas y medidas implementadas en Chile, con énfasis en el Programa Tranque, y las aplicaciones de la teledetección en el monitoreo de metales pesados.

2.1. Programa Tranque y medidas en Chile

Para abordar este desafío, se implementa el Plan Nacional de Depósitos de Relaves para una Minería Sostenible, liderado por el Ministerio de Minería, con el objetivo de gestionar los depósitos desde su construcción hasta su cierre. Este plan considera depósitos activos, inactivos y abandonados, y se organiza en tres ejes principales. Entre sus propuestas se incluyen la creación de un observatorio nacional para monitoreo y fiscalización continua mediante tecnología satelital, y la modificación del artículo 2° del Decreto Ley N° 3.525 para ampliar las funciones de Sernageomin ([División de Desarrollo Sostenible del Ministerio de Minería, 2019](#)).

En paralelo, se desarrolla el Programa Tranque, liderado por la Fundación Chile entre diciembre de 2017 y abril de 2024, considerado uno de los mayores avances técnicos en esta área ([Ada Contreras, 2022](#)). Este programa establece un sistema estandarizado de monitoreo y alerta temprana, centrado en la estabilidad física de los depósitos y en su impacto ambiental ([Equipo Programa Tranque, 2018](#)). Para alcanzar este objetivo, se integran diversas tecnologías para capturar, procesar y analizar datos en tiempo real, destacando tres componentes principales.

- **Sensores locales:** Se instalan sensores en puntos críticos para medir variables como pH, oxígeno disuelto, metales pesados y temperatura. Los datos recolectados se transmiten de forma continua a una plataforma central, lo que permite detectar anomalías y riesgos potenciales de manera eficiente.
- **Imágenes satelitales y drones:** Estas herramientas permiten evaluar la estabilidad estructural y detectar cambios físicos en la superficie de los depósitos, como desplazamientos, agrietamientos o variaciones topográficas, con alta precisión.
- **Modelos predictivos:** Los datos recopilados se procesan mediante algoritmos predictivos que calculan la probabilidad de fallas estructurales o eventos de

contaminación. Estos modelos optimizan las acciones preventivas y contribuyen a reducir los riesgos ambientales.

En los depósitos activos, el sistema también integra datos operativos proporcionados por las compañías mineras, que suelen ser voluminosos y complejos. Estos datos son procesados en una plataforma avanzada de gestión, lo que permite realizar análisis en tiempo real y mejorar la toma de decisiones. Además, esta integración contribuye a identificar riesgos operativos de manera temprana, incrementando la seguridad y eficiencia en la gestión de relaves ([Transparent Tailings Initiative Team, 2019](#)).

En diciembre de 2018, se implementa un sistema piloto del Programa Tranque en el depósito de relaves El Mauro, operado por minera Los Pelambres. Este piloto evalúa el monitoreo local y centralizado bajo condiciones reales, optimizando componentes como sensores, transmisión de datos y algoritmos predictivos. En abril de 2024, se concluye el desarrollo del sistema de monitoreo y alerta temprana, que opera actualmente bajo la supervisión de Sernageomin. Este sistema, conectado al depósito El Mauro, evidencia los avances en Chile respecto al monitoreo continuo de depósitos de relaves ([FundaciónChile, 2024](#)). El Programa Tranque ha sido un referente para la formulación de políticas públicas. La Política Nacional Minera 2050, por ejemplo, establece que, para 2030, el 100 % de los depósitos operativos de la gran y mediana minería deben contar con un plan integral de monitoreo ([Ministerio de Minería, 2021](#)). Este avance reafirma el compromiso de Chile con una minería más segura, sostenible y tecnológicamente avanzada.

2.2. Teledetección y monitoreo de metales en depósitos de relaves

La teledetección es una herramienta clave para monitorear concentraciones de metales pesados en suelos, utilizando sensores con diferentes resoluciones espectrales. Estos sensores capturan imágenes que se clasifican en RGB, multiespectrales e hiperspectrales. Estudios recientes han aplicado esta tecnología en diversos entornos, incluyendo tierras agrícolas para estimar metales como Cr, Cu, Pb, Cd y Ni ([F. Mirzaei et al., 2021](#); [L. Wang et al., 2022](#); [Tan et al., 2020](#)), áreas urbanas montañosas ([Yang et al., 2021](#)) y sitios mineros de oro, plomo-zinc y cobre ([Choe et al., 2008](#); [Tan et al., 2021](#); [Song et al., 2020](#)). Esto demuestra la versatilidad de la teledetección para monitorear distintas condiciones del suelo, complementada por algoritmos de regresión que mejoran su precisión.

2.2.1. Estimación de metales con imágenes hiperespectrales

Las imágenes hiperespectrales, debido a su alta resolución espectral, ofrecen información detallada para el análisis cuantitativo de metales pesados en el suelo. Esta tecnología desempeña un papel crucial al identificar concentraciones específicas mediante algoritmos de regresión, proporcionando resultados precisos y consistentes ([Lin et al., 2019](#); [Z.E. Mashimbye, 2012](#)).

En algunos estudios, se emplean sensores *in situ* para capturar imágenes hiperespectrales y estimar concentraciones de metales pesados en hojas de uva. Se analizan un total de 420 muestras, identificando la influencia de metales como Cu, Zn, Pb, Cr y Cd en las características espectrales de las regiones visible e infrarrojo cercano. Los resultados indican que la estimación de Zn es más precisa que la de Cu, con valores

de R^2 de 0.7 y 0.52, respectivamente (M. Mirzaei *et al.*, 2019). Estos resultados evidencian el potencial de los sensores hiperespectrales para aplicaciones específicas.

En contra parte, diversos sensores hiperespectrales también han sido instalados en satélites y drones. El sensor HyMAP destaca por su alta resolución espectral y espacial, logrando estimaciones precisas de metales pesados (Choe *et al.*, 2008; Y. Wang *et al.*, 2022). Por su parte, HySpex proporciona datos detallados en áreas pequeñas con una resolución espacial de 1 metro (Lassalle *et al.*, 2021). El satélite Gaofen-5 ha demostrado eficacia para predecir concentraciones de Cu en suelos mineros (Sun *et al.*, 2023), mientras que drones equipados con espectrómetros hiperespectrales han sido utilizados para mapeos locales a alta resolución (Gan *et al.*, 2023). Estos dispositivos permiten detectar metales como As, Cd, Cu, Pb, Ni y Zn en suelos de diversas características.

Junto a los sensores hiperespectrales, se aplican diferentes algoritmos para estimar metales pesados en diversos contextos, reflejando variabilidad en sus resultados. Por ejemplo, se utiliza la Regresión de Mínimos Cuadrados Parciales (PLSR) para mapear propiedades del suelo como arcilla, arena y carbonato de calcio. Este método logra alta precisión en la estimación del carbonato de calcio ($R^2 = 0.76$), mientras que muestra menor exactitud para la arena ($R^2 = 0.2$) (Gomez *et al.*, 2012). Además, al combinar PLSR con algoritmos genéticos (GA-PLSR), se mejoran las estimaciones para Ni y Zn en suelos de áreas industriales, alcanzando valores de R^2 entre 0.4 y 0.8. Sin embargo, con datos de menor resolución espectral, se observa pérdida de precisión y menor sensibilidad para detectar concentraciones bajas (Y. Wang *et al.*, 2022). Estos resultados destacan la importancia de la calidad del detalle espectral en la eficacia del modelo.

Otro enfoque ampliamente utilizado es la Regresión Lineal Múltiple (MLR), cuyo desempeño varía según el metal y las condiciones del área analizada. En terrenos industriales, se observan estimaciones de Cu y Zn con altos valores de R^2 entre 0.8 y 0.9 (Lassalle *et al.*, 2021). En contraste, en suelos mineros de oro, con 69 muestras recolectadas, metales como Pb, Zn y Cu, obtienen valores de R^2 menores (entre 0.5 y 0.65), mientras que el As alcanza un R^2 de 0.876 (Choe *et al.*, 2008). Al mismo tiempo, se aplican transformaciones logarítmicas para mitigar sesgos hacia valores extremos en los contenidos de metales, logrando mejorar las estimaciones en ambos casos. Estos resultados indican que el desempeño de los modelos depende más de las propiedades del metal y del suelo que del algoritmo utilizado.

Finalmente, un estudio que analiza 66 muestras en el depósito de relaves Tongkuangyu utiliza el satélite GF5 y el método PLSR para estimar concentraciones de Cu mediante la segmentación de datos según umbrales de concentración. Se identifica una alta correlación entre las bandas de 2209 nm y 907 nm y el contenido de Cu ($r = 0.76$). Además, la segmentación mejora la precisión de las estimaciones, incrementando el R^2 de 0.59 a 0.77 (Yin *et al.*, 2021). Estos hallazgos indican que la segmentación representa una estrategia prometedora y destacan la importancia de validar este enfoque en distintas regiones geográficas.

2.2.2. Estimación de metales pesados con imágenes multiespectrales

Las imágenes multiespectrales representan una alternativa viable para el monitoreo por teledetección de metales pesados en suelos. A diferencia de las imágenes hiperespectrales, estas poseen menor resolución espectral, lo que limita su capacidad para identificar características químicas específicas (Chen *et al.*, 2022; Y. Wang *et al.*,

2024). Sin embargo, las imágenes hiperespectrales presentan baja resolución espacial y limitada cobertura temporal, lo que las hace menos adecuadas para estudios que requieran analizar áreas extensas de forma continua. En este contexto, las imágenes multiespectrales son preferidas debido a su mayor rentabilidad, facilidad de procesamiento y amplia cobertura espacial y temporal. Estas características impulsan su uso en el monitoreo de metales pesados en suelos, a pesar de su menor resolución espectral (Xu *et al.*, 2024; Lovynska *et al.*, 2024).

Los datos multiespectrales son obtenidos principalmente a través de satélites como Landsat 7, 8, y 9, utilizados para identificar concentraciones como Cu, Pb y Zn, además de analizar la distribución espacio-temporal de metales pesados (D'Emilio *et al.*, 2012; Mouazen *et al.*, 2021). Por ejemplo, un estudio utiliza Landsat 8 para mapear la acumulación espacio-temporal de metales tóxicos en diferentes escalas (Peng *et al.*, 2016; T. Shi *et al.*, 2021; Xu *et al.*, 2024). Otro satélite ampliamente utilizado es Sentinel-2, que permite estimar concentraciones y mapear distribuciones de metales pesados (F. Mirzaei *et al.*, 2021; L. Wang *et al.*, 2022; T. Shi *et al.*, 2021). Además, estudios han integrado imágenes de Sentinel-2A y Landsat 8, mejorando la precisión de predicciones. En el caso del Zn, el R^2 incrementa de 0.19 con Landsat 8 a 0.53 al integrar ambos satélites (Khosravi *et al.*, 2022). Así, las imágenes multiespectrales resultan fundamentales para identificar una amplia gama de metales en suelos, incluidos As, Cd, Co, Cr, Cu, Fe, Hg, Mn, Ni, Pb, Sb, V y Zn.

En paralelo, la relación entre longitudes de onda específicas y áreas de absorción, en este tipo de imágenes, ha permitido el desarrollo de diversos algoritmos estadísticos y de aprendizaje automático para estimar concentraciones de metales. Por ejemplo, la Regresión Lineal Múltiple por Pasos (SMLR) y Mínimos Cuadrados Parciales (PLS) han mostrado correlaciones moderadas entre valores medidos y predichos. A modo ejemplo, PLS alcanza un R^2 de 0.563 y 0.564 para Ni y Zn, respectivamente, mientras que SMLR obtiene valores similares de 0.564 y 0.570 (Guan *et al.*, 2019). En otro estudio, que analiza 435 muestras de suelo urbano, los modelos Random Forest (RF), PLSR y Support Vector Machine (SVM) son comparados. RF obtiene mejores resultados con un R^2 de 0.78 para Ni y 0.8 para Cu, superando a PLSR y SVM que obtienen un R^2 de 0.4 y 0.55 para el Ni, respectivamente, y de 0.55 y 0.5 para el Cu (Mouazen *et al.*, 2021). Estos resultados reflejan la importancia del algoritmo en función del contexto.

Estudios en áreas mineras cupríferas con 138 muestras de suelo evalúan PLSR, ANN (Red Neuronal Artificial) y SVR (Support Vector Regression). PLSR muestra el mejor desempeño para Cu ($R^2 = 0.493$), superando a ANN ($R^2 = 0.262$) y SVR ($R^2 = 0.279$) (Fang *et al.*, 2018). Otros trabajos, que analizan 516 muestras a 20 cm de profundidad en áreas urbanas, evalúan modelos híbridos como LASSO-GA-BPNN (Least Absolute Shrinkage and Selection Operator con algoritmos genéticos y redes neuronales de retropropagación). Este modelo muestra menor error en comparación con SVR, RF y técnicas de interpolación espacial (S. Shi *et al.*, 2022). Además, un estudio con 971 muestras de suelos cercanos a lagos concluye que PLSR y BPNN tienen un desempeño similar, con R^2 de 0.526 y 0.161 para As y Hg en el caso de PLSR, frente a R^2 de 0.499 y 0.156 en BPNN (Zhao *et al.*, 2021).

Por otra parte, las técnicas geoestadísticas son herramientas útiles en ausencia de datos auxiliares o cuando las variables objetivo presentan patrones uniformes. Estas técnicas han logrado estimaciones con un R^2 de 0.62 para Cu, destacándose como métodos efectivos de interpolación espacial (Zeng *et al.*, 2021). Estas herramientas

complementan los algoritmos mencionados al proporcionar estimaciones en áreas no muestreadas.

En conclusión, la precisión de los modelos depende de factores como las características y concentraciones de los metales, así como de la cantidad y variabilidad de las muestras. En consecuencia, cada modelo posee fortalezas y limitaciones que deben evaluarse en función del contexto y los objetivos del estudio.

2.2.3. Factores adicionales utilizados en teledetección de metales pesados

La reflectancia difusa de los metales pesados en los depósitos de relaves representa un desafío relevante en la teledetección. Este fenómeno está influenciado por propiedades del material presentes en cada depósito, como las estructuras cristalinas de los minerales, la concentración de metales, el contenido de humedad y el grado de oxidación. Estas características condicionan la forma en que los metales reflejan la luz y dificultan su detección precisa mediante métodos espectrales.

Por ejemplo, un estudio muestra que las predicciones de As mediante espectroscopía son más precisas en hojas de arroz ($R^2 = 0.54$) que en suelos ($R^2 = 0.42$) utilizando PLSR (Khosravi *et al.*, 2022). Otro análisis destaca que la mineralogía de los suelos mineros afecta las señales espectrales, observándose mayor adsorción de metales como Cu y Zn en arcillas en comparación con minerales tipo *skarn* (Kim *et al.*, 2020). Estos hallazgos destacan que las propiedades físicas y químicas del suelo desempeñan un papel crucial en la variabilidad de las características espectrales.

