

Prospección Geológica Mediante la Aplicación de Técnicas de Machine Learning,  
Geoestadística y Procesamiento de Imágenes Satelitales en la Franja Metalogenética XV  
(Perú)

Miguel Ángel Ávila Marín, Lucas Quintero Marín

Trabajo de Grado para Optar el título de Geólogo

Director

Julián David De Bedout Ordóñez

Magíster en Ingeniería Geológica

Codirectores

Julián David Rojas Delgado

Geólogo

José Pedro Mora Ortiz

Magíster en Geofísica y Sistemas de Información Geográfica

Universidad Industrial de Santander  
Facultad de Ingenierías Físicoquímicas  
Escuela de Geología

Bucaramanga

2021

**Dedicatoria**

*A Dios, quien fue mi roca y fortaleza en los momentos difíciles*  
*A Norma Marín, mi madre, quien creyó en mí y nunca faltó en darme aliento*  
*A Sergio Quintero, mi padre, quien supo corregirme cuando fue necesario*  
*A mis hermanas, Manuela y Victoria, mis amigas y compañeras de vida*  
*A mis amigos, por ayudarme a recorrer el camino y hacerlo llevadero*  
*A la gente que amo y siempre llevaré en mí.*

*Lucas Quintero Marín.*

### **Dedicatoria**

*A Dios que ha sido bueno conmigo y me permitió vivir y completar esta etapa de mi vida*

*A mis padres Henry Ávila y Leorenis Marín, por apoyarme, darme su amor incondicional y ser mi ejemplo a seguir en todo momento*

*A mis hermanos Yurley, Shirley, Giselle y Jesús David, por siempre estar cuando los necesité y formar parte de mi familia, mi mayor bendición*

*A mis amigos, por hacer de esta etapa la mejor*

*A mi novia Laura, por enseñarme lo que es un amor sincero*

*A todas las personas con las que coincidí a lo largo de estos años y que llevo siempre en el corazón.*

*Miguel Ángel Ávila Marín*

### **Agradecimientos**

A la Universidad Industrial de Santander por acogernos y prepararnos durante estos años de formación.

A nuestro codirector Julián Rojas quien trabajó hombro a hombro con nosotros y nos acompañó de manera paciente y constante durante todo el desarrollo del proyecto.

A nuestro director Julián De Bedout por sus oportunas correcciones y el conocimiento compartido desde el inicio de este trabajo.

A nuestro codirector Jose Pedro Mora por la asesoría brindada en momentos cruciales.

A la escuela de geología y sus profesores por aportar a nuestro proceso de aprendizaje.

A nuestros amigos, Andrés, Felipe, Giovanni, Jhoan, María, Mario y Omar por hacer de la universidad una etapa memorable.

A Carlos, Juandiego y Felipe por su amistad sincera.

A Jorge, Eduardo y Ángel por su excelente amistad y compañía.

## Tabla de Contenido

Introducción.....	15
1. Objetivos.....	16
1.1. Objetivo General .....	16
1.2. Objetivos Específicos .....	16
2. Localización.....	17
3. Metodología .....	18
4. Marco Geológico.....	26
4.1. Marco Estructural.....	26
4.2. Estratigrafía.....	28
4.2.1. Grupo Yura .....	30
4.2.2. Formación Soraya.....	31
4.2.3. Formación Mara .....	31
4.2.4. Formación Ferrobamba.....	32
5. Marco Teórico.....	32
5.1. Caracterización Estadística .....	32
5.1.1. Test de Normalidad de Anderson-Darling.....	33
5.1.2. Diagramas de Dispersión (Scatter Plots) .....	34
5.1.3. Regresión Lineal Simple.....	35
5.1.4. Regresión Lineal Múltiple .....	37
5.1.5. Valores Atípicos (Outliers) .....	37
5.1.6. Coeficiente de Correlación Lineal de Pearson.....	38
5.1.7. Matriz de Correlación Lineal .....	40
5.1.8. Dendogramas (Cluster Analysis) .....	41
5.2. <i>Machine Learning</i> .....	42
5.2.1. Redes Neuronales Artificiales.....	43
5.3. Imágenes Satelitales .....	46
5.3.1. Principios de la Radiación Electromagnética.....	47
5.3.2. Espectro Electromagnético .....	49
5.3.3. Interacción de la Energía Electromagnética con los Objetos en Superficie Terrestre	50
5.3.4. Curvas de Reflectancia Espectral.....	52
5.3.5. Sistemas de Sensores Remotos .....	53
5.3.6. Bandas Espectrales .....	56
5.3.7. Tipos de Resolución de un Sistema Sensor .....	57