En respuesta a esta complejidad, diversos estudios han investigado la relación entre el contenido de metales pesados en suelos y los espectros de la vegetación en crecimiento para mejorar las estimaciones. Un enfoque incluye el uso de índices espectrales como herramientas para estimar concentraciones metálicas. Aunque no infalibles, estos índices han demostrado ser valiosos al combinarse con enfoques avanzados como ANN o PLSR (F. Wang *et al.*, 2018). Por ejemplo, el Índice de Vegetación de Diferencia Normalizada ha permitido correlacionar las reflectancias espectrales con el contenido de metales en hojas, destacándose las imágenes hiperespectrales por su eficacia para mapear la contaminación en sitios industriales (D'Emilio *et al.*, 2012; Guan *et al.*, 2019; Lassalle *et al.*, 2021). Además, el Índice de Vegetación Mejorado ha sido útil para evaluar la contaminación por Cd, Pb y As en cultivos de arroz, mientras que otros índices como el Índice de Agua de Diferencia Normalizada y FA-PAR han permitido detectar variaciones fenológicas relacionadas con el estrés por metales pesados en estos cultivos (Zhang *et al.*, 2018; Zhou *et al.*, 2017).

En algunos estudios, se han adoptado enfoques alternativos para mejorar la precisión en la estimación de metales pesados. Un ejemplo incluye el uso de imágenes satelitales de GF-2 y PLSR, donde se utilizan metales pesados como variables predictoras adicionales. En este estudio se recolectan 22 muestras de un depósito de relave. Así, este enfoque identifica una correlación de 0.5 entre el Ni con las bandas B1–B4, lo que permite emplear este elemento como variable inicial para modelar otros metales relacionados. Pese a las mejoras en la estimación para el Ni de un R^2 de 0.357 a 0.443, el modelo tiene dificultades en la estimación de metales con baja concentración, como As y Hg, donde se obtiene un R^2 de 0.168 y 0.176, respectivamente. Los resultados sugieren que el uso de imágenes hiperespectrales y aumentar el número de muestras podrían ampliar las capacidades predictivas al incorporar factores adicionales de modelado (Rui *et al.*, 2021).

Con el objetivo de explorar nuevas alternativas, un estudio combina datos de elevación (altitud, pendiente y aspecto) con bandas espectrales en un modelo basado en algoritmos genéticos y redes neuronales de retropropagación. Este enfoque logra incrementar los valores de R^2 de 0.41 a 0.59 para Cu, de 0.29 a 0.35 para Pb y de 0.28 a 0.42 para As (Yang *et al.*, 2021). Asimismo, un análisis comparativo entre PLS y SMLR para estimar metales como Cu, Cr, Mn, Ni, V y Zn, demuestra que SMLR alcanza mejores resultados, aumentando el R^2 de 0.325 a 0.368 para Cu al incluir datos de elevación (Guan *et al.*, 2019).

Finalmente, se ha evaluado el uso de imágenes térmicas para detectar estrés en cultivos de arroz causado por metales pesados. Un modelo espacio-temporal (TSFS) clasifica niveles de contaminación con más del 90 % de precisión, demostrando que estas imágenes, aunque no estiman directamente el contenido metálico, podrían ser una valiosa variable adicional en modelos predictivos futuros (Liu *et al.*, 2016).

En conclusión, tomando en cuenta los antecedentes presentados, los diferentes modelos y metodologías presentan precisiones variables según el contexto de cada estudio. Las concentraciones y distribuciones de los metales, junto con las características de las áreas muestreadas, son factores clave que influyen en los resultados obtenidos, más que el modelo seleccionado. Por ello, la selección de enfoques debe considerar tanto las características del sitio como los objetivos específicos del análisis. Lo que abre la posibilidad de realizar estimaciones de metales en depósitos de relaves en Chile con datos históricos de Sernageomin.

Capítulo 3

Hipótesis y objetivos

Hipótesis: Los metales pesados con mayores concentraciones y menor variabilidad en su distribución pueden ser estimados con mayor precisión en depósitos de relaves utilizando imágenes satelitales.

Objetivo: El objetivo general del presente trabajo es estimar las concentraciones de Cu, Ni, Zn, Co y As en depósitos de relaves de la región de Coquimbo usando imágenes satelitales multiespectrales.

Para la correcta ejecución de este objetivo se proponen los siguientes **objetivos específicos**:

- Compilar y procesar datos sobre las concentraciones de metales pesados y las dimensiones superficiales de los depósitos de relaves.
- Analizar y comparar el desempeño de distintos modelos de regresión en la estimación de concentraciones de metales pesados en depósitos de relaves.
- Analizar la influencia del tamaño de los depósitos y de su estado operativo (activo, inactivo o abandonado) en la precisión de la estimación de las concentraciones de metales pesados.

Capítulo 4

Metodología

Para compilar las características de metales y depósitos de relaves, estimar sus concentraciones y analizar la influencia del tamaño y el estado operativo de los relaves, se utiliza un flujo de seis pasos (Figura 4.1). Primero, se delimita el área de estudio a los depósitos de relaves de la región de Coquimbo, caracterizada por su intensa actividad minera y diversidad climática (1). Segundo, se obtienen datos *in situ* de concentraciones químicas a partir de estudios previos realizados por Sernageomin, asegurando la estandarización de las muestras (2). Tercero, se adquieren imágenes satelitales multiespectrales y se aplican técnicas de preprocesamiento para garantizar la calidad de los datos espectrales (3). Cuarto, se implementan modelos de regresión lineal como PLSR, Generalized Linear Models (GLM) y Generalized Additive Models (GAM) para explorar relaciones no lineales en la estimación de metales pesados (4). Quinto, se evalúa el impacto de la incorporación de variables categóricas y físicas (5). Sexto, se analizan métricas como R^2 , RMSE y Bias para seleccionar el modelo con mayor precisión (6).

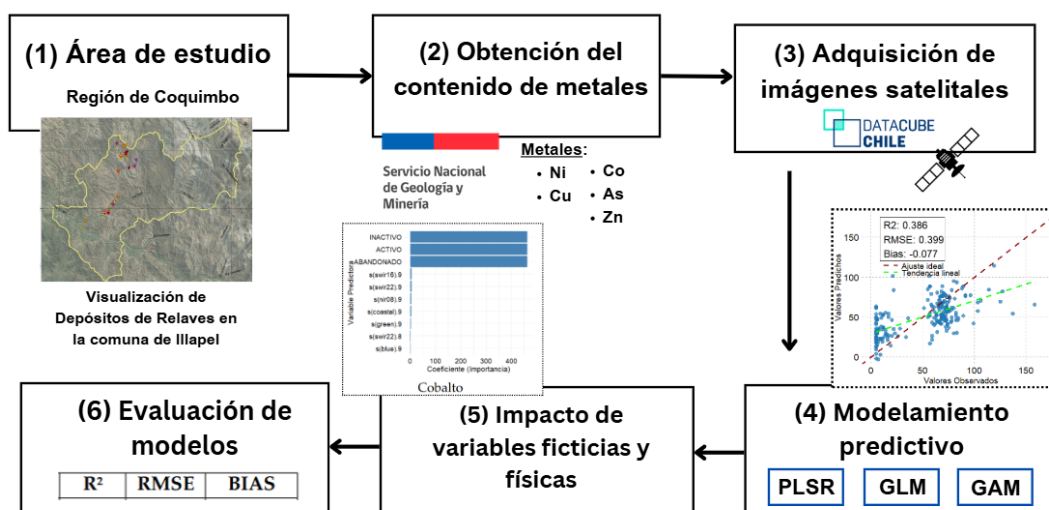


FIGURA 4.1: Flujo de trabajo. Se presenta la metodología dividida en 6 pasos: (1) área de estudio; (2) Obtención del contenido de metales; (3) adquisición de imágenes satelitales; (4) modelamiento predictivo; (5) impacto de variables ficticias y físicas; (6) evaluación de modelos.

4.1. Área de estudio

El área de estudio abarca depósitos de relaves ubicados en la región de Coquimbo, Chile. El área seleccionada incluye 14 de las 15 comunas, extendiéndose de norte a sur. La mayor concentración de depósitos se encuentra en las comunas de Andacollo, Illapel, La Serena, Punitaqui, Vicuña y La Higuera, que representan el 78,3 % de las muestras analizadas. En la Figura 4.2 se visualiza la distribución por tipo de depósitos en la región de Coquimbo.

Se consideran depósitos de relaves generados por actividades mineras relacionadas con la extracción de oro, cobre o combinaciones de ambos recursos. La masa de estos depósitos varía considerablemente, desde 2.017 toneladas (correspondientes al depósito La Fortuna 1, propiedad de Miguel Aguirre Bustamante) hasta 416 millones de toneladas (depositadas en Hipógeno, de la empresa Teck). En promedio, los depósitos tienen una masa de 3,7 millones de toneladas, con una alta dispersión reflejada en una desviación estándar de 36,9 millones de toneladas. Además, los depósitos se encuentran en distintos estados operativos, clasificándose como inactivos (201 depósitos), activos (21 depósitos) y abandonados (6 depósitos).

La región de Coquimbo se caracteriza por una diversidad climática y topográfica, lo que influye directamente en las características de los depósitos. En el sector central, representado por la comuna de Andacollo, predomina un clima estepario (clasificación Köppen: BS) (Beck *et al.*, 2018), con una precipitación anual promedio de 186 mm y una temperatura media anual de 15,8 °C. La altitud promedio en esta comuna es de 840 m.s.n.m., con una dispersión reflejada en una desviación estándar de 725 m.s.n.m. En la comuna de La Higuera, situada al norte de la región, el clima es desértico frío (Köppen: BWk) (Beck *et al.*, 2018), con una precipitación anual promedio de 104 mm y una temperatura media anual de 16,2 °C. La altitud promedio en esta zona es de 629 m.s.n.m., alcanzando un máximo de 2.769 m.s.n.m. Por su parte, en el extremo sur de la región, representado por la comuna de Illapel, el clima es estepario frío (Köppen: BSk) (Beck *et al.*, 2018), con una precipitación anual de 286 mm y una temperatura media anual de 15,6 °C. La altitud promedio en esta comuna es de 976 m.s.n.m., con un valor máximo de 4.055 m.s.n.m.



FIGURA 4.2: Depósitos de relaves de la región de Coquimbo, Chile, utilizados en esta investigación. En total se consideran 21 depósitos en estado de Activo, 6 en estado de Abandonado y 201 en estado de Inactivo.

4.2. Obtención y selección de metales pesados

Las concentraciones de metales analizadas se obtuvieron del Reporte Geoquímico 2023 de Sernageomin (Sernageomin, 2023). Esta fuente ofrece una visión integral de la composición química de los depósitos en la región de Coquimbo, constituyendo un recurso esencial para este estudio. Se seleccionan exclusivamente las muestras provenientes de cubetas, ya que estas reflejan con mayor precisión las concentraciones químicas predominantes en la superficie de los relaves. Esto resulta relevante, dado que las imágenes satelitales capturan principalmente información superficial de los depósitos (Lee *et al.*, 2016). Además, para garantizar la consistencia de las observaciones, se toma una sola muestra de cubeta por depósito, considerando que los casos con múltiples muestras disponibles son minoritarios y presentan valores intra-depósito similares. Todas las concentraciones se expresan en gramos por tonelada (g/t).

En este trabajo, los metales que se analizan, con excepción del Co, se consideran elementos CCA debido a su toxicidad para el ser humano y el medio ambiente, su capacidad de bioacumulación y su transporte a largas distancias. Estos elementos se seleccionan por su relevancia ambiental y porque han sido objeto de análisis en estudios previos, lo que garantiza su factibilidad de estimación y permite realizar comparaciones con enfoques actuales. A continuación, se detallan los metales analizados:

1. **Cobre (Cu):** Destaca por ser el más abundante de los 56 elementos químicos reportados. En cuanto a su impacto ambiental, su exceso puede provocar clorosis, necrosis y reducción del crecimiento vegetal, efectos que se reflejan en indicadores espectrales (Canadian Council of Ministers of the Environment, 1999). Además, Chile es el principal productor de cobre del mundo, por lo que se destaca la oportunidad de reprocesamiento de este metal (ConsejoMinero, 2024).
2. **Níquel (Ni):** Aunque es esencial en pequeñas cantidades para las plantas, en concentraciones elevadas genera toxicidad que afecta procesos fisiológicos clave, como la fotosíntesis, efectos detectables mediante cambios en la reflectancia espectral (Canadian Council of Ministers of the Environment, 2015). Por otro lado, la demanda por este metal aumentará a medida que se acelere la tendencia hacia los vehículos eléctricos, dado que es un componente clave de sus baterías (BHP, 2024).
3. **Zinc (Zn):** Posee propiedades tóxicas para la biota terrestre y acuática. Además, en concentraciones elevadas en suelos agrícolas pueden alterar la calidad del agua afectando a la salud humana y ambiental (Canadian Council of Ministers of the Environment, 2018). Por otro lado, es uno de los metales pilares y predominantes en las faenas mineras en Chile (Sernageomin, 2024b).
4. **Cobalto (Co):** Este metal no posee un impacto ambiental característico, sin embargo, posee una relevancia económica e industrial que lo posicionan como un “Elemento Estratégico” por la Unión Europea (JRI Ingeniería S.A., 2020).
5. **Arsénico (As):** Aunque el As no posee una relevancia económica comparable a otros metales, se considera altamente tóxico, persistente, móvil y bioacumulable, representando una amenaza ambiental y de salud pública significativa (Canadian Council of Ministers of the Environment, 2001).

La selección de estos metales no solo responde a su abundancia en los depósitos de relaves e impacto ambiental, sino también a su relevancia en el contexto actual de desarrollo tecnológico y sostenibilidad. Según el Banco Mundial y Church y Crawford, al menos 24 minerales críticos, entre ellos Cu, Ni, Co, Zn y elementos de tierras raras, se emplean en tecnologías de energía renovable, almacenamiento de energía e instalaciones eléctricas (World Bank Group, 2017; Hafner y Tagliapietra, 2020). Esto subraya la importancia de estudiar estos elementos en el marco de la transición hacia economías bajas en carbono.

Un aspecto crucial en el manejo de los datos consiste en abordar las concentraciones por debajo del umbral mínimo detectable por los instrumentos de medición *in situ*. En estos casos, se adopta un enfoque conservador, asignando el valor mínimo medible del equipo a dichas muestras. Por ejemplo, cuando el reporte indica una concentración " < 5 ", se asigna un valor de 5 g/t. Este procedimiento garantiza la consistencia en los datos sin excluir información relevante.

4.3. Adquisición y preprocesamiento de imágenes satelitales

La adquisición de datos remotos se lleva a cabo mediante el satélite Landsat 8, lanzado en 2013 por la Administración Nacional de Aeronáutica y el Espacio (NASA) y el Servicio Geológico de los Estados Unidos. Este satélite está diseñado para monitorear y gestionar recursos naturales, capturando imágenes multispectrales y térmicas a través de sus sensores OLI y TIRS. En esta investigación, se utiliza el sensor OLI, que abarca 9 bandas espectrales con longitudes de onda entre 0,43 y 2,29 micrómetros, una resolución espacial de 30 metros y una cobertura de área de 185 x 185 km por imagen (T.D. Acharya, 2015).

Las imágenes satelitales se obtienen a través de la plataforma Data Cube Chile (<https://datacubechile.cl/>), que facilita la descarga, procesamiento y análisis de datos de Landsat 8. La plataforma aplica un preprocesamiento inicial que incluye correcciones geométricas y radiométricas, asegurando la calidad de los datos. Para este estudio, se seleccionan imágenes de la colección C2L2, específicamente del producto *Surface Reflectance*, que proporciona valores de reflectancia corregidos atmosféricamente.