5.3.8. Combinación de Bandas .....	58
5.3.9. Teledetección Aplicada a la Minería .....	59
5.3.10. ASTER (Advanced Spaceborne Thermal Emission and Refelection Radiometer) .	60
5.3.11. Cruces de Bandas en ASTER.....	61
6.3.12. Pre-Procesamiento de Imágenes.....	61
6. Resultados.....	64
6.1. Análisis Estadísticos.....	64
6.1.1. Cluster Analysis-Estadística Multivariable .....	65
6.1.2. Dendogramas .....	65
6.1.3. Matriz de Correlación Lineal de Pearson .....	67
6.1.4. Diagramas de cajas (Boxplots).....	68
6.1.5. Diagramas de Dispersión .....	69
6.1.6. Diagramas de Dispersión 2D y Regesión Lineal .....	70
6.1.7. Interpolación - IDW .....	73
6.2. Imágenes Satelitales .....	79
6.2.1. Landsat-8 .....	79
6.1.2. Procesamiento de imágenes ASTER.....	93
6.3. Machine Learning .....	98
6.3.1. Resultados de la adquisición y preprocesamiento de datos de entrada .....	98
6.3.2. Resultados de la aplicación del modelo.....	107
7. Conclusiones.....	116
Referencias Bibliográficas.....	119

**Lista de figuras**

<b>Figura 1.</b> Mapa de localización de la zona de estudio, Franja Metalogenética XV.....	17
<b>Figura 2.</b> Esquema general de las fases principales del proyecto.....	18
<b>Figura 3.</b> Diseño de muestreo sistemático (Tomado de Vargas et al., 2017).....	20
<b>Figura 4.</b> Histograma de Cu, uno de los elementos analizados, a la izquierda el histograma de los datos crudos mostrando distribución log-normal, a la derecha se observa histograma posterior a transformación logarítmica mostrando su nueva distribución normal.....	21
<b>Figura 5.</b> Bandas presentes en el sensor Landsat-8 OLI / TIRS (Tomado de USGS, 2017).	23
<b>Figura 6.</b> Componentes de los sistemas de Pórfidos-Skarn en la zona de estudio, mostrando los procesos constituyentes, elementos de interés y mapas predictores utilizados como inputs en el entrenamiento del algoritmo de Redes Neuronales, esquema (Tomado y modificado de Sun et al., 2019).....	24
<b>Figura 7.</b> Mapa simplificado de dominios Estructurales en la región de Cusco, Abancay y Sucuani, indicando la ubicación de la zona de estudio (Tomado y modificado de Carlotto, 2006).....	28
<b>Figura 8.</b> Columna estratigráfica del distrito Tintaya, mostrando los pulsos de dioritas, monzonitas y diques máficos emplazados en dos etapas. Esta columna se tomó como la tipo (Tomado y modificado de Maher, 2010).....	30
<b>Figura 9.</b> Diagrama de dispersión tradicional. a) Correlación positiva, b) correlación negativa y c) ninguna correlación (Tomado de Spiegel, 2009).....	34
<b>Figura 10.</b> Diagrama de dispersión 3D.....	35
<b>Figura 11.</b> Puntos correspondientes a observaciones del modelo de regresión lineal simple, se observa la recta de regresión y las distancias a puntos fuera de ella (Tomado de Devore, 2016).....	36
<b>Figura 12.</b> Distribución de $\epsilon$ , donde 0 equivale a la media aritmética de los datos (Tomado de Devore, 2016).....	37
<b>Figura 13.</b> Diagrama de caja modificado. Mostrando un ejemplo de las velocidades de datos en aeropuertos para erizon en Mbps, los valores atípicos son los que están por fuera del diagrama (Tomado de Triola, 2018).....	38
<b>Figura 14.</b> Matriz de correlación, ejemplo de la base de datos del CHSI (Tomado de Zadeh, 2019).....	40
<b>Figura 15.</b> Dendograma de un conjunto de datos (Tomado de Forina et al., 2002).....	41
<b>Figura 16.</b> Arquitectura de una red neuronal artificial con múltiples capas o layers (Tomado de Abraham, 2005).....	46
<b>Figura 17.</b> Longitud de onda y amplitud de una onda típica (Modificado de Demirel & Düzgün, 2011).....	48
<b>Figura 18.</b> Frecuencia de ondas (Modificado de Demirel & Düzgün, 2011).....	48
<b>Figura 19.</b> Espectro electromagnético (Tomado de Ordoñez, 2017).....	50
<b>Figura 20.</b> Distribución de la energía solar total entrante con longitudes de onda menores a 4.5 $\mu\text{m}$ (Tomado de Demirel & Düzgün, 2011).....	51
<b>Figura 21.</b> Reflexión especular y difusa (Tomado de Demirel & Düzgün, 2011).....	52
<b>Figura 22.</b> Curvas de reflectancia espectral típicas para características comunes en la Tierra (Tomado de Demirel & Düzgün, 2011).....	53

<b>Figura 23.</b> Clasificación de sistemas de sensores remotos con 3 perspectivas diferentes (Tomado de Demirel & Düzgün, 2011). .....	54
<b>Figura 24.</b> Sensores pasivos (a) y sensores activos (b) (Tomado de Demirel & Düzgün, 2011).....	55
<b>Figura 25.</b> Componentes de los sensores remotos (Tomado de Demirel & Düzgün, 2011)	56
<b>Figura 26.</b> Comparación de las bandas de los distintos sensores que se encuentran a lo largo de las diferentes longitudes de onda (Tomado de USGS, 2014). .....	56
<b>Figura 27.</b> Tipos de resolución que se presentan en una imagen (Modificado de Humboldt, 2019).....	58
<b>Figura 28.</b> Ejemplo de una imagen NDVI representando la cubierta vegetal global en varios tonos de verde (Tomado de Khorram et al., 2012). .....	59
<b>Figura 29.</b> Distribución de bandas ASTER y Landsat (Tomado de Kalinowski & Oliver, 2004).....	60
<b>Figura 30.</b> Compilación de cruces de bandas comunes con imágenes ASTER, se utilizan para resaltar aspectos de interés geológico utilizados en minería (Tomado de Kalinowski & Oliver, 2004).....	61
<b>Figura 31.</b> Principales correcciones en el pre-procesamiento de imágenes (Modificado de Khorram et al., 2012). .....	62
<b>Figura 32.</b> Clasificación no supervisada del área de la Bahía de San Francisco (Tomado de Khorram et al., 2012). .....	63
<b>Figura 33.</b> Resumen de parámetros estadísticos de datos geoquímicos procesados de sedimentos activos en la Franja Metalogenética XV. Para cada elemento se generaron datos estadísticos de valor mínimo y máximo (Maximum and minimum), cuartil 1 y 3 (1 st Quartile-3rd), mediana (Median), Desviación estándar (StDev) y el P-Valor, A-Valor generados mediante el Test de Anderson-Darling. ....	64
<b>Figura 34.</b> Dendograma de todos los elementos menores y traza (a excepción del Fe) en ppm estudiados. ....	66
<b>Figura 35.</b> Dendograma filtrado de elementos estudiados en ppm. Se observan 2 grupos de elementos con distribución similar (Grupo 1: Au, Ag, Cu, Mo, As, Zn, Cd, Cr, Sn, W, Ti, V, Zr, Y; Grupo 2: Ba, In, Cs, Tl, Li, Ga, P, Ge, Nb, Ce, Lu, Rb, U, Th, Hf). ....	67
<b>Figura 36.</b> Matriz de correlación lineal de Pearson entre elementos geoquímicos estudiados. Los intervalos resaltados son: alto (>0.7), media (0.4-0.7), baja (-4-(+4)), media inversa ((-0.4)-(-0.7) y alta inversa (<-0.7). ....	68
<b>Figura 37.</b> Comparación de diagramas de cajas de los diferentes elementos de interés en ppm, organizados por grupos de correlación geoquímica. Grupo 1 (azul) y Grupo 2 (amarillo). ....	69
<b>Figura 38.</b> Diagramas de dispersión de tríos de elementos mejor correlacionados. En orden descendente las mejores correlaciones fueron (Li-Ga-P) > (Ce-Lu-Rb) > (U-Th-Hf) > (Li-Lu-Rb).....	70
<b>Figura 39.</b> Gráfico de dispersión mostrando ecuación de regresión lineal y correlación entre Ce y Lu. Se observa un coeficiente de determinación del 84.6% y desviación estándar de 6.89. Estos dos elementos son los que mejor se correlacionan de todos.....	71
<b>Figura 40.</b> Gráfico de dispersión mostrando ecuación de regresión lineal y correlación entre Li y Ga. Se observa un coeficiente de determinación de 67.6% y desviación estándar de 3.81. ....	71