Para alinear temporalmente las imágenes con los datos muestreados en terreno, se solicita a Sernageomin las fechas de muestreo de los depósitos. Así, se seleccionan las imágenes más cercanas a cada muestreo, con registros entre las 14:38 y 14:39 horas, como se detalla a continuación:

- **Muestreos de terreno:** 3 y 15 de septiembre, 16 de octubre, 6 de noviembre, y 9 de diciembre de 2015; 1 de febrero, 1 de marzo y 9 de mayo de 2016; y 11 de diciembre de 2017.
- **Imágenes asociadas:** 2 de octubre, 19 de noviembre y 5 de diciembre de 2015; 22 de enero, 23 de febrero y 14 de junio de 2016; y 10 de diciembre de 2017.

Para extraer la reflectancia de los depósitos exclusivamente, se delimita el área superficial de cada depósito mediante interpretación visual en Google Earth Pro, utilizando las coordenadas proporcionadas por Sernageomin, y cuidando el detalle de la forma y tamaño. La delimitación se realiza con imágenes capturadas entre mayo de 2015 y diciembre de 2017, asegurando coherencia temporal con los datos muestreados (Sernageomin, 2023). Los polígonos generados, con áreas entre 281,4 m² y

2.821.566 m^2 , se exportan en formato GeoJSON y se emplean como máscaras en las imágenes satelitales correspondientes.

Una vez delimitados los polígonos, se cargan las bandas Blue, Green, Red, Nir08, Coastal, Swir16 y Swir22 en el sistema de referencia EPSG:4326. Además, se eliminan nubes, nieve y píxeles sin datos mediante la máscara de calidad "qa_pixel". Por cada banda, se calcula la reflectancia promedio aplicando las geometrías de los polígonos como máscaras sobre las imágenes. Los resultados de los 228 depósitos analizados se consolidan en un único dataframe con los valores promedio de reflectancia de las siete bandas seleccionadas. Este procesamiento se realiza utilizando las librerías Datasets y Rasterio en Python 3.10.12 (Gillies, 2018). Adicionalmente, se incluyen seis índices espectrales para aumentar la sensibilidad del modelo a patrones relevantes. Los índices utilizados son:

- **NDVI** (Índice de Vegetación de Diferencia Normalizada):

$$NDVI = \frac{Nir08 - Red}{Nir08 + Red}$$

- **DVI** (Índice de Vegetación por Diferencia):

$$DVI = Nir08 - Red$$

- **RVI** (Índice de Vegetación por Razón):

$$RVI = \frac{Nir08}{Red}$$

- **SBI** (Índice de Suelo Brillante)

$$SBI = \sqrt{Nir08^2 + Red^2}$$

- **EVI** (Índice de Vegetación Mejorado):

$$EVI = 2,5 \times \frac{Nir08 - Red}{Nir08 + 6 \times Red - 7,2 \times Blue + 1}$$

- **NDWI** (Índice de Agua de Diferencia Normalizada):

$$NDWI = \frac{Green - Nir08}{Green + Nir08}$$

4.4. Modelamiento predictivo

El modelamiento predictivo se realiza utilizando modelos de regresión basados en Partial Least-Squares Regression (PLSR) y Generalized Linear Models (GLM) para estimar las concentraciones de Ni, Cu, Zn, Co y As. El modelo PLSR se emplea ampliamente en teledetección de metales pesados debido a su capacidad para manejar la multicolinealidad y variables altamente correlacionadas (Lovynska *et al.*, 2024).

Este enfoque reduce la dimensionalidad de los datos mediante componentes latentes que maximizan la covarianza entre las variables predictoras y la variable de respuesta. Estas componentes capturan la mayor varianza relevante de las variables espectrales, optimizando la capacidad predictiva del modelo. En estudios de relaves, donde el número de variables espectrales supera al de muestras disponibles, PLSR ofrece modelos robustos y confiables que correlacionan la reflectancia espectral con concentraciones químicas, incluso en condiciones de ruido o datos incompletos (Wold *et al.*, 2001).

Por su parte, el modelo GLM se utiliza para analizar relaciones entre variables cuando los residuos de los modelos no siguen una distribución normal. Este enfoque se adapta a distribuciones de la familia exponencial, como Binomial, Poisson y Gamma, mediante un predictor lineal y una función de enlace. El predictor lineal combina las variables independientes de forma aditiva, mientras que la función de enlace ajusta el valor esperado de la respuesta a su distribución específica. Estas características hacen que GLM sea adecuado para manejar datos heterogéneos y sesgados, como las concentraciones metálicas en depósitos de relaves. Su capacidad para enfrentar colinealidad y variabilidad no constante lo convierte en una herramienta eficaz para mejorar la precisión de las estimaciones (Olsson, 2006).

La inclusión de ambos modelos responde al objetivo de comparar modelos de regresión en el contexto de los datos analizados. El modelo GLM ofrece flexibilidad en la distribución de la variable de respuesta, lo que resulta relevante considerando su desempeño demostrado en estudios previos. Por ejemplo, en un análisis sobre la riqueza de especies de plantas vasculares, GLM con distribución Binomial negativa supera a Random Forest debido a su capacidad para manejar datos de conteo con heterocedasticidad (Lopatin *et al.*, 2016). Aunque este estudio no utiliza datos de conteo, GLM se considera adecuado para explorar su capacidad en datos con características similares.

La selección de variables se realiza generando múltiples combinaciones de variables predictoras basadas en bandas e índices espectrales para los modelos PLSR y GLM. En la primera etapa, se incluyen variables con correlaciones de Pearson superiores a 0.2. Si todas las correlaciones son menores, se priorizan aquellas con valores iguales o superiores a 0.1. Para los metales con correlaciones inferiores a 0.1, se consideran todas las variables debido a su baja asociación inicial. Estos criterios se definen dadas las correlaciones presentadas en este estudio. Posteriormente, se añaden iterativamente otras bandas e índices espectrales para complementar las variables seleccionadas en la etapa inicial. Las combinaciones se generan dinámicamente mediante la función *combn* del lenguaje R, lo que permite explorar configuraciones simples y complejas (Carey, 2024). Además, se realiza un análisis de significancia para cada correlación con el fin de evaluar su relevancia estadística, lo que permite seleccionar únicamente aquellas variables que contribuyen significativamente a la precisión de los modelos.

Una vez generadas las combinaciones de variables para cada metal, se aplican tres experimentos. Cada experimento se realiza tanto para PLSR como para GLM, generando un total de 16 modelos por cada metal.

1. El primer experimento evalúa todas las combinaciones posibles de variables, seleccionando la mejor para cada metal según el valor de R^2 . En este proceso, cada modelo (PLSR y GLM) identifica combinaciones específicas de variables que maximizan su desempeño. Esto permite a cada modelo aprovechar sus

fortalezas al utilizar las variables más adecuadas para la estimación de cada metal.

2. En el segundo experimento, se emplea la mejor combinación de variables obtenida en el primer experimento para evaluar el impacto de las transformaciones logarítmica y raíz cuadrada en la variable de respuesta. Estas transformaciones, aplicadas antes del ajuste del modelo, buscan manejar relaciones no lineales y estabilizar varianzas sesgadas. Las predicciones se devuelven a su escala original para garantizar la comparabilidad entre modelos.
3. En el tercer experimento, se utiliza nuevamente la mejor combinación de variables seleccionada en el primer experimento. En esta etapa, se evalúan distintas transformaciones aplicadas a las variables independientes con el objetivo de mejorar la relación entre las variables predictoras y la respuesta. A continuación, se describen las transformaciones implementadas junto con sus respectivas justificaciones:
 - **Logarítmica y raíz cuadrada:** Reducir la influencia de valores extremos y estabilizar distribuciones sesgadas.
 - **Brightness Normalization:** Se identifica que esta normalización mejora la robustez del modelo PLSR frente a interferencias causadas por sombras (Feilhauer *et al.*, 2010).
 - **Standard Normalization:** Eliminar sesgos relacionados con diferentes escalas.
 - **Derivadas espectrales:** Se determina que estas técnicas son herramientas precisas para eliminar el ruido de fondo y resolver espectros superpuestos en teledetección (Demetriades-Shah *et al.*, 1990).

Estas transformaciones abordan problemas de escala, ruido y gradientes espectrales, optimizando las condiciones para el ajuste de los modelos. Además, se implementa un manejo de datos para tratar variables con valores negativos que dificultan la aplicación de ciertas transformaciones. Este enfoque garantiza la consistencia en el análisis y la comparabilidad de los resultados.

En los modelos GLM, se realiza un análisis de heterocedasticidad preliminar para seleccionar la familia de distribución y la función de enlace más adecuadas. Este análisis busca evaluar si los residuos de un modelo tienen una varianza no constante (Ling, 1984). Este proceso utiliza los criterios AIC/BIC y gráficos de residuos estandarizados. Los criterios AIC/BIC se emplean para comparar modelos equilibrando ajuste y complejidad, favoreciendo aquellos con valores más bajos (Penny, 2012). Por su parte, los gráficos de residuos estandarizados complementan el análisis al proporcionar una validación visual del comportamiento de los residuos y su concordancia con las suposiciones del modelo. Estudios similares aplican esta metodología para comparar modelos en la estimación de metales pesados (Yang *et al.*, 2021; Ling, 1984).

Adicionalmente, se complementan los modelos lineales PLSR y GLM con Generalized Additive Models (GAM), que permiten modelar relaciones no lineales mediante funciones suavizadas $s(x)$ (Hastie, 1992). Antes de implementar GAM, se evalúa la correlación de Spearman entre cada metal y las variables predictoras. Esta correlación, que captura relaciones no lineales y monótonas, se compara con la correlación de Pearson, la cual se basa únicamente en asociaciones lineales (Hauke y Kossowski,

2011). Las variables predictoras que presentan una correlación de Spearman mayor que la de Pearson se consideran candidatas para incluirse en el modelo GAM con una función suavizada $s(x)$. Este enfoque permite identificar relaciones complejas entre los metales y sus predictores, mejorando tanto la capacidad de modelado como la interpretación de los resultados.

4.5. Impacto de variables ficticias y físicas en la estimación de metales

Se analiza el impacto de incluir predictores nuevos basados en variables ficticias y físicas en la estimación de concentraciones de metales en depósitos de relaves. Este enfoque permite evaluar posibles relaciones entre el tamaño de los polígonos y las concentraciones de metales, además de capturar diferencias asociadas a los estados operativos de los depósitos, los cuales pueden ser fuentes de heterogeneidad. Este tipo de variables se utiliza en estudios que buscan capturar la influencia de factores categóricos. Por ejemplo, evaluar cómo diferentes categorías de uso y tipo de suelo afectan las concentraciones de metales (Wu *et al.*, 2020).

En este contexto, se incorporan las siguientes nuevas variables:

- **Categoría por Tamaño:** Representa la categorización del área de cada depósito por cuartiles (Q1, Q2, Q3 o Q4).
- **Estado Operativo:** Representa la categorización por estado operativo del depósito (Activo, Inactivo o Abandonado).
- **Área en m^2 :** Corresponde al área superficial de cada depósito representada como una variable continua.

A partir de la mejor combinación identificada para cada metal en el primer experimento, se generan combinaciones dinámicas que integran a las tres nuevas variables. Estas combinaciones se incluyen en los modelos PLSR y GLM, siguiendo la metodología utilizada previamente. Los modelos se ajustan considerando las familias de distribución identificadas previamente.

4.6. Evaluación de modelos

Los modelos generados se evalúan mediante validación cruzada k -fold con 10 particiones. Este método divide los datos en diez subconjuntos, utilizando uno para validación y los restantes para entrenamiento en cada iteración. Este enfoque permite estimar el desempeño predictivo de manera robusta, minimizando el riesgo de sobreajuste y optimizando el uso de datos, especialmente en conjuntos pequeños (Fushiki, 2011).

Los modelos se ajustan dinámicamente para cada combinación de variables, seleccionando el número óptimo de componentes latentes mediante el método *onesigma* (R. Wehrens y der Heijden, 2024). Este método identifica configuraciones cuyo error de predicción se encuentra dentro de una desviación estándar del mínimo error observado. De esta manera, se garantiza un balance entre la simplicidad del modelo y su capacidad predictiva.

El desempeño predictivo de los modelos se mide utilizando R^2 , RMSE estandarizado y Bias. El R^2 mide el porcentaje de varianza explicada por el modelo. El RMSE

evalúa el error promedio entre las predicciones y los valores observados, estandarizándose por la mediana para facilitar la comparación entre metales. El Bias calcula el sesgo sistemático en las predicciones, expresado como $1 - \beta_1$, donde β_1 es el coeficiente de una regresión entre predicciones y valores observados. Estas métricas se calculan en cada iteración de la validación cruzada y se promedian para obtener un desempeño global. La comparación de estas métricas permite determinar qué modelo, entre PLSR y GLM, ofrece mayor precisión en la estimación de concentraciones de metales pesados.

El desempeño de los modelos globales que incluyen variables categóricas se evalúa sin dividir previamente los datos por categoría. Este enfoque captura relaciones complejas entre las variables predictoras y la respuesta, maximizando el uso de la información disponible. Posteriormente, se analizan métricas como R^2 , RMSE y Bias dentro de cada categoría de la variable categórica. Esto permite identificar diferencias en el desempeño entre categorías, favoreciendo interpretaciones más precisas sin comprometer la coherencia del modelo global.

Después del cálculo de las métricas en cada experimento, se evalúa nuevamente la heterocedasticidad de cada metal mediante gráficos de residuos estandarizados. Además, en el caso de PLSR, se incorpora el gráfico Q-Q de residuos para evaluar la normalidad. Este gráfico compara la distribución de los residuos con una distribución normal teórica, identificando posibles desviaciones en las colas o en el rango central. Por este motivo, no se aplica en GLM, ya que, en este modelo, los residuos siguen la distribución de la familia asignada previamente, la que no necesariamente es Gaussiana. Estas evaluaciones complementan el análisis al verificar si, además de mejoras en las métricas, se logran mejoras en la heterocedasticidad y la normalidad de las variables de respuesta.

Capítulo 5

Resultados

En esta sección se presentan los resultados obtenidos de los modelos PLSR, GLM y GAM utilizados para estimar las concentraciones de Ni, Cu, Zn, Co y As a partir de imágenes del satélite Landsat 8. Se describen los resultados de tres experimentos principales: 1) la distribución y selección de variables; 2) los cambios en la estimación de cada metal al probar modelos lineales y no lineales; y 3) las variaciones en las estimaciones al incorporar las variables de estado operativo y tamaño de los depósitos de relaves.

5.1. Distribución de metales y pre-selección de variables

El análisis de histogramas evidencia patrones en las concentraciones de Cu, Ni, Zn, Co y As, como sesgos, asimetrías y bimodalidades (Figura 5.1). Estos patrones influyen en la selección de familias de distribución para los modelos GLM y GAM, destacando la alta variabilidad de las concentraciones. Este análisis justifica la aplicación de transformaciones y ajustes específicos. En la Figura 5.1 se presentan los histogramas utilizados para el análisis de heterocedasticidad y la elección de funciones de enlace, facilitando la selección de modelos adecuados.

El Ni muestra una distribución bimodal con picos en 5 g/t y 67 g/t, mientras que el Cu presenta un sesgo positivo con una desviación estándar alta. En el Zn se observa un sesgo extremo hacia valores mayores, con casos atípicos marcados. El Co exhibe un sesgo positivo con un pico en 5 g/t. Finalmente, el As presenta un sesgo positivo con un pico en 20 g/t. Estas características justifican evaluar las familias Gamma, Binomial Negativa y Gaussiana, adecuadas para datos con sesgo positivo y sobredispersión (Bliss C. I., 1953; Stacy, 1962).