<b>Figura 41.</b> Gráfico de dispersión mostrando ecuación de regresión lineal y correlación entre el U y Th. Se observa un coeficiente de determinación de 61.5 % y desviación estándar de 0.579.....	72
<b>Figura 42.</b> Gráfico de dispersión mostrando ecuación de regresión lineal y correlación entre el Zn y Cd. Se observa un coeficiente de determinación de 55.8% y desviación estándar de 86.92.....	72
<b>Figura 43.</b> Histogramas de los elementos no encontrados en el Atlas Geoquímico del Perú y algunos de los incluidos pero realizados nuevamente para corroborar los límites de valores de fondo y umbrales. A la izquierda se encuentran los elementos mostrando distribución logarítmica, a la derecha nuevos histogramas de elementos normalizados. Esta transformación es de vital importancia para el cálculo de los valores de fondo y umbrales de los elementos y para la unificación posterior mediante redes neuronales.....	75
<b>Figura 44.</b> Mapas de Anomalías Geoquímicas en ppm de elementos analizados. Aunque todos los elementos presentan anomalías las mayores son de Fe, Cu, Au, Cr, Pb y Zn. Casi todos los elementos presentan valores anómalos en la zona suroeste de la franja metalogenética XV. Los elementos con menores anomalías fueron Cd y Sb. ....	76
<b>Figura 45.</b> Segundo Bloque de anomalías geoquímicas de elementos analizados en ppm.	77
<b>Figura 46.</b> A, librería espectral de los minerales epidota, calcita, moscovita, caolinita, clorita y alunita. B, librería espectral de los minerales limonita, jarosita, hematita y goethita (Clark et al., 1997).....	80
<b>Figura 47.</b> Imagen Landsat-8 LC80050702019218LGN00 con combinaciones de bandas RGB 4/2,6/7,5 y RGB 4/2,6/7,10 respectivamente, escala 1:800000. ....	81
<b>Figura 48.</b> Acercamiento de la imagen Landsat-8 LC80050702019218LGN00 mostrando en detalle depósitos tipo skarn.....	82
<b>Figura 49.</b> Imagen Landsat-8 LC80050692019218LGN00 con combinaciones de bandas RGB 4/2,6/7,5 y RGB 4/2,6/7,10 respectivamente.....	83
<b>Figura 50.</b> Acercamiento de la imagen Landsat-8 LC80050702019218LGN00 mostrando en detalle depósitos tipo skarn (Círculos morados), pórfidos de Cu-M (rombos rojos), vetas (triángulos verdes) y epitermales (círculos fucsias) de intermedia sulfuración. Se utilizó la misma combinación en falso color 4/2, 6/7, 10. ....	84
<b>Figura 51.</b> Imagen Landsat-8 LC80040702020150LGN00 con combinaciones de bandas RGB 4/2,6/7,5 y RGB 4/2,6/7,10 respectivamente. Se observan claramente 2 tonalidades predominantes y como con la banda 10 se resaltan de mejor manera zonas ricas en óxidos de Fe.....	85
<b>Figura 52.</b> Acercamiento de la imagen Landsat-8 LC80050702019218LGN00 mostrando en detalle depósitos clasificados como vetas, pórfidos de Cu-Mo, epitermales indiferenciados y de intermedia sulfuración. ....	86
<b>Figura 53.</b> Acercamiento de la imagen Landsat-8 LC80050702019218LGN00 mostrando en detalle depósitos clasificados como vetas, pórfidos de Cu-Mo y skarns, la mina conocida como las Bambas se muestra en el rectángulo negro. La combinación de bandas fue la misma que en las anteriores imágenes.....	87
<b>Figura 54.</b> Vista panorámica hacia el noroeste del depósito Ferrobamba (Tomado de Sanchez et al., 2019). ....	88
<b>Figura 55.</b> Acercamiento de la imagen Landsat-8 LC80040702020150LGN00 mostrando en detalle yacimientos de skarns (Círculos morados) y vetas (triángulos amarillos). La mina conocida como Constancia se encuentra delimitada por el rectángulo negro. ....	89

<b>Figura 56.</b> Imagen Landsat-8 LC80040692020150LGN00 con combinaciones de bandas RGB 4/2,6/7,5 y RGB 4/2,6/7,10 respectivamente.....	90
<b>Figura 57.</b> Acercamiento de la imagen Landsat-8 LC80040692020150LGN00 mostrando en detalle depósitos clasificados como vetas, pórfidos de Cu-Mo y skarns. ....	90
<b>Figura 58.</b> Imagen Landsat-8 LC80030702020239LGN00 con combinaciones de bandas RGB 4/2,6/7,5 y RGB 4/2,6/7,10 respectivamente.....	91
<b>Figura 59.</b> Acercamiento de la imagen Landsat-8 LC80030702020239LGN00 mostrando en detalle yacimientos de skarns, los distritos mineros de Antapaccay (delimitado por el rectángulo rojo) y Tintaya (delimitado por el rectángulo negro).....	92
<b>Figura 60.</b> Mapa de alteraciones de óxidos de hierro. Se delimitaron las zonas en respuesta al procesamiento de las imágenes satelitales. El mapa muestra la distancia desde cualquier punto a una zona de alteración de óxidos de hierro. Las zonas más alejadas se encuentran al noreste en el dominio de la Cordillera Oriental. ....	93
<b>Figura 61.</b> Ubicación de Imágenes ASTER utilizadas. Estas se escogieron posterior a los análisis de imágenes Landsat-8 y de los análisis geoestadísticos y geológicos, donde se predeterminaron zonas de interés.....	94
<b>Figura 62.</b> A. índice de alteración Argílica avanzada, las zonas amarillas-rojas indican la alta absorción de los minerales Alunita-pirofilita en la banda 5 (Mars & Rowan, 2006). B. índice de minerales alterados con contenido de OH (OHI). C. Composición en falso color para cálculo de alteración fílica. D. Combinación 432 de la misma zona de la imagen C. .	96
<b>Figura 63.</b> E. combinación en falso color RGB para identificar alteración argílica, R: 4/5, G:5/6, B:6/7. Los yacimientos confirmados en la zona sirvieron de guía para identificar las tonalidades asociadas. F. la relación (6+9)/(7/8) se utilizó para encontrar zonas con contenido de epidota, clorita, anfibolita, las zonas de mayor reflectancia se observan en rojo y están cerca de yacimientos de tipo skarn confirmados. G y H. Relaciones de banda en RGB para identificar zonas de óxidos férricos (2/1). Óxidos ferrosos (1/2 + 5/2) y arcillas 5/7. ....	97
<b>Figura 64.</b> Mapa de alteraciones hidrotermales. Algunas fueron tomadas de la bases de datos del Geocatmin, (2019).....	98
<b>Figura 65.</b> Resultado del análisis de patrón de puntos mostrando la distancia de vecinos cercanos y la probabilidad de encontrar un nuevo yacimiento desde otro ya identificado. Distancia mínima[m]: 275.072, distancia máxima[m]: 65730.036. ....	99
<b>Figura 66.</b> Localización de depósitos y selección de puntos de no depósito. Se realizó un buffer de 15.591 metros que representa el límite. Los puntos de no yacimiento se generaron mediante la herramienta de ArcGis Create Random Points, fuera del buffer realizado. ...	100
<b>Figura 67.</b> Arriba. Mapa geológico utilizado para el cálculo de distancias a intrusivos y contactos Intrusivos-Calizas donde se resaltan los intrusivos (relacionados al batolito de Abancay) y las calizas (principalmente de Ferrobamba), en azul las demás litologías. Abajo. Mapa de proximidad a intrusivos aflorando en la zona de estudio. Los datos fueron tomados del Mapa Geológico del Perú a escala 1:100000 del INGEMMET.....	102
<b>Figura 68.</b> Mapa de proximidad a contactos Intrusivos-Calizas. Se realizó en base al Mapa Geológico a escala 1:100.000 del Perú. ....	103
<b>Figura 69.</b> Los datos se adquirieron de Mapas Geológicos de la zona a escala 1:1M, 1:100K y 1:50K. Esa información se completó realizando una extracción de lineamientos en Arcgis con las ayuda de imágenes satelitales y DEMs. ....	104
<b>Figura 70.</b> Mapa de densidad de lineamientos. Se generó con las fallas y lineamientos de la Figura 69.....	105