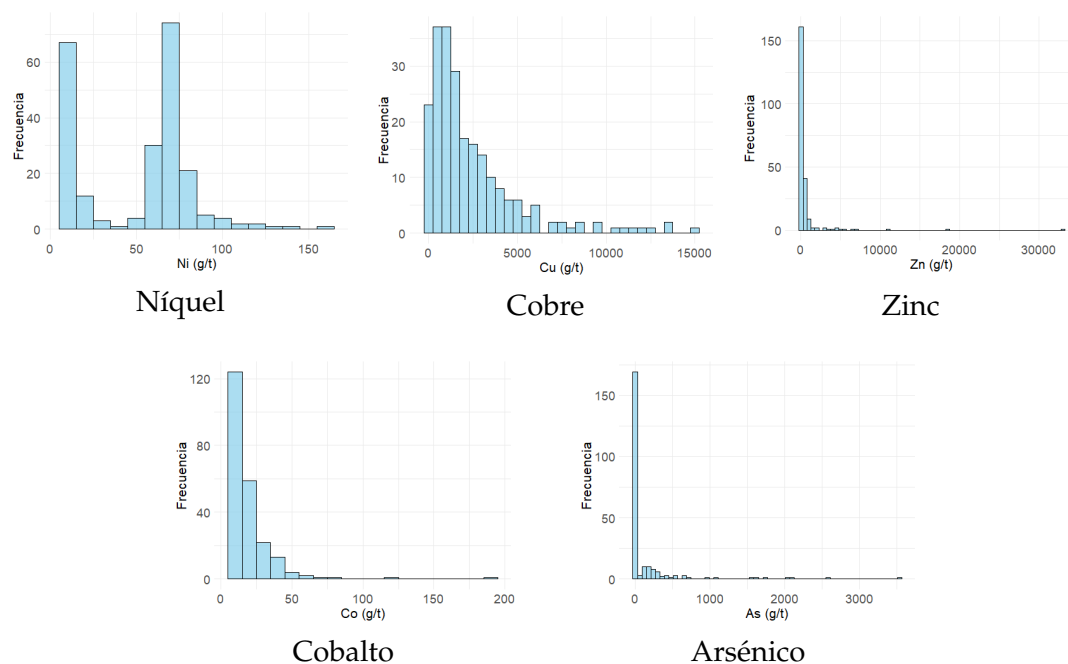


FIGURA 5.1: Histogramas de cada variable de respuesta.

En la Tabla 5.1 se presentan las estadísticas descriptivas de los metales, destacando elevados coeficientes de variabilidad. Por ejemplo, el Cu presenta un coeficiente de variabilidad del 271.92 %, mientras que el Zn alcanza el 421.86 %. Además, el Cu muestra las concentraciones más altas, seguido del Zn. Sin embargo, este último exhibe valores elevados debido a casos extremos, lo que evidencia un sesgo en su distribución.

	Ni	Cu	Zn	Co	As
Min (g/t)	5.00	33.0	5.00	5.00	10.00
1Q (g/t)	11.00	735.5	54.75	5.00	20.00
2Q (g/t)	65.00	1563.0	120.50	15.00	20.00
Media (g/t)	50.17	2446.3	684.87	18.49	150.38
3Q (g/t)	72.00	3200.8	291.25	23.00	84.96
Sd (g/t)	33.194	2710.126	2737.379	18.541	408.915
CV	66.16 %	110.78 %	421.86 %	100,27 %	271.92 %
Max (g/t)	158.00	15039.0	32834.00	190.00	3519.16
Skewness	-0.036	2.174	8.773	4.736	5.080

CUADRO 5.1: Estadísticas descriptivas de cada variable de respuesta.

En la Tabla 5.2 se presentan los valores de AIC y BIC para distintas familias de distribución, donde los modelos con menores valores de AIC y BIC representan un mejor equilibrio entre ajuste y complejidad (Penny, 2012). En el caso del Ni, las familias evaluadas muestran resultados similares. Para el Cu, la Binomial Negativa obtiene los mejores valores de AIC/BIC, aunque la familia Gamma resulta comparable. En el Zn, la Binomial Negativa presenta un ajuste superior al resto de las familias. Para el Co, los valores de AIC/BIC indican que la familia Gamma es ligeramente mejor que la Binomial Negativa. Finalmente, en el As, la Binomial Negativa muestra un ajuste levemente superior al de la familia Gamma.

	Métrica	Gaussiana (identity)	Gamma (log)	Gaussiana (log)	Negativa Binomial
Ni	AIC	2176.740	2164.680	2176.740	2164.989
	BIC	2207.604	2195.545	2207.604	2195.853
Cu	AIC	4258.290	4023.039	4258.290	4020.283
	BIC	4289.154	4053.903	4289.154	4051.148
Zn	AIC	4270.870	4023.039	4266.766	3228.428
	BIC	4301.734	4053.903	4297.630	3259.292
Co	AIC	1975.446	1726.438	1975.287	1735.073
	BIC	2006.310	1757.302	2006.151	1765.937
As	AIC	3389.993	2618.103	3321.065	2613.135
	BIC	3420.857	2648.967	3351.929	2643.999

CUADRO 5.2: Comparación de las distintas familias de distribución y funciones de enlace para cada variable de respuesta con valores AIC y BIC. Donde las celdas destacadas corresponden a los modelos seleccionados para cada variable de respuesta.

El análisis de gráficos de residuos estandarizados, mostrado en la Figura 5.2, complementa la evaluación de los modelos. Para el Ni, la familia seleccionada fue la Gaussiana con enlace *identity*, dado que presenta la menor tendencia sistemática, aunque persiste una curvatura. En el Cu, la Binomial Negativa fue seleccionada por manejar mejor la tendencia sistemática observada, además se observa una mayor dispersión de los residuos. En el Zn, los gráficos confirman que la Binomial Negativa ofrece el mejor ajuste, con una leve tendencia sistemática y una menor dispersión de valores bajos. Para el Co, tanto Gamma como Binomial Negativa manejan adecuadamente la homocedasticidad, pero se selecciona Gamma debido a su mejor desempeño en AIC/BIC. Finalmente, en el As, cada familia presenta una tendencia sistemática marcada, como se observa en el gráfico de residuos estandarizados del Arsénico, sin embargo la Binomial Negativa muestra levemente la menor tendencia sistemática en los residuos.

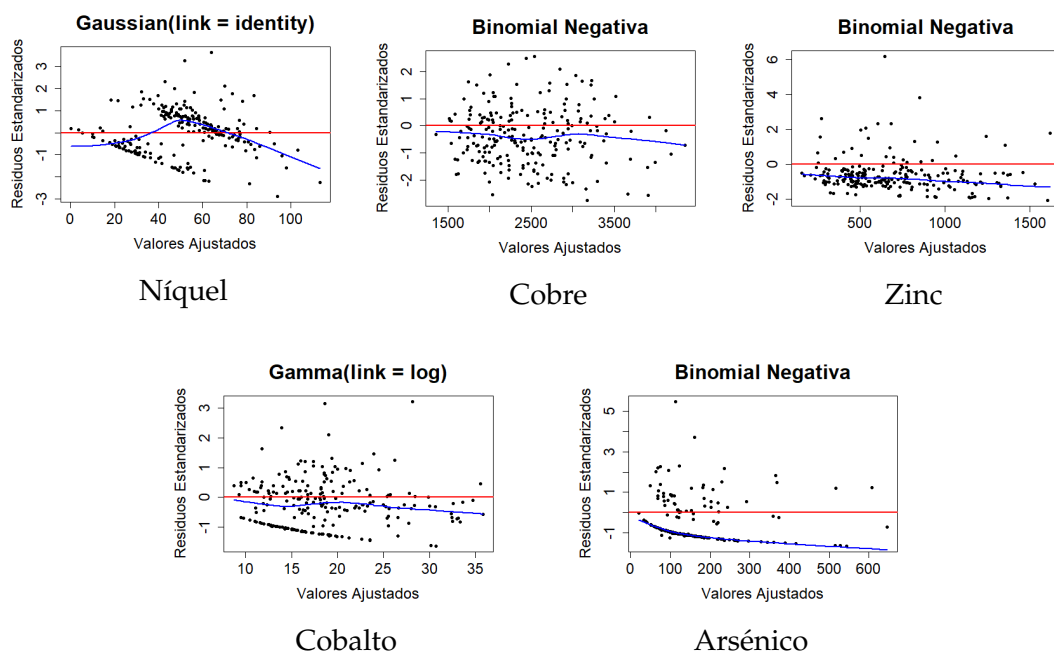


FIGURA 5.2: Gráfico de residuos estandarizados de las familias de distribución seleccionadas para cada variable de respuesta.

El análisis de correlación muestra que el Ni presenta correlaciones moderadas con bandas espectrales, con valores superiores a 0.18. En los demás metales, las correlaciones son menores a 0.2, aunque Nir08 y Swir16 muestran un mejor desempeño que las demás variables predictoras. En el Zn se observa una relación inversa con las bandas espectrales, e inferior a 0.1 en magnitud. La Figura 5.3 ilustra estas correlaciones, que guían la selección de variables predictoras.

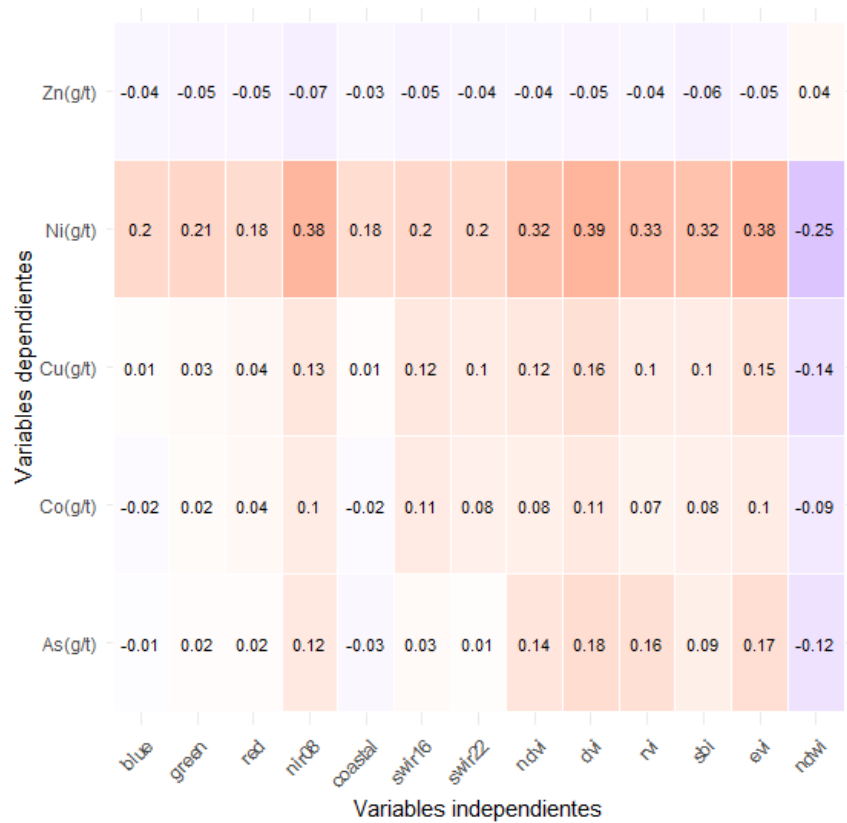


FIGURA 5.3: Correlación de Pearson entre las variables predictivas con cada banda e índice espectral.

El análisis de significancia estadística de las correlaciones entre las variables de respuesta y las bandas espectrales se detalla en la Tabla 5.3. En este análisis, el Ni destaca por presentar una correlación altamente significativa con la banda Nir08 ($p < 0,001$), mientras que las demás bandas muestran niveles de significancia moderados ($p < 0,01$). En contraste, las correlaciones entre las variables predictoras y las concentraciones de Cu, Zn, Co y As son bajas y no alcanzan un nivel de significancia estadística de $p < 0,05$. A pesar de esto, resulta necesario incluir variables predictoras en la construcción de los modelos para garantizar su funcionalidad. Para ello, se aplican PLSR y GLM, seleccionando las variables con mayor potencial predictivo relativo, independientemente de su correlación inicial o significancia estadística.

Por este motivo, es esencial interpretar con cautela los resultados obtenidos durante la selección de variables. Las relaciones identificadas entre las variables predictoras y de respuesta podrían no reflejar patrones subyacentes sólidos, lo que limita la certeza de las predicciones y las posiciona como aproximaciones tentativas en lugar de conclusiones definitivas. Este enfoque permite abordar las limitaciones en los datos disponibles, pero también resalta la importancia de llevar a cabo investigaciones adicionales para validar la robustez y eficacia de los modelos desarrollados.

Variable	Ni	Cu	Zn	Co	As
Blue	0.003	0.851	0.576	0.752	0.859
Green	0.001	0.651	0.468	0.722	0.813
Red	0.005	0.533	0.486	0.563	0.808
Nir08	$3,229 \times 10^{-9}$	0.005	0.311	0.137	0.008
Coastal	0.006	0.828	0.632	0.733	0.617
Swir16	0.002	0.067	0.463	0.107	0.655
Swir22	0.001	0.136	0.515	0.243	0.848

CUADRO 5.3: Grado de significancia de la correlación de Pearson entre cada variable de respuesta con las bandas espectrales.

5.2. Resultados para cada modelo

Para la selección de variables, se elaboran modelos predictivos para cada metal utilizando combinaciones entre las variables respuesta y los predictores seleccionados previamente. En el caso de Ni y Co, se evalúan 2,047 combinaciones de variables. Para Cu y Zn, al disponer de más variables base, las combinaciones evaluadas son menores, con un total de 1,023 y 63 configuraciones, respectivamente. En As, al contar con menos variables base, se pueden añadir más variables en las iteraciones, alcanzando un total de 4,095 combinaciones.

Estas configuraciones permiten identificar las variables más relevantes para cada modelo. En Ni, se destacan las bandas Green, Nir08 y Swir16, mientras que en Cu, las bandas Nir08 y Swir16, junto con el índice NDWI, resultan seleccionadas. En Zn, los índices SBI y NDWI se seleccionan en ambos modelos. Para Co, los índices espectrales son relevantes únicamente en PLSR, mientras que en As, los índices RVI y SBI resultan comunes en ambos enfoques.

De esta manera, las mejores combinaciones de variables para PLSR y GLM en cada metal se presentan en la Tabla 5.4.

Modelo	Metal	Combinación
PLSR	Ni	Green, Nir08, Swir16, NDVI y NDWI.
	Cu	Green, Blue, Red, Nir08, Coastal, Swir16, Swir22, NDVI, RVI, SBI y NDWI.
	Zn	Green, Blue, Red, Nir08, Coastal, Swir16, Swir22, SBI y NDWI.
	Co	Blue, Nir08, Coastal, Swir16, NDVI, DVI, EVI y NDWI.
	As	Red, Nir08, Coastal, Swir16, Swir22, RVI y SBI.
GLM	Ni	Green, Blue, Nir08, Coastal, Swir16, Swir22, RVI, SBI y EVI.
	Cu	Nir08, Swir16, Swir22, EVI y NDWI.
	Zn	Blue, Green, Red, Nir08, Coastal, Swir16, Swir22, RVI, SBI y NDWI.
	Co	Blue, Green, Nir08, Coastal, Swir16 y Swir22.
	As	Nir08, Swir16, NDVI, DVI, RVI y SBI.

CUADRO 5.4: Combinaciones de variables seleccionadas para cada variable de respuesta según PLSR y GLM.

En la Tabla 5.5 se presentan las métricas obtenidas para las mejores combinaciones de PLSR y GLM en cada variable de respuesta, mostrando el impacto de las transformaciones aplicadas a las variables predictoras, según los tres experimentos explicados en la Sección 4.4. En esta evaluación, y corresponde a la variable de respuesta y x a las combinaciones seleccionadas en la Tabla 5.4.

Para garantizar valores positivos durante las transformaciones de las variables predictoras, se suma “1” a todas las variables. En términos generales, en la Tabla 5.5 los modelos presentan valores bajos de R^2 en el conjunto de datos de testeo. Para el Ni, R^2 varía entre 0.047 y 0.386; en el caso de Cu, Zn y Co, los valores no superan 0.141; mientras que para el As, R^2 oscila entre 0.030 y 0.222. Estas métricas reflejan diferencias en el desempeño de los modelos según el metal analizado.

Para el Ni, GLM muestra mayor consistencia que PLSR, logrando un menor RMSE y un R^2 de hasta 0.386 sin transformaciones. Por su parte, PLSR alcanza su mejor R^2 (0.349) también sin transformaciones, aunque tiende a sobreestimar los valores. Estos resultados evidencian que GLM modela el Ni con mayor precisión.

En el Cu, GLM alcanza un R^2 ligeramente superior al de PLSR, mientras que este último presenta un menor RMSE. GLM subestima los valores, mientras que PLSR tiende a sobreestimarlos. El mejor modelo de GLM incluye una transformación logarítmica en la variable de respuesta, alcanzando un R^2 de 0.076, mientras que PLSR logra su mejor R^2 (0.069) sin transformaciones.

Para el Zn, PLSR se ajusta mejor que GLM tanto en R^2 como en RMSE. PLSR muestra su mejor desempeño con la Standard Normalization, alcanzando un R^2 de 0.109. Por su parte, GLM obtiene un R^2 de 0.075 utilizando transformaciones logarítmicas en las variables predictoras y la respuesta.

En el Co, GLM presenta mejores valores de R^2 , llegando a 0.141, en comparación con PLSR, que alcanza un R^2 de 0.106, aunque GLM muestra un mayor RMSE. Ambos

modelos tienden a sobreestimar los valores en magnitudes similares. Para ambos enfoques, la Brightness Normalization proporciona los mejores resultados, destacando su utilidad para este metal.

En el As, GLM demuestra un mejor desempeño en términos de R^2 , mientras que los valores de RMSE son similares entre ambos enfoques. PLSR alcanza un R^2 de 0.164 sin transformaciones, mientras que el mejor modelo de GLM utiliza una transformación logarítmica en la respuesta, logrando un R^2 de 0.222.

En resumen, GLM muestra un mejor desempeño para Ni, Cu, Co y As, mientras que PLSR resulta más preciso para Zn. Entre los metales, el Ni destaca por su R^2 más alto y su RMSE estandarizado más bajo, evidenciando un comportamiento más predecible. En contraste, el Cu presenta los valores de R^2 más bajos. El Zn y el Co registran R^2 similares, aunque el Zn exhibe un RMSE estandarizado mayor. Por último, el As logra el segundo R^2 más alto, pero con el RMSE estandarizado más elevado entre los metales. Estos resultados resaltan las fortalezas y limitaciones de cada modelo según las características específicas de los metales analizados.

Metal	Modelo	PLSR			GLM		
		R^2	RMSE	Bias	R^2	RMSE	Bias
Ni	$y \sim x$	0.349	0.416	0.199	0.386	0.399	-0.077
	$\log(y) \sim x$	0.265	0.543	0.295	0.047	$2,863 \times 10^{52}$	$5,844 \times 10^{51}$
	$\sqrt{y} \sim x$	0.319	0.444	0.259	0.326	3681.843	2924.84
	$y \sim \log(x)$	0.334	0.420	0.205	0.384	0.400	-0.075
	$y \sim \sqrt{x}$	0.346	0.416	0.201	0.385	0.399	-0.076
	$y \sim \text{brightness}(x)$	0.147	0.489	0.223	0.302	0.426	-0.319
	$y \sim \text{standard}(x)$	0.240	0.454	0.223	0.332	0.429	-0.018
	$y \sim \text{derivadas}(x)$	0.144	0.493	0.282	0.386	0.399	-0.077
Cu	$y \sim x$	0.069	1.701	0.517	0.076	1.844	-1062.217
	$\log(y) \sim x$	0.035	1.781	0.711	0.076	1.844	-1062.217
	$\sqrt{y} \sim x$	0.035	1.680	0.597	0.035	1.754	-532.4083
	$y \sim \log(x)$	0.035	1.675	0.494	0.070	1.843	-1064.549
	$y \sim \sqrt{x}$	0.037	1.677	0.495	0.070	1.843	-1064.606
	$y \sim \text{brightness}(x)$	0.049	1.682	0.500	0.053	1.837	-1074.209
	$y \sim \text{standard}(x)$	0.023	1.681	0.496	0.076	1.840	-1078.948
	$y \sim \text{derivadas}(x)$	0.060	1.678	0.492	0.076	1.843	-1062.217
Zn	$y \sim x$	0.097	15.79	0.487	0.057	16.690	-520.145
	$\log(y) \sim x$	0.081	15.17	0.882	0.057	16.690	-520.145
	$\sqrt{y} \sim x$	0.060	14.93	0.746	0.069	16.500	-375.658
	$y \sim \log(x)$	0.089	16.46	0.582	0.075	16.694	-521.227
	$y \sim \sqrt{x}$	0.097	16.43	0.573	0.073	16.693	-521.337
	$y \sim \text{brightness}(x)$	0.081	16.49	0.597	0.038	16.734	-520.599
	$y \sim \text{standard}(x)$	0.109	16.47	0.576	0.057	16.691	-520.145
	$y \sim \text{derivadas}(x)$	0.071	16.45	0.452	0.057	16.691	-520.145
Co	$y \sim x$	0.066	1.08	0.326	0.120	1.120	0.313
	$\log(y) \sim x$	0.060	1.08	0.482	0.050	$1,147 \times 10^{17}$	$2,370 \times 10^{16}$
	$\sqrt{y} \sim x$	0.065	1.06	0.423	0.111	447.771	369.582
	$y \sim \log(x)$	0.101	1.07	0.327	0.123	1.120	0.313
	$y \sim \sqrt{x}$	0.096	1.07	0.326	0.123	1.120	0.313
	$y \sim \text{brightness}(x)$	0.106	1.07	0.326	0.141	1.140	0.326
	$y \sim \text{standard}(x)$	0.090	1.06	0.326	0.120	1.120	0.097

Metal	Modelo	PLSR			GLM		
		R ²	RMSE	Bias	R ²	RMSE	Bias
	$y \sim \text{derivadas}(x)$	0.073	1.08	0.348	0.120	1.120	0.313
As	$y \sim x$	0.164	17.97	0.640	0.222	17.280	-105.260
	$\log(y) \sim x$	0.143	17.64	0.894	0.222	17.280	-105.260
	$\sqrt{y} \sim x$	0.133	17.28	0.818	0.212	16.560	-73.516
	$y \sim \log(x)$	0.049	17.35	0.663	0.216	20.332	-82.283
	$y \sim \sqrt{x}$	0.067	17.29	0.649	0.219	17.957	-97.264
	$y \sim \text{brightness}(x)$	0.075	17.13	0.651	0.201	19.356	-88.081
	$y \sim \text{standard}(x)$	0.030	17.64	0.703	0.222	17.277	-105.260
	$y \sim \text{derivadas}(x)$	0.104	17.32	0.662	0.222	17.277	-105.260

CUADRO 5.5: Resultados de R^2 , RMSE y Bias por cada modelo aplicado a cada variable de respuesta utilizando PLSR y GLM. Los valores corresponden a resultados en el conjunto de datos de testeo durante la validación 10-fold. Las celdas destacadas en azul representan a los mejores modelos de cada variable de respuesta utilizando PLSR, mientras que las celdas destacadas en verde, representan los mejores modelos para cada variable de respuesta utilizando GLM.

Los gráficos de residuos estandarizados de los mejores modelos para cada metal se presentan en la Figura 5.4. En el caso del Ni, los residuos muestran una curvatura marcada, lo que indica que PLSR no logra capturar completamente las relaciones entre las variables. Para el Cu, se identifica una tendencia sistemática leve, con menor dispersión en los valores residuales bajos. En el Zn, la dispersión de los residuos aumenta con valores ajustados mayores, acompañada de una tendencia sistemática más pronunciada. En el Co, los valores ajustados se concentran principalmente en rangos bajos, mostrando una tendencia sistemática notable. Por último, en el As, los valores ajustados se agrupan en rangos pequeños, excepto por un único valor ajustado cercano a 600, que genera una curvatura fuerte en los residuos. Estos patrones evidencian las dificultades para modelar algunos metales, incluso tras aplicar transformaciones y seleccionar combinaciones óptimas de variables.

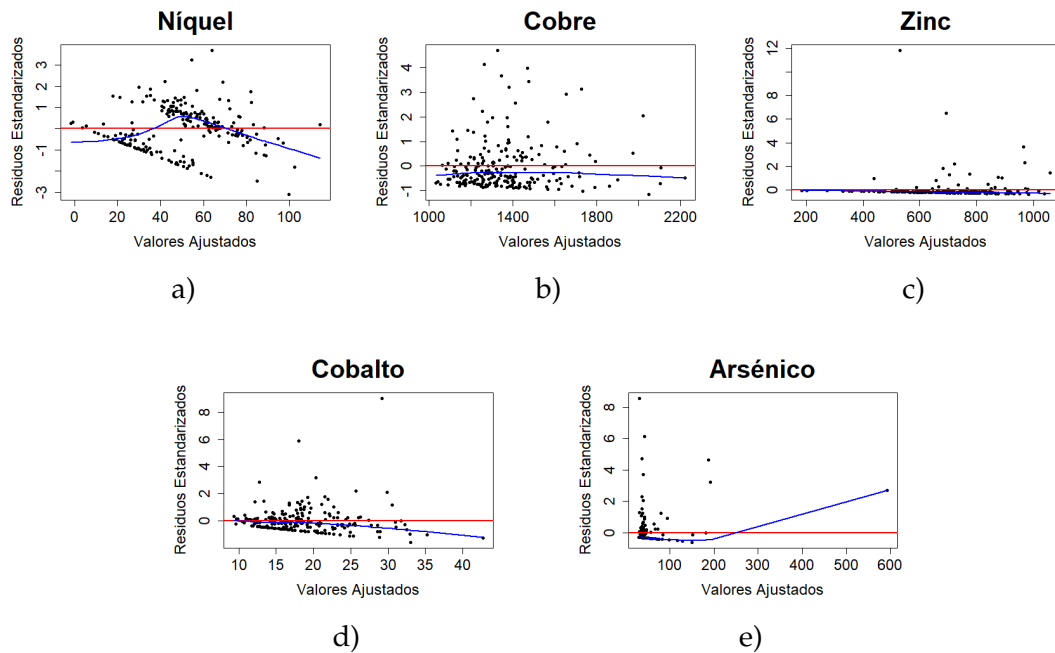


FIGURA 5.4: Gráficos de residuos estandarizados para los mejores modelos ajustados a cada variable de respuesta. a) Modelo $y \sim x$ de GLM para Ni; b) Modelo $\log(y) \sim x$ de GLM para Cu; c) Modelo $y \sim \text{standard}(x)$ de PLSR para Zn; d) Modelo $y \sim \text{brightness}(x)$ de GLM para Co; e) Modelo $\log(y) \sim x$ de GLM para As. Estos gráficos permiten evaluar visualmente el ajuste de cada modelo y detectar tendencias sistemáticas o patrones en los residuos.

Por lo tanto, las transformaciones aplicadas muestran un impacto limitado en el desempeño de los modelos. Solo en el Zn y el Co se evidencian mejoras en R^2 , aunque estas no son suficientes para reducir la heterocedasticidad de manera significativa. Esto sugiere que las transformaciones no capturan adecuadamente las relaciones subyacentes entre las variables predictoras y las respuestas.

Ante la persistencia de heterocedasticidad en los modelos, se exploran relaciones no lineales mediante el uso de GAM. Los resultados de correlaciones de Spearman, mostrados en la Tabla 5.6, evidencian asociaciones no lineales más intensas en ciertos metales. Estas asociaciones justifican el uso de funciones suavizadas $s(x)$ en GAM para modelar dichas relaciones utilizando los modelos seleccionados con GLM en la Tabla 5.5.

En particular, las correlaciones de Spearman destacan relaciones más fuertes en magnitud para Zn y Co en comparación con Pearson. En el As, las variables Swir16 y NDVI presentan mejores correlaciones con Spearman. Por otro lado, para el Ni y Cu, las correlaciones de Pearson son mayores que las de Spearman en cada relación. Estas variables se seleccionan para incluir funciones $s(x)$ en GAM, optimizando el ajuste del modelo.

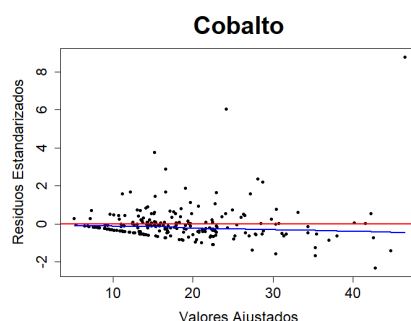
Variable	Ni	Cu	Zn	Co	As
Blue	0.21	0	-0.18	0.04	-0.18
Red	0.18	0.03	-0.22	0.1	-0.17
Green	0.22	0.02	-0.21	0.09	-0.17
Nir08	0.33	0.11	-0.15	0.24	-0.05
Coastal	0.19	0	-0.17	0.03	-0.17
Swir16	0.15	0.08	-0.25	0.18	-0.12
Swir22	0.15	0.07	-0.19	0.15	-0.11
NDVI	0.16	0.13	0.13	0.24	0.16
DVI	0.31	0.17	0.05	0.34	0.1
RVI	0.16	0.13	0.13	0.24	0.16
SBI	0.28	0.08	-0.18	0.19	-0.09
EVI	0.29	0.16	0.08	0.32	0.12
NDWI	-0.11	-0.13	-0.1	-0.22	-0.14

CUADRO 5.6: Correlación de Spearman entre cada variable de respuesta con sus variables predictoras. El color verde indica que la correlación de Spearman es mayor que la correlación de Pearson para dicha relación, mientras que el color amarillo y rojo indican que la correlaciones son similares e inferiores, respectivamente. Las celdas en color blanco corresponden a las variables que no son seleccionadas como variables predictoras para dicha variable de respuesta.

Los resultados de GAM muestran que solo el Co mejora su R^2 y reduce su RMSE en comparación con GLM (Figura 5.5a). No obstante, el gráfico de residuos presentado en 5.5b de la Figura 5.5 evidencia una mayor dispersión en valores ajustados altos para el Co, junto con una menor tendencia sistemática. Estos hallazgos destacan el potencial de GAM para abordar relaciones no lineales y desafíos en la modelación de datos sesgados.

	Modelo	R^2	RMSE	BIAS
Zn	$y \sim \text{standard}(x)$	0.070	15.507	-518.217
Co	$y \sim \text{brightness}(x)$	0.214	1.050	-0.258
As	$\log(y) \sim x$	0.122	16.410	-105.879

(A) Métricas del modelo GAM para Zn, Co y As.