**Figura 71.** Mapa de proximidad a intersección de fallas en roca encajante (calizas). Se utilizaron datos del mapa geológico del Perú a escala 1:50.000. .... 106

**Figura 72.** Mapa de proximidad a pliegues, la información se tomó del mapa geológico del Perú a escala 1:100k del INGEMMET. Yacimientos de Skarn de cobre como el Corocochuayco en el distrito de Tintaya y el de Accha, las mineralizaciones se encuentran en la zona de charnela o flancos pliegues (Boni et al., 2009; Maher, 2010). .... 107

**Figura 73.** Código realizado en Jupyter Notebook de entrenamiento de la red neuronal artificial. .... 108

**Figura 74.** Código realizado en Jupyter Notebook de predicción del algoritmo de RNA. 108

**Figura 75.** El código recibió el nombre de ‘clf’, ‘MLPClassifier’: es el nombre del algoritmo supervisado utilizado, ‘activation’: es la función de activación de las capas ocultas, ‘solver’: determina el peso de la optimización, ‘alpha’: es un parámetro de regular. .... 109

**Figura 76.** Mapa de predictividad de ocurrencia mineral..... 111

**Figura 77.** Acercamiento a zonas con probabilidad de ocurrencia mineral mayor al 95%.112

**Figura 78.** Ubicación y descripción de zonas con probabilidad mayor al 95% de ocurrencias minerales donde no hay yacimientos confirmados hasta la fecha. .... 114

**Figura 79.** Tabla resumen de las propiedades de las zonas prospectadas con probabilidad superior al 95% de encontrar ocurrencias minerales. .... 115

### **Lista de Abreviaciones**

AD: Árboles de Decisión

Alt: Alteración

BAY: Batolito Andahuaylas-Yauri

Fm: Formación

MLP: Multi Layer Perceptron

RNA: Redes Neuronales Artificiales

SVMs: Support Vector Machines

## Resumen

**Título:** Prospección geológica mediante la aplicación de técnicas de machine learning, geoestadística y procesamiento de imágenes satelitales en la franja metalogenética XV (Perú)\*

**Autores:** Miguel Ángel Ávila Marín\*\* y Lucas Quintero Marín\*\*

**Palabras Clave:** Machine learning, redes neuronales artificiales, procesamiento de imágenes satelitales, análisis geoestadísticos, prospección minera, Landsat-8, ASTER.