(B) Gráfico de residuos estandarizados para el Co.

FIGURA 5.5: Resultados del modelo GAM. a) Resultados de R^2 , RMSE y Bias del modelo GAM para Zn, Co y As, en el conjunto de datos de testeo. b) Gráfico de residuos estandarizados para Co usando GAM.

Finalmente, los gráficos de resultados observados frente a los predichos para los mejores modelos de cada metal, presentados en la Figura 5.6, se utilizan para evaluar visualmente el ajuste de los modelos. Estos gráficos muestran la proximidad de las predicciones a los valores observados, junto con los valores de R^2 , que oscilan entre 7.6 % y 38.9 %. En general, los mejores modelos para cada metal tienden a subestimar las concentraciones estimadas. Además, en el caso de níquel y arsénico, se identifican dificultades para predecir valores altos con precisión.

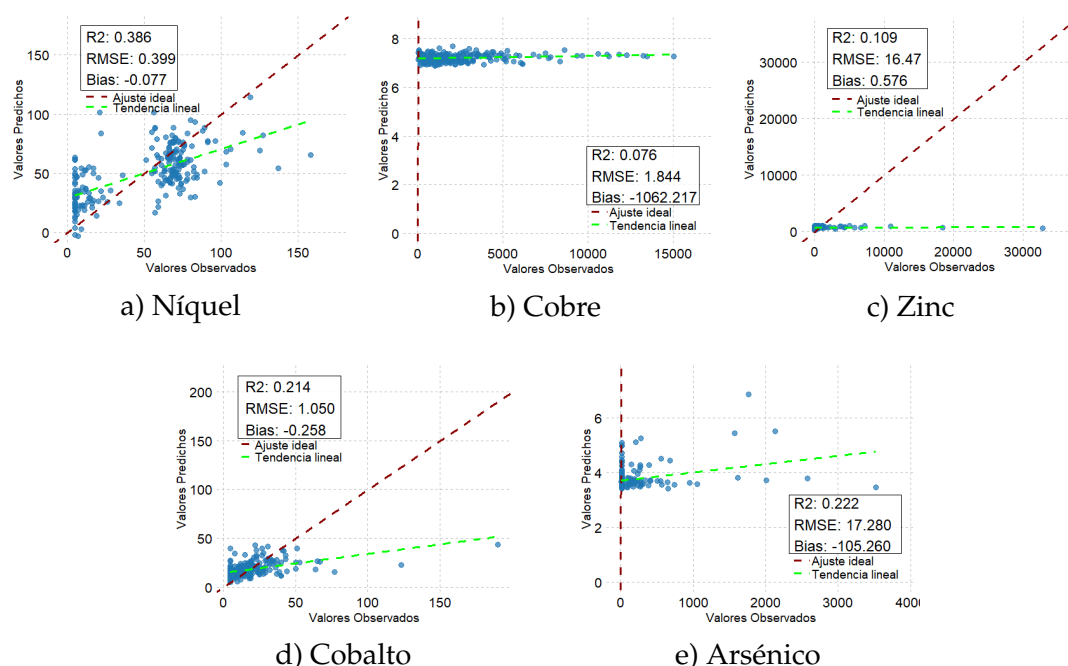


FIGURA 5.6: Gráficos de los valores reales versus valores predichos de cada variable de respuesta. a) Modelo $y \sim x$ de GLM para Ni; b) Modelo $\log(y) \sim x$ de GLM para Cu; c) Modelo $y \sim \text{standard}(x)$ de PLSR para Zn; d) Modelo $y \sim \text{brightness}(x)$ de GAM para Co; e) Modelo $\log(y) \sim x$ de GLM para As.

5.3. Influencia del tamaño y estado de los depósitos en la precisión

Los modelos actuales presentan bajos valores de R^2 y persistencia de heterocedasticidad. Para abordar estas limitaciones, se incorporan variables explicativas adicionales. Estas variables buscan segmentar los datos en subconjuntos homogéneos, capturar gradientes relacionados con el tamaño y mejorar la capacidad predictiva de los modelos, preservando la estabilidad de las estimaciones. La combinación de variables predictoras, descrita en la Tabla 5.4, se diversifica en tres configuraciones para cada metal. En la primera configuración, se incorpora la variable ficticia categórica por estado operativo. La segunda configuración añade las variables ficticias categóricas por tamaño. La tercera incluye el área en m^2 .

Los resultados, presentados en la Tabla 5.7, evidencian mejoras en dos casos específicos. Para el Ni, el modelo GLM incrementa su R^2 a 0.397 y reduce su RMSE al incorporar el área en m^2 . Para el Co, el modelo GAM mejora su R^2 a 0.218 al añadir las variables ficticias del estado operativo. Estos resultados destacan el potencial de

las nuevas variables en casos específicos, aunque sus beneficios son limitados en general.

Metal	Variable incorporada	PLSR			GLM		
		R ²	RMSE	Bias	R ²	RMSE	Bias
Ni	Estado Operativo	0.328	0.428	-0.670	0.387	0.397	0.133
	Categoría por Tamaño	0.222	0.459	-0.660	0.382	0.400	-0.186
	Área en m^2	0.327	0.437	0.279	0.397	0.397	0.087
Cu	Estado Operativo	0.037	1.708	-5.211	0.060	1.850	-1043.571
	Categoría por Tamaño	0.035	1.719	-12.645	0.052	1.841	-1050.088
	Área en m^2	0.048	1.695	0.473	0.076	1.844	-1061.305
Zn	Estado Operativo	0.097	16.873	0.635	0.045	40.393	795.069
	Categoría por Tamaño	0.102	16.847	1.101	0.023	18.544	122.296
	Área en m^2	0.098	16.873	0.784	0.056	24.384	135.446
Co*	Estado Operativo	0.016	1.085	-0.047	0.218	1.051	-0.295
	Categoría por Tamaño	0.041	1.084	26.655	0.193	1.058	-0.287
	Área en m^2	0.060	1.080	0.184	0.212	1.054	-0.150
As	Estado Operativo	0.069	17.941	-8.150	0.216	17.288	-104.882
	Categoría por Tamaño	0.031	18.086	-0.010	0.200	17.302	-104.970
	Área en m^2	0.105	17.727	-0.027	0.221	17.262	-105.206

CUADRO 5.7: Resultados de R^2 , RMSE y Bias por cada configuración de variables aplicado a cada variable de respuesta utilizando PLSR, GLM y GAM, en el conjunto de datos de testeó. * El Co se modela utilizando GAM en lugar de GLM. Las celdas destacadas en verde representan los modelos que muestran una mejora con respecto a los resultados descritos en la Sección 5.2.

El análisis de sensibilidad detalla la influencia de las variables predictoras en el mejor modelo de cada metal (Figura 5.7). Para el Ni, Nir08 presenta el mayor coeficiente positivo, seguido de Green y Coastal, mientras que Blue y SBI muestran los coeficientes más negativos. En el Cu, todos los coeficientes son positivos, destacando EVI como la variable más relevante, seguida de Swir16, mientras que Swir22, Nir08 y NDWI tienen impactos menores. Para el Zn, Green y SBI presentan los mayores impactos negativos, mientras que Red y Nir08 muestran coeficientes positivos relevantes. Las demás variables tienen impactos menores, con coeficientes de menor magnitud y menor relevancia en el modelo.

En el Co, las categorías “Inactivo”, “Activo” y “Abandonado” presentan impactos positivos similares, y las funciones suavizadas resaltan la importancia de Nir08, Swir16 y Swir22. Sin embargo, variables como Blue y Swir22 también exhiben relaciones negativas en ciertas partes de sus relaciones no lineales. Finalmente, en el As, Nir08 tiene un impacto negativo significativo, mientras que SBI presenta un efecto positivo moderado. Variables como DVI y NDVI muestran contribuciones positivas menores.

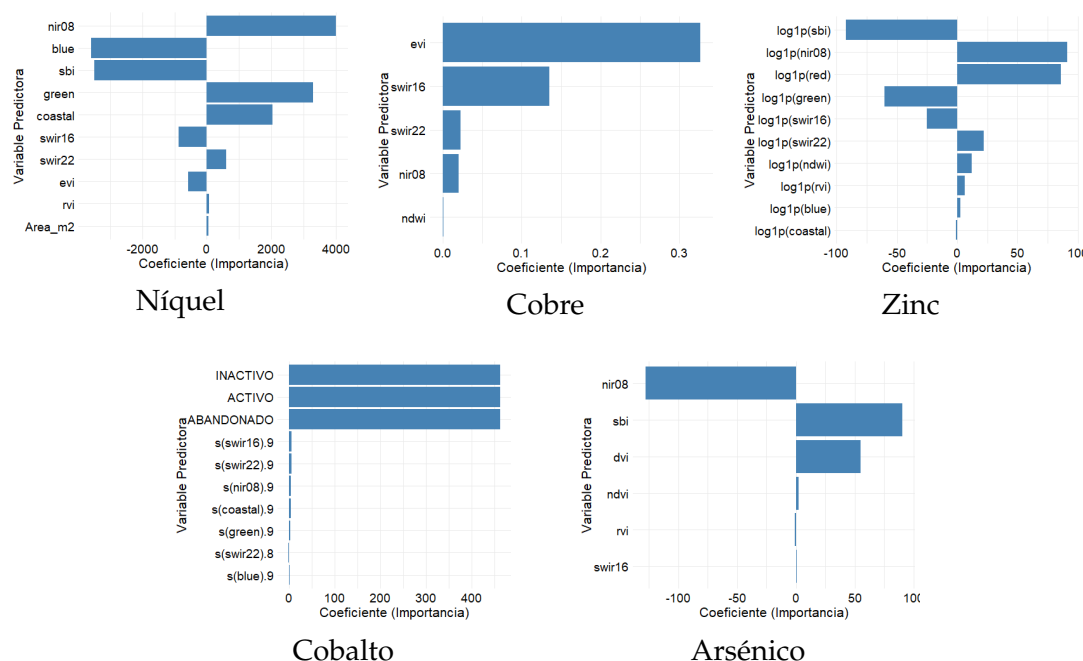


FIGURA 5.7: Importancia de las variable en los mejores modelos por metal. Los valores corresponden a los coeficientes Beta de cada variable predictora.

La incorporación de variables ficticias y del área en m^2 permite identificar patrones relacionados con el tamaño y el estado operativo de los depósitos. Sin embargo, los resultados muestran mejoras limitadas en las métricas y no logran reducir la heterocedasticidad de manera significativa. A pesar de ello, la variable ficticia del estado operativo evidencia una alta sensibilidad para el Co, destacando su relevancia en la estimación de este metal. Por este motivo, se analizan las métricas por categoría del estado operativo para evaluar diferencias en el desempeño del modelo.

En la Tabla 5.8, se presentan los resultados según el estado operativo del modelo. Para la categoría “Inactivo”, el modelo explica el 27,2 % de la variabilidad en los datos. En contraste, las categorías “Abandonado” y “Activo” explican el 12,1% y 8,3 %, respectivamente. Sin embargo, “Abandonado” y “Activo” cuentan con solo 6 y 21 muestras, lo que limita la capacidad de modelado en estos grupos.

	Estado Operativo	Muestras	R ²	RMSE	Bias
Co	Inactivo	201	0.272	1.115	-0.662
	Activo	21	0.083	0.763	1.688
	Abandonado	6	0.121	0.430	1.380

CUADRO 5.8: Resultados de R^2 , RMSE y Bias para el Co, en el conjunto de datos de testeo, dividido en cada una de sus categorías del estado operativo.

Capítulo 6

Discusión

En este capítulo se analizan los resultados descritos en el Capítulo 5, donde se utilizan PLSR, GLM y GAM para estimar las concentraciones de metales pesados a partir de imágenes satelitales de depósitos de relaves.

6.1. Distribución de los metales pesados en depósitos de relaves

La distribución de cada metal se analiza para identificar patrones en sus concentraciones. Según la Tabla 5.1, ningún metal presenta una distribución normal, y todos, excepto el Ni, superan un coeficiente de variabilidad del 100 %. En particular, los valores máximos del Cu y Zn se ubican en una escala de magnitud diez veces mayor que su tercer cuartil, con valores de 15.039 g/t y 32.834 g/t, respectivamente. Este patrón genera distribuciones fuertemente sesgadas hacia la derecha, especialmente para el Cu y Zn. En comparación, otros estudios reportan coeficientes de variabilidad inferiores al 50 % y distribuciones sin sesgos marcados (Guan *et al.*, 2019). Algunos trabajos adicionales indican coeficientes de variabilidad entre 24 % y 33 %, con distribuciones leptocúrticas cuyos valores extremos se mantienen dentro de la misma escala que el resto de los datos (Mouazen *et al.*, 2021).

En contraste, investigaciones previas reportan coeficientes de variabilidad más altos, entre 82 % y 90 %, pero con valores extremos dentro de una misma escala, a diferencia de los resultados presentados en esta tesis (Zhao *et al.*, 2021; Yin *et al.*, 2021). Estas diferencias pueden explicarse por las condiciones del área de estudio y las características del muestreo. En dichos trabajos, las muestras se obtienen de suelos mineros o urbanos, con áreas de estudio que abarcan sectores de 640 km² o depósitos de relaves activos de 2,188 km de longitud y 0,512 km de ancho. En contraste, esta investigación abarca la región de Coquimbo, con una superficie de 40.580 km², representando un contexto geográfico mucho más amplio. Este enfoque incluye depósitos con variabilidad significativa en tamaño y estado operativo, factores que influyen directamente en las concentraciones de metales. En particular, los depósitos activos, que concentran mayores actividades humanas, tienden a presentar concentraciones más altas debido a la acumulación reciente de residuos. Futuras investigaciones pueden explorar la variabilidad intra-depósito mediante un muestreo más denso dentro de los relaves. Además, dado que esta tesis considera una muestra por depósito, habiendo casos con más de una muestra disponible, se recomienda en estudios futuros utilizar el promedio de estas muestras para mejorar la representatividad de cada depósito.

6.2. Selección e importancia de variables

Las correlaciones de Pearson presentadas en la Figura 5.3 muestran que solo el Ni presenta una correlación moderada, destacando la banda Nir08 con un valor de 0.38. En contraste, el resto de los metales muestra correlaciones bajas, cercanas a 0.1. Estudios previos reportan correlaciones para Ni y Zn cercanas a 0.4 con bandas como Blue, Green y Red (Guan *et al.*, 2019). Otro estudio analiza las correlaciones del As, Cu y Zn con las bandas Blue, Green y Red, observando que solo el As alcanza valores entre 0.321 y 0.401, mientras que el Cu y Zn presentan valores menores a 0.015, a pesar de que estos últimos muestran mayores concentraciones (Zhao *et al.*, 2021). Adicionalmente, otro estudio presenta correlaciones similares para Ni, Cu y Zn, donde el Ni muestra valores entre 0.35 y 0.5 con bandas espectrales, mientras que el Cu y Zn no superan 0.2 (Rui *et al.*, 2021).