**Descripción:** La franja metalogenética XV se ubica al sureste de Perú, entre la Cordillera Oriental y el Altiplano de las regiones de Ayacucho, Cusco y Puno. Esta ha sido reconocida por la presencia de yacimientos de tipo Skarn Fe-Cu y pórfidos de Cu-Au (Mo-Zn). En respuesta a la necesidad de realizar nuevas prospecciones minerales optimizando la relación de costos y beneficios se realizó la prospección geológica utilizando datos libres de sedimentos activos y geología del INGEMMET, imágenes satelitales ASTER y Landsat-8 del Servicio Geológico de los Estados Unidos. Con los datos de sedimentos activos se realizaron análisis de estadística multivariable (Correlación lineal de Pearson, dendogramas y matrices de correlación lineal) para determinar zonas de anomalías y asociaciones geoquímicas. Con el tratamiento de imágenes satelitales mediante combinaciones y relaciones de banda se delimitaron áreas con alteraciones hidrotermales y de óxidos de hierro. En total se generaron 19 mapas que representan fuente, transporte, trampas físicas o procesos de depositación química asociados a sistemas de Pórfidos-Skarn, estos se utilizaron como capas de entrada y se integraron mediante machine learning usando redes neuronales artificiales. Las redes neuronales demostraron una gran utilidad en la prospección minera de yacimientos de tipo pórfido y skarn al generar un mapa en el que las zonas delimitadas de alta y muy alta prospectividad ocuparon el 16% de la zona de estudio y abarcaron el 84% de los yacimientos confirmados.

---

\* Trabajo de Grado

\*\* Facultad de Ingenierías Físicoquímicas. Escuela de Geología. Director: Julián David De Bedout Ordóñez. Magíster en Ingeniería Geológica. Codirectores: Julian David Rojas Delgado. Geólogo. Jose Pedro Mora Ortiz. Magíster en Geofísica y Sistemas de Información Geográfica.

### Abstract

**Title:** Geological prospecting through the application of machine learning, geostatistics and satellite image processing techniques in the metallogenic belt XV (Perú) \*

**Author:** Miguel Ángel Ávila Marín\*\*, Lucas Quintero Marín\*\*.

**Key Words:** Machine learning, artificial neural networks, satellite image processing, geostatistical analysis, mineral prospecting, Landsat-8, ASTER.

**Description:** The metallogenic belt XV is located in southeastern Peru, between the Eastern Cordillera and the Altiplano of the regions of Ayacucho, Cusco and Puno. It has been recognized by the presence of Fe-Cu skarn and Cu-Au (Mo-Zn) porphyry deposits. In response to the need to carry out new mineral prospecting optimizing the cost-benefit ratio, geological prospecting was carried out using free data of active sediments and geology from INGEMMET, ASTER and Landsat-8 satellite images from the U.S. Geological Survey. Multivariate statistical analyses (Pearson linear correlation, dendograms and linear correlation matrices) were performed on the active sediment data to determine anomaly zones and geochemical associations. With the treatment of satellite images by means of combinations and band ratios, areas with hydrothermal and iron oxide alterations were delimited. In total 19 maps representing source, transport, physical traps or chemical deposition processes associated with Porphyry-Skarn systems were generated, these were used as input layers and integrated by machine learning using artificial neural networks. The neural networks proved to be very useful in prospecting for porphyry and skarn deposits by generating a map in which the delimited zones of high and very high prospectivity occupied 16% of the study area and covered 84% of the confirmed deposits.

---

\* Bachelor Thesis

\*\* Physicochemical Engineering Faculty. Geology School. Director: Julián David De Bedout Ordóñez Master in Geological Engineering. Co-directors: Julian David Rojas Delgado. Geologist. Jose Pedro Mora Ortiz. Master in Geophysicist and Geographic Information Systems.

## Introducción

El trabajo se desarrolla bajo la modalidad de investigación con el fin de realizar la prospección geológica de la franja metalogenética XV de Perú, la cual está localizada al suroeste del país, entre la Cordillera Occidental y el Altiplano de las regiones de Ayacucho, Cusco y Puno (Carlotto et al., 1999; Perelló *et al.*, 2003). Ha sido reconocida por la presencia de yacimientos de tipo skarn Fe-Cu y recientemente se ha destacado como una provincia importante de pórfidos de Cu-Au (Mo-Zn) (Carlotto *et al.*, 1999; Perelló *et al.*, 2003).

La creciente complejidad a la hora de hallar depósitos y zonas mineralizadas hacen de la prospección minera un proceso largo y costoso, esto genera la necesidad de implementar nuevas formas de estudio y el replanteamiento de los métodos actuales, intentando disminuir precios y aumentando las posibilidades de hallazgo.

Para Perú la minería representa un 10% de su PBI y el 61% de sus exportaciones según datos del Anuario Minero del Perú 2019 (Sanchez *et al.*, 2019). Este se encuentra geológicamente dividido en lo que se ha llamado franjas metalogenéticas, lo que permite tener una noción de qué mineralizaciones y materias primas se pueden encontrar y donde hallarlas. El objetivo entonces es, con la ayuda de los métodos tradicionales como el procesamiento de imágenes satelitales, la geoestadística, geoquímica y el aporte de las nuevas técnicas de *machine learning*, definir zonas de interés económico con un menor riesgo exploratorio.

## **1. Objetivos**

### **1.1 Objetivo General**

Reconocer mediante técnicas de geoestadística, imágenes satelitales y machine learning, zonas de prospección minera en la franja metalogénica XV al sur de Perú.

### **1.2 Objetivos Específicos**

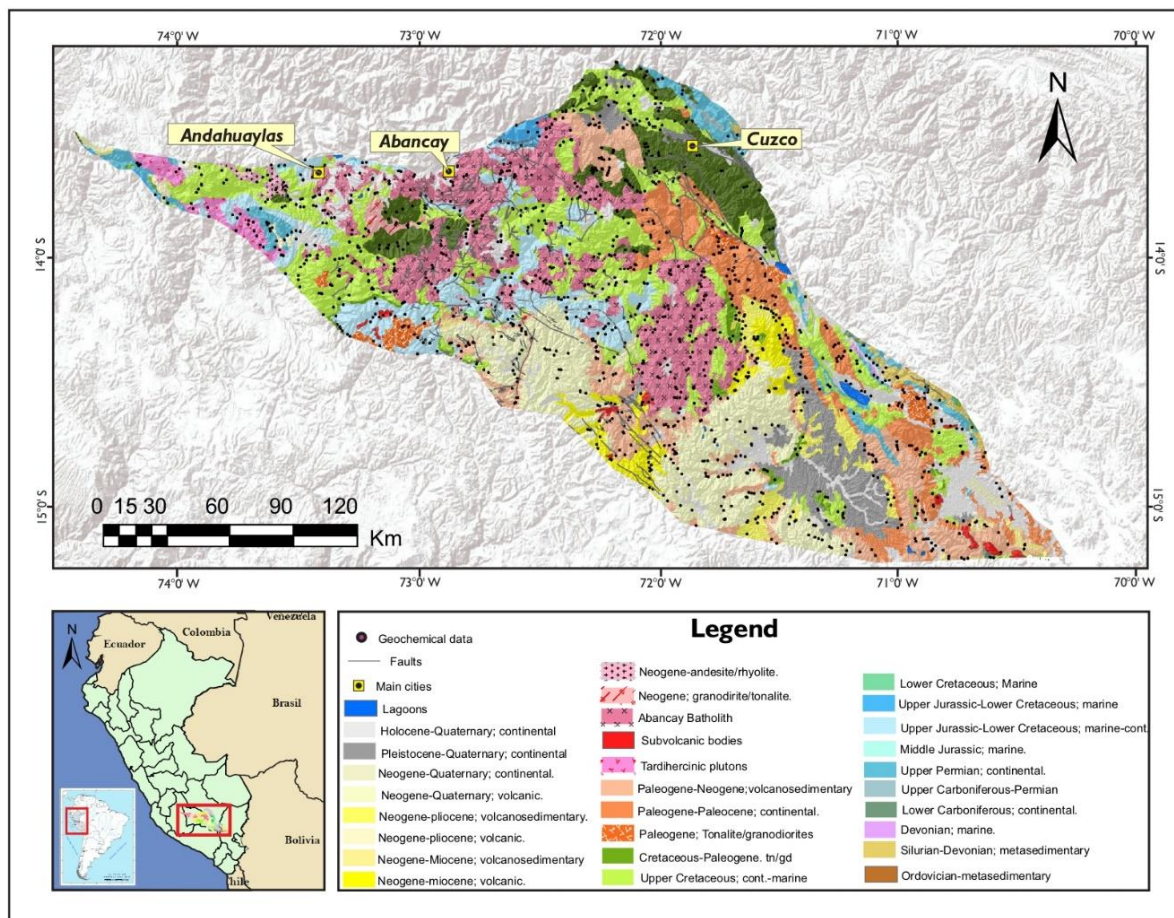
- Estudiar las influencias litológicas, estructurales y de temporalidad en los procesos de mineralización de la franja metalogénica XV.
- Comparar y examinar individualmente los resultados obtenidos mediante machine learning y el procesamiento de imágenes satelitales.
- Analizar la utilidad de los modelos predictivos de machine learning en la prospección mineral de depósitos de tipo pórfidos y skarn.

## 2. Localización

La zona de estudio ubicada en la **Figura 1**, corresponde a la Franja metalogenética XV (INGEMMET, 2009). Se encuentra al suroeste del Perú, entre la Cordillera Occidental y el Altiplano de las regiones de Ayacucho, Cusco y Puno (Carlotto et al., 1999; Perelló *et al.*, 2003). La franja se extiende desde el noroeste del Lago Titicaca hasta el sur del municipio de Ayacucho en los departamentos de Ayacucho, Apurímac, Cusco, Arequipa y Puno, abarcando una distancia de 469 Km y un área de aproximadamente 36.000Km<sup>2</sup>.

Se encuentra controlada por los sistemas de fallas Urcos-Sicuani-Ayaviri y Cusco-Lagunillas-Mañazo, Abancay-Andahuaylas-Totos-Chincheros-Lincapa (Carlotto *et al.*, 2005).