De este modo, estas investigaciones consideran únicamente metales con una correlación moderada o alta con las bandas espectrales. En esta tesis, la selección de variables incluye metales con baja correlación, aplicando PLSR y GLM. En comparación con otros trabajos, la Figura 5.7 muestra que las bandas Nir08 y Swir16, frecuentemente subestimadas, resultan clave para las estimaciones de metales en este contexto. Además, tanto en esta investigación como en otros estudios, se observa que la incorporación de índices espectrales mejora marginalmente la precisión en algunos casos (Zhao *et al.*, 2021), indicando que su aporte depende del entorno en que se ubica cada metal. De esta manera, las correlaciones observadas en este estudio reflejan características específicas del sitio, influenciadas por la variabilidad en los suelos y las reflectancias. Este resultado limita la aplicabilidad generalizada de los modelos y resalta la importancia de analizar condiciones locales. En este caso, al utilizar una imagen que representa todo el depósito de relaves, que contiene una amplia variedad de metales y minerales, se pierde precisión en la reflectancia, en contraste con tener una imagen específica del punto donde se encuentra el metal.

Se menciona que la profundidad exacta de las muestras tomadas por Sernageomin se desconoce. Aunque se informa que el muestreo se realiza en la cubeta de cada depósito, no se especifica la profundidad precisa. Este aspecto puede afectar los datos obtenidos por Landsat 8 si existe heterogeneidad en los muestreos realizados por depósito de relaves. Como referencia, otros estudios realizan muestreos con profundidades que oscilan entre 10 y 30 cm.

6.3. Análisis de la estimación de los modelos

La estimación de concentraciones de metales pesados muestra un desempeño bajo en general, aunque se identificaron aspectos relevantes. En la Tabla 5.5 se observa que PLSR y GLM presentan resultados similares, destacando que los mejores modelos lineales obtienen un R^2 de 0.386 para Ni, 0.076 para Cu, 0.109 para Zn, 0.141 para Co y 0.222 para As. Como comparación, un estudio previo reporta valores de R^2 de 0.566 para Ni, 0.582 para Cu y 0.570 para Zn utilizando PLSR, con correlaciones espectrales cercanas a 0.55 (Guan *et al.*, 2019). Esto indica que valores de correlación más altos incrementan la precisión en la estimación de metales.

Los errores más significativos reportados en (Guan *et al.*, 2019) ocurren en muestras con concentraciones extremas, evidenciando limitaciones predictivas en rangos altos y bajos, una observación consistente con los resultados de esta investigación. Un estudio adicional reporta mayores valores de RMSE para Cu, atribuidos a su alta

variabilidad en el área de estudio, un fenómeno también identificado en este análisis (Yang *et al.*, 2021). Estos hallazgos sugieren que las características extremas de los datos impactan directamente en la precisión de las estimaciones.

En adición, un estudio con distribuciones similares a las de esta tesis, pero sin valores extremos en escalas superiores a la media, reporta un R^2 de 0.4 para Ni, 0.2 para Cu, 0.1 para Co y 0.8 para As (Mouazen *et al.*, 2021). A excepción del As, los resultados de esta tesis son comparables, incluso al incluir valores extremos. Esto refuerza la utilidad de GLM para abordar distribuciones no normales y mejorar las estimaciones bajo estas condiciones.

Inclusive, en la Tabla 5.5 se identifican mejoras marginales al aplicar transformaciones a las variables predictoras. La Brightness Normalization incrementa el R^2 de Co de 0.120 a 0.141, mientras que la Standard Normalization aumenta el R^2 de Zn de 0.097 a 0.109. Estos resultados son consistentes con un estudio que utiliza reflectancia para estimar concentraciones de pigmentos en plantas (Feilhauer *et al.*, 2010). En otro análisis, una transformación logarítmica mejora el R^2 de PLSR para As de 0.431 a 0.460 (Zhao *et al.*, 2021). Estas observaciones refuerzan la utilidad de las transformaciones para optimizar la relación entre las variables predictoras y la respuesta, aunque no generan cambios sustanciales.

Otro estudio con correlaciones similares a las de esta tesis reporta que el metal con mayor correlación (Ni, $r=0.5$) alcanza un R^2 de 0.357, mientras que Cu y Zn logran valores de R^2 de 0.275 y 0.111, respectivamente, con correlaciones menores a 0.2 (Rui *et al.*, 2021). Estos resultados se asemejan a los de esta tesis, sin embargo, en el estudio mencionado, se opta por modelar los R^2 de Cu y Zn a partir del Ni para mejorar las estimaciones. Este enfoque no resulta viable en esta tesis debido al alto sesgo en los datos, lo que transferiría errores a otros modelos. Por esta razón, se utiliza GLM para modelar las distribuciones residuales. Este enfoque demuestra ser eficaz para manejar distribuciones sesgadas con valores extremos superiores a la media, logrando resultados comparables a los observados en otros estudios.

La aplicación del modelo no lineal GAM mejora el R^2 del Co, pasando de 0.141 con GLM a 0.214 (Figura 5.5a). Este resultado sugiere que las bandas e índices espectrales mantienen una relación no lineal con el Co. En estudios previos, modelos no lineales como Random Forest también reportan mejores resultados que PLS, aumentando el R^2 del Ni de 0.4 a 0.8 (Mouazen *et al.*, 2021). Otro análisis muestra que para As y Hg, PLSR alcanza R^2 de 0.536 y 0.268, respectivamente, mientras que el modelo Backpropagation Neural Network incrementa estos valores en 0.05 en promedio (Zhao *et al.*, 2021). De este modo, aunque en esta tesis GAM incrementa el R^2 del Co, este comportamiento no se observa en los demás metales. Esto sugiere que, a pesar de detectar indicios de no linealidad, el modelo GAM no mejora la estimación en otros metales, probablemente debido a la presencia de altos valores extremos que dificultan su ajuste.

Estos resultados destacan que los modelos no lineales pueden mejorar las estimaciones de metales pesados, aunque las mejoras suelen ser modestas. En entornos con baja multicolinealidad, como en imágenes multiespectrales, los modelos no lineales se utilizan como una herramienta efectiva para abordar relaciones no lineales y aumentar la precisión predictiva.

6.4. Análisis de las variables categóricas

La inclusión de variables ficticias categóricas muestra resultados limitados, con una ligera mejora del R^2 del Co de 0.214 a 0.218 al incorporar la variable de estado operativo (Tabla 5.7). Por otro lado, la incorporación de la variable continua de área en m^2 incrementa el R^2 del Ni de 0.386 a 0.397. Estos resultados sugieren que variables físicas, como el área, contribuyen marginalmente a mejorar las estimaciones de ciertos metales, mientras que su impacto en otros es insignificante.

Las estimaciones del Co varían al categorizar las muestras según el estado operativo (Tabla 5.8). Los depósitos inactivos presentan un R^2 de 0.272, superior al obtenido al considerar todos los depósitos. Sin embargo, las categorías restantes incluyen pocas muestras, lo que limita la confiabilidad de sus resultados. Un estudio similar divide las muestras en dos grupos (> 100 mg/kg de Cu) y reporta un incremento en el R^2 de 0.59 a 0.77 al usar PLSR con espectros en laboratorio (Yin *et al.*, 2021). Estos hallazgos sugieren que concentraciones homogéneas y sin valores extremos mejoran las estimaciones de R^2 .

En este análisis, agrupar los depósitos por estado operativo contribuye ligeramente a reducir la influencia de valores extremos, como los generados por depósitos activos en crecimiento o aquellos abandonados. Además, las categorías reflejan diferencias en el flujo fluvial, donde velocidades bajas favorecen la sedimentación de metales debido al efecto de la humedad sobre la reflectancia espectral. Por ejemplo, un estudio previo con Random Forest reporta una disminución en las concentraciones predichas a lo largo de un canal fluvial debido al aumento estacional del volumen de agua (Yin *et al.*, 2021).

Por lo tanto, factores estacionales podrían afectar las estimaciones en este estudio, considerando que las imágenes Landsat 8 y los muestreos se realizan en distintos meses, con diferencias de hasta 30 días. Estas variaciones climatológicas impactan negativamente las estimaciones al alterar las condiciones del terreno respecto al momento del muestreo.

Capítulo 7

Conclusiones

Esta tesis evaluó la capacidad de estimar concentraciones de metales pesados en distintos depósitos de relaves utilizando imágenes satelitales de Landsat 8. Para ello, se emplearon modelos de regresión lineal y no lineal, con el objetivo de identificar estrategias efectivas para manejar distribuciones complejas en las variables de respuesta. Los resultados obtenidos ofrecen nuevas perspectivas sobre el potencial de las imágenes multiespectrales para el monitoreo de estos depósitos.

Ante distribuciones altamente sesgadas en las concentraciones de metales, los modelos GLM demuestran mayor precisión en las estimaciones, independientemente de la magnitud de las concentraciones. Este enfoque resulta prometedor para depósitos abandonados con características homogéneas, donde la uniformidad reduce el sesgo y mejora la precisión. Sin embargo, no se valida la hipótesis planteada, ya que las concentraciones de metales no son factores determinantes en la mejora de las estimaciones. En cambio, las correlaciones y la presencia de valores extremos a mayor escala desempeñan un papel más significativo. Además, los resultados indican que la correlación entre bandas espectrales y metales pesados depende más del sector geográfico que de la magnitud de las concentraciones, condicionando la precisión según la ubicación.

Se comprueba que los modelos GLM y GAM superan a PLSR en contextos de distribuciones no normales, y que la inclusión de variables ficticias mejora la precisión en ciertos metales. Estos hallazgos resaltan la utilidad de metodologías adaptativas para estimar metales pesados mediante imágenes satelitales. A pesar de los bajos valores de R^2 obtenidos, los resultados son comparables a estudios previos, incluso considerando la mayor complejidad de los datos utilizados en esta tesis. Esto evidencia la validez de las metodologías aplicadas y subraya los desafíos asociados al modelado de datos altamente sesgados con imágenes de baja resolución espectral.

Como investigación futura, se recomienda incrementar el número de puntos de muestreo por depósito para reducir la cantidad de depósitos analizados y aumentar las muestras por unidad. Esta estrategia busca mejorar la normalidad en las distribuciones y la precisión en las estimaciones. Además, se sugiere ampliar el monitoreo hacia tierras raras, elementos de creciente interés tecnológico que son abundantes en los depósitos de relaves y representan una oportunidad significativa para futuras investigaciones.

Finalmente, se identifica que los datos de Landsat 8 presentan limitaciones debido a su baja resolución espectral, lo que afecta la precisión en la estimación de metales como el As, cuyos espectros de absorción no siempre coinciden con las bandas disponibles. El uso de imágenes hiperespectrales podría abordar esta limitación, aunque

actualmente no se dispone de una cobertura temporal adecuada para los períodos de muestreo realizados por Sernageomin. La integración de estas herramientas en el futuro permitiría mejorar la precisión en la estimación de concentraciones metálicas y optimizar el monitoreo remoto y eficiente de los depósitos de relaves.

Referencias

- Ada Contreras, E. O. V. G., Cristóbal Muñoz. (2022). *Monitoreo del estado de los relaves mineros en Chile*. (Disponible en: <https://www.cochilco.cl/web/sustentabilidad/>)
- A. Pica-Téllez, F. M. S. B. M. F. M. I. K. D. R. O. R. D. I. S., R. Garreaud. (2020). Informe Proyecto ARClím: Atlas de Riesgos Climáticos para Chile. *Centro de Ciencia del Clima y la Resiliencia, Centro de Cambio Global UC y MeteoData para el Ministerio del Medio Ambiente a través de La Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ)*.
- Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., y Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, 5(1), 180214. doi: 10.1038/sdata.2018.214
- BHP. (2024). *¿cuál es el futuro del níquel?* (Disponible en: <https://www.bhp.com/es/what-we-do/products/nickel/>)
- Bliss C. I., F. R. A. (1953). Fitting the negative binomial distribution to biological data. *Biometrics*, 9(2), 176–200. doi: 10.2307/3001850
- Canadian Council of Ministers of the Environment. (1999). *Canadian soil quality guidelines for the protection of environmental and human health: Copper* (1999). in: *Canadian environmental quality guidelines*, 1999. Winnipeg: Canadian Council of Ministers of the Environment.
- Canadian Council of Ministers of the Environment. (2001). *Canadian soil quality guidelines for the protection of environmental and human health: Arsenic (inorganic)* (1997). updated in: *Canadian environmental quality guidelines*, 1999. Winnipeg: Canadian Council of Ministers of the Environment.
- Canadian Council of Ministers of the Environment. (2015). *Canadian soil quality guidelines for the protection of environmental and human health: nickel* (2015). in: *Canadian environmental quality guidelines*, 1999. Winnipeg: Canadian Council of Ministers of the Environment.
- Canadian Council of Ministers of the Environment. (2018). *Canadian soil quality guidelines for the protection of environmental and human health: zinc* (2018). in: *Canadian environmental quality guidelines*, 1999. Winnipeg: Canadian Council of Ministers of the Environment.
- Carey, V. (2024). *combn: Generate all combinations of n elements, taken m at a time*. (Disponible en: <https://www.rdocumentation.org/packages/utils/versions/3.6.2/topics/combn/>)
- Chen, Q., Zhao, Z., Zhou, J., Zhu, R., Xia, J., Sun, T., ... Chao, J. (2022). ASTER and GF-5 Satellite Data for Mapping Hydrothermal Alteration Minerals in the Longtoushan Pb-Zn Deposit, SW China. *Remote Sensing*, 14(5), 1253. doi: 10.3390/rs14051253
- Choe, E., Van Der Meer, F., Van Ruitenbeek, F., Van Der Werff, H., De Smeth, B., y Kim, K.-W. (2008). Mapping of heavy metal pollution in stream sediments using combined geochemistry, field spectroscopy, and hyperspectral remote sensing: A case study of the Rodalquilar mining area, SE Spain. *Remote Sensing of Environment*, 112(7), 3222–3233. doi: 10.1016/j.rse.2008.03.017
- Chuvieco, E. (2017). Teledetección, s.i.g. y cambio global. *Geographicalia*, 29, 33–56. doi: 10.26754/ojs.geoph/geoph.1992291833
- ConsejoMinero. (2024). *Cifras actualizadas de la minería*. (Disponible en: <https://consejominero.cl/mineria-en-chile/cifras-actualizadas-de-la-mineria/>)

- Demetriades-Shah, T. H., Steven, M. D., y Clark, J. A. (1990). High resolution derivative spectra in remote sensing. *Remote Sensing of Environment*, 33(1), 55–64. doi: 10.1016/0034-4257(90)90055-Q
- División de Desarrollo Sostenible del Ministerio de Minería. (2019). *Plan nacional de depósitos de relaves para una minería sostenible*. (Disponible en: <https://www.minmineria.gob.cl/>)
- D’Emilio, M., Macchiato, M., Ragosta, M., y Simoniello, T. (2012). A method for the integration of satellite vegetation activities observations and magnetic susceptibility measurements for monitoring heavy metals in soil. *Journal of Hazardous Materials*, 241–242, 118–126. doi: 10.1016/j.jhazmat.2012.09.021
- Equipo Programa Tranque, F. C. (2018). *AVANCES Y RETOS PARA LA gestión DE LOS depósitos DE RELAVES EN CHILE*.
- Fang, Y., Xu, L., Peng, J., Wang, H., Wong, A., y Clausi, D. A. (2018). RETRIEVAL AND MAPPING OF HEAVY METAL CONCENTRATION IN SOIL USING TIME SERIES LANDSAT 8 IMAGERY. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-3, 335–340. doi: 10.5194/isprs-archives-XLII-3-335-2018
- Feilhauer, H., Asner, G. P., Martin, R. E., y Schmidtlein, S. (2010). Brightness-normalized Partial Least Squares Regression for hyperspectral data. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 111(12-13), 1947–1957. doi: 10.1016/j.jqsrt.2010.03.007
- FundaciónChile. (2024). *La ruta del programa*. (Disponible en: <https://fch.cl/iniciativa/programa-tranque/>)
- Fushiki, T. (2011). Estimation of prediction error by using k-fold cross-validation. *Stat Comput*, 21, 137–146. doi: 10.1007/s11222-009-9153-8
- Gan, W., Zhang, Y., Xu, J., Yang, R., Xiao, A., y Hu, X. (2023). Spatial Distribution of Soil Heavy Metal Concentrations in Road-Neighboring Areas Using UAV-Based Hyperspectral Remote Sensing and GIS Technology. *Sustainability*, 15(13), 10043. doi: 10.3390/su151310043
- Gillies, S. (2018). *rasterio documentation*.
- Gomez, C., Lagacherie, P., y Coulouma, G. (2012). Regional predictions of eight common soil properties and their spatial structures from hyperspectral Vis–NIR data. *Geoderma*, 189–190, 176–185. doi: 10.1016/j.geoderma.2012.05.023
- Guan, Q., Zhao, R., Wang, F., Pan, N., Yang, L., Song, N., ... Lin, J. (2019). Prediction of heavy metals in soils of an arid area based on multi-spectral data. *Journal of Environmental Management*, 243, 137–143. doi: 10.1016/j.jenvman.2019.04.109
- Hafner, M., y Tagliapietra, S. (Eds.). (2020). *The Geopolitics of the Global Energy Transition* (Vol. 73). Springer International Publishing. doi: 10.1007/978-3-030-39066-2
- Hastie, T. (1992). *Statistical models in s (1st ed.)*. Routledge. doi: 10.1201/9780203738535
- Hauke, J., y Kossowski, T. (2011). Comparison of Values of Pearson’s and Spearman’s Correlation Coefficients on the Same Sets of Data. *QUAGEO*, 30(2), 87–93. doi: 10.2478/v10117-011-0021-1
- Hernández Fernández, A. B. (2021). *Colapso del tranque de relaves Las Palmas durante el sismo del Maule 2010*. doi: 10.7764/tesisUC/ING/57461
- JRI Ingeniería S.A., E. L. (2020). *Reprocesamiento de relaves y recuperación de elementos de valor d288-man-gn-p2-001*.
- Khosravi, V., Gholizadeh, A., y Saberioon, M. (2022). Soil toxic elements determination using integration of Sentinel-2 and Landsat-8 images: Effect of fusion

- techniques on model performance. *Environmental Pollution*, 310, 119828. doi: 10.1016/j.envpol.2022.119828
- Kim, H., Yu, J., Wang, L., Jeong, Y., y Kim, J. (2020). Variations in Spectral Signals of Heavy Metal Contamination in Mine Soils Controlled by Mineral Assemblages. *Remote Sensing*, 12(20), 3273. doi: 10.3390/rs12203273
- Lassalle, G., Fabre, S., Credo, A., Hédacq, R., Dubucq, D., y Elger, A. (2021). Mapping leaf metal content over industrial brownfields using airborne hyperspectral imaging and optimized vegetation indices. *Scientific Reports*, 11(1), 2. doi: 10.1038/s41598-020-79439-z
- Lee, Z., Shang, S., Qi, L., Yan, J., y Lin, G. (2016). A semi-analytical scheme to estimate Secchi-disk depth from Landsat-8 measurements. *Remote Sensing of Environment*, 177, 101–106. doi: 10.1016/j.rse.2016.02.033
- Lin, X., Su, Y.-C., Shang, J., Sha, J., Li, X., Sun, Y.-Y., ... Jin, B. (2019). Geographically Weighted Regression Effects on Soil Zinc Content Hyperspectral Modeling by Applying the Fractional-Order Differential. *Remote Sensing*, 11(6), 636. doi: 10.3390/rs11060636
- Ling, R. F. (1984). Residuals and influence in regression. *Technometrics*, 26(4), 413–415. doi: 10.1080/00401706.1984.10487996
- Liu, M., Liu, X., Zhang, B., y Ding, C. (2016). Regional heavy metal pollution in crops by integrating physiological function variability with spatio-temporal stability using multi-temporal thermal remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 51, 91–102. doi: 10.1016/j.jag.2016.05.003
- Lopatin, J., Dolos, K., Hernández, H., Galleguillos, M., y Fassnacht, F. (2016). Comparing Generalized Linear Models and random forest to model vascular plant species richness using LiDAR data in a natural forest in central Chile. *Remote Sensing of Environment*, 173, 200–210. doi: 10.1016/j.rse.2015.11.029
- Lovynska, V., Bayat, B., Bol, R., Moradi, S., Rahmati, M., Raj, R., ... Montzka, C. (2024). Monitoring Heavy Metals and Metalloids in Soils and Vegetation by Remote Sensing: A Review. *Remote Sensing*, 16(17), 3221. doi: 10.3390/rs16173221
- Ministerio de Minería. (2021). *Evaluación ambiental estratégica de la política nacional minera 2050*.
- Mirzaei, F., Abbasi, Y., y Sohrabi, T. (2021). Modeling the distribution of heavy metals in lands irrigated by wastewater using satellite images of Sentinel-2. *The Egyptian Journal of Remote Sensing and Space Science*, 24(3), 537–546. doi: 10.1016/j.ejrs.2021.03.002
- Mirzaei, M., Verrelst, J., Marofi, S., Abbasi, M., y Azadi, H. (2019). Eco-Friendly Estimation of Heavy Metal Contents in Grapevine Foliage Using In-Field Hyperspectral Data and Multivariate Analysis. *Remote Sensing*, 11(23), 2731. doi: 10.3390/rs11232731
- Mouazen, A. M., Nyarko, F., Qaswar, M., Tóth, G., Gobin, A., y Moshou, D. (2021). Spatiotemporal Prediction and Mapping of Heavy Metals at Regional Scale Using Regression Methods and Landsat 7. *Remote Sensing*, 13(22), 4615. doi: 10.3390/rs13224615
- Olsson, U. (2006). *Generalized linear models: An applied approach*. Studentlitteratur : Btj [distributör] ; eLib [distributör].
- Peng, Y., Kheir, R., Adhikari, K., Malinowski, R., Greve, M., Knadel, M., y Greve, M. (2016). Digital Mapping of Toxic Metals in Qatari Soils Using Remote Sensing and Ancillary Data. *Remote Sensing*, 8(12), 1003. doi: 10.3390/rs8121003
- Penny, W. (2012). Comparing Dynamic Causal Models using AIC, BIC and Free Energy. *NeuroImage*, 59(1), 319–330. doi: 10.1016/j.neuroimage.2011.07.039

- Rigol, J. P. (2000). *Aplicación de sistemas de información geográfica y teledetección en exploración minera*.
- Rui, W., Shuang, W., Kan, W., Shiqiao, H., Ruijie, W., Bo, L., ... Xinpeng, D. (2021). Estimation and Spatial Analysis of Heavy Metals in Metal Tailing Pond Based on Improved PLS With Multiple Factors. *IEEE Access*, 9, 64880–64894. doi: 10.1109/ACCESS.2021.3073933
- R. Wehrens, H. V. d. V., y der Heijden, G. V. (2024). *selectncomp: Suggestions for the optimal number of components in pcr and pls models*. (Disponible en: <https://www.rdocumentation.org/packages/pls/versions/2.8-5/topics/selectNcomp>)
- Sacristán Romero, F. (2005). La teledetección satelital y los sistemas de protección ambiental. *Civilizar*, 5(9), 1. doi: 10.22518/16578953.701
- Sernageomin. (2023). *Geoquímica de superficie de depósitos de relaves de Chile*. (Disponible en: <https://www.sernageomin.cl/datos-publicos-deposito-de-relaves/>)
- Sernageomin. (2024a). *Depósitos de relaves*. (Disponible en: <https://www.sernageomin.cl/preguntas-frecuentes-sobre-relaves/>)
- Sernageomin. (2024b). *Recursos minerales*. (Disponible en: <https://www.sernageomin.cl/recursos-minerales/>)
- Sernageomin, Agencia de la Cooperación Internacional de Japón. (2007). *Guía metodológica para la evaluación de estudios y declaraciones de impacto ambiental*. (Disponible en: <https://repositorio.sernageomin.cl/items/85f47e88-4bb3-4f11-8614-1ffcfa67fe86>)
- Shi, S., Hou, M., Gu, Z., Jiang, C., Zhang, W., Hou, M., ... Xi, Z. (2022). Estimation of Heavy Metal Content in Soil Based on Machine Learning Models. *Land*, 11(7), 1037. doi: 10.3390/land11071037
- Shi, T., Hu, X., Guo, L., Su, F., Tu, W., Hu, Z., ... Wu, G. (2021). Digital mapping of zinc in urban topsoil using multisource geospatial data and random forest. *Science of The Total Environment*, 792, 148455. doi: 10.1016/j.scitotenv.2021.148455
- Silva Rotta, L. H., Alcântara, E., Park, E., Negri, R. G., Lin, Y. N., Bernardo, N., ... Souza Filho, C. R. (2020). The 2019 Brumadinho tailings dam collapse: Possible cause and impacts of the worst human and environmental disaster in Brazil. *International Journal of Applied Earth Observation and Geoinformation*, 90, 102119. doi: 10.1016/j.jag.2020.102119
- Song, W., Song, W., Gu, H., y Li, F. (2020). Progress in the Remote Sensing Monitoring of the Ecological Environment in Mining Areas. *International Journal of Environmental Research and Public Health*, 17(6), 1846. doi: 10.3390/ijerph17061846
- Stacy, E. W. (1962). A generalization of the gamma distribution. *The Annals of Mathematical Statistics*, 33(3), 1187–1192. doi: stable/2237889
- Sun, W., Liu, S., Wang, M., Zhang, X., Shang, K., y Liu, Q. (2023). Soil copper concentration map in mining area generated from AHSI remote sensing imagery. *Science of The Total Environment*, 860, 160511. doi: 10.1016/j.scitotenv.2022.160511
- Tan, K., Ma, W., Chen, L., Wang, H., Du, Q., Du, P., ... Li, H. (2021). Estimating the distribution trend of soil heavy metals in mining area from HyMap airborne hyperspectral imagery based on ensemble learning. *Journal of Hazardous Materials*, 401, 123288. doi: 10.1016/j.jhazmat.2020.123288
- Tan, K., Wang, H., Chen, L., Du, Q., Du, P., y Pan, C. (2020). Estimation of the spatial distribution of heavy metal in agricultural soils using airborne hyperspectral imaging and random forest. *Journal of Hazardous Materials*, 382, 120987. doi:

- 10.1016/j.jhazmat.2019.120987
- T.D. Acharya, I. Y. (2015). Exploring landsat 8. *International Journal of IT, Engineering and Applied Sciences Research (IJIEASR)*, 4(4).
- Transparent Tailings Initiative Team. (2019). *Physical stability monitoring and assessment standard for tailings storage facilities a proposal from chile*.
- Vela-Almeida, D., León, M., y Lewinsohn, J. L. (2021). *Indicadores de sostenibilidad en la minería metálica*.
- Wang, F., Gao, J., y Zha, Y. (2018). Hyperspectral sensing of heavy metals in soil and vegetation: Feasibility and challenges. *ISPRS Journal of Photogrammetry and Remote Sensing*, 136, 73–84. doi: 10.1016/j.isprsjprs.2017.12.003
- Wang, L., Zhou, Y., Liu, J., Liu, Y., Zuo, Q., y Li, Q. (2022). Exploring the potential of multispectral satellite images for estimating the contents of cadmium and lead in cropland: The effect of the dimidiate pixel model and random forest. *Journal of Cleaner Production*, 367, 132922. doi: 10.1016/j.jclepro.2022.132922
- Wang, Y., Zhang, X., Sun, W., Wang, J., Ding, S., y Liu, S. (2022). Effects of hyperspectral data with different spectral resolutions on the estimation of soil heavy metal content: From ground-based and airborne data to satellite-simulated data. *Science of The Total Environment*, 838, 156129. doi: 10.1016/j.scitotenv.2022.156129
- Wang, Y., Zou, B., Chai, L., Lin, Z., Feng, H., Tang, Y., ... Zou, H. (2024). Monitoring of soil heavy metals based on hyperspectral remote sensing: A review. *Earth-Science Reviews*, 254, 104814. doi: 10.1016/j.earscirev.2024.104814
- Wold, S., Sjöström, M., y Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2), 109–130. doi: 10.1016/S0169-7439(01)00155-1
- World Bank Group. (2017). *The Growing Role of Minerals and Metals for a Low Carbon Future*. World Bank, Washington, DC.
- Wu, Z., Chen, Y., Han, Y., Ke, T., y Liu, Y. (2020). Identifying the influencing factors controlling the spatial variation of heavy metals in suburban soil using spatial regression models. *Science of The Total Environment*, 717, 137-212. doi: 10.1016/j.scitotenv.2020.137212
- Xu, X., Wang, Z., Song, X., Zhan, W., y Yang, S. (2024). A remote sensing-based strategy for mapping potentially toxic elements of soils: Temporal-spatial-spectral covariates combined with random forest. *Environmental Research*, 240, 117570. doi: 10.1016/j.envres.2023.117570
- Yang, Y., Cui, Q., Jia, P., Liu, J., y Bai, H. (2021). Estimating the heavy metal concentrations in topsoil in the Daxigou mining area, China, using multispectral satellite imagery. *Scientific Reports*, 11(1), 11718. doi: 10.1038/s41598-021-91103-8
- Yin, F., Wu, M., Liu, L., Zhu, Y., Feng, J., Yin, D., ... Yin, C. (2021). Predicting the abundance of copper in soil using reflectance spectroscopy and GF5 hyperspectral imagery. *International Journal of Applied Earth Observation and Geoinformation*, 102, 102420. doi: 10.1016/j.jag.2021.102420
- Z.E. Mashimbye, J. N. W. D. C. A. V. N. D. T., M.A. Cho. (2012). Model-Based Integrated Methods for Quantitative Estimation of Soil Salinity from Hyperspectral Remote Sensing Data: A Case Study of Selected South African Soils. *Pedosphere*, 22(5), 640–649. doi: 10.1016/S1002-0160(12)60049-6
- Zeng, L., Wang, Y., Jing, L., y Cheng, Q. (2021). Quantitative determination of auxiliary information for mapping soil heavy metals and soil contamination risk assessment. *Applied Geochemistry*, 130, 104964. doi: 10.1016/j.apgeochem.2021.104964

- Zerené, G. S., y Muñoz, C. I. (2017). *La política de aguas en Chile y la gran minería del cobre*.
- Zhang, B., Liu, X., Liu, M., y Meng, Y. (2018). Detection of Rice Phenological Variations under Heavy Metal Stress by Means of Blended Landsat and MODIS Image Time Series. *Remote Sensing*, 11(1), 13. doi: 10.3390/rs11010013
- Zhao, H., Liu, P., Qiao, B., y Wu, K. (2021). The Spatial Distribution and Prediction of Soil Heavy Metals Based on Measured Samples and Multi-Spectral Images in Tai Lake of China. *Land*, 10(11), 1227. doi: 10.3390/land10111227
- Zhou, G., Liu, X., Zhao, S., Liu, M., y Wu, L. (2017). Estimating FAPAR of Rice Growth Period Using Radiation Transfer Model Coupled with the WOFOST Model for Analyzing Heavy Metal Stress. *Remote Sensing*, 9(5), 424. doi: 10.3390/rs9050424