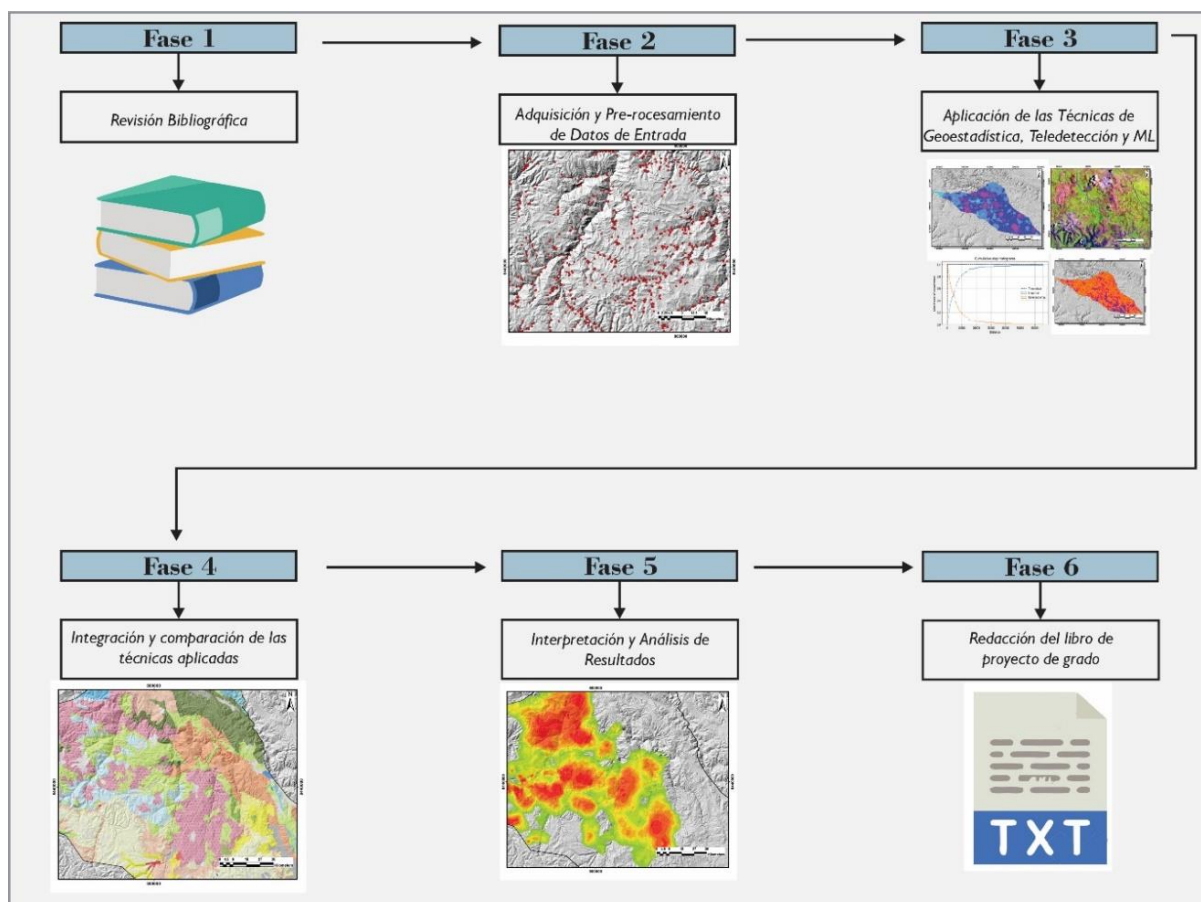
**Figura 1.** Mapa de localización de la zona de estudio, Franja Metalogenética XV.



### 3. Metodología

Este proyecto tiene la siguiente estrategia de trabajo, en orden de cumplir con los objetivos trazados, se plantea una metodología compuesta por 6 fases que abarcan todo el proceso desde el momento en que se adquiere la información hasta la finalización y entrega del documento final. En la **Figura 2** se presenta un esquema general de dicha metodología.

**Figura 2.** Esquema general de las fases principales del proyecto.



#### Fase 1: Revisión Bibliográfica

En la primera fase se realizó una búsqueda literaria sobre los principios geoestadísticos, geoquímicos, de procesamiento de imágenes satelitales y Machine Learning, esto con la finalidad de adquirir un dominio del tema de trabajo, técnicas aplicadas y se generó un marco teórico de referencia sobre el que se aplicó el componente práctico. También se investigó la geología de la Franja Metalogenética XV, para esto se recurrió a diferentes fuentes

bibliográficas que facilitaron el entendimiento de la zona a fin de realizar una correlación entre el estudio realizado y lo encontrado en la literatura.

## **Fase 2: Adquisición y pre-procesamiento de datos de entrada**

Para la realización del proyecto se reúnen un compilado de datos que son los *inputs* o datos de entrada, estos consisten principalmente en datos geoquímicos libres de sedimentos activos de la Franja Metalogenética XV, información geológica del Perú en formato shape, e imágenes satelitales libres ASTER 1T y Landsat 8. Estos datos crudos se sometieron a pre-procesamientos antes de aplicar los análisis geoestadísticos, procesamiento de las imágenes satelitales y aplicación de los algoritmos de Redes Neuronales Artificiales.

### ***Fase 2.1: Análisis Geoestadísticos***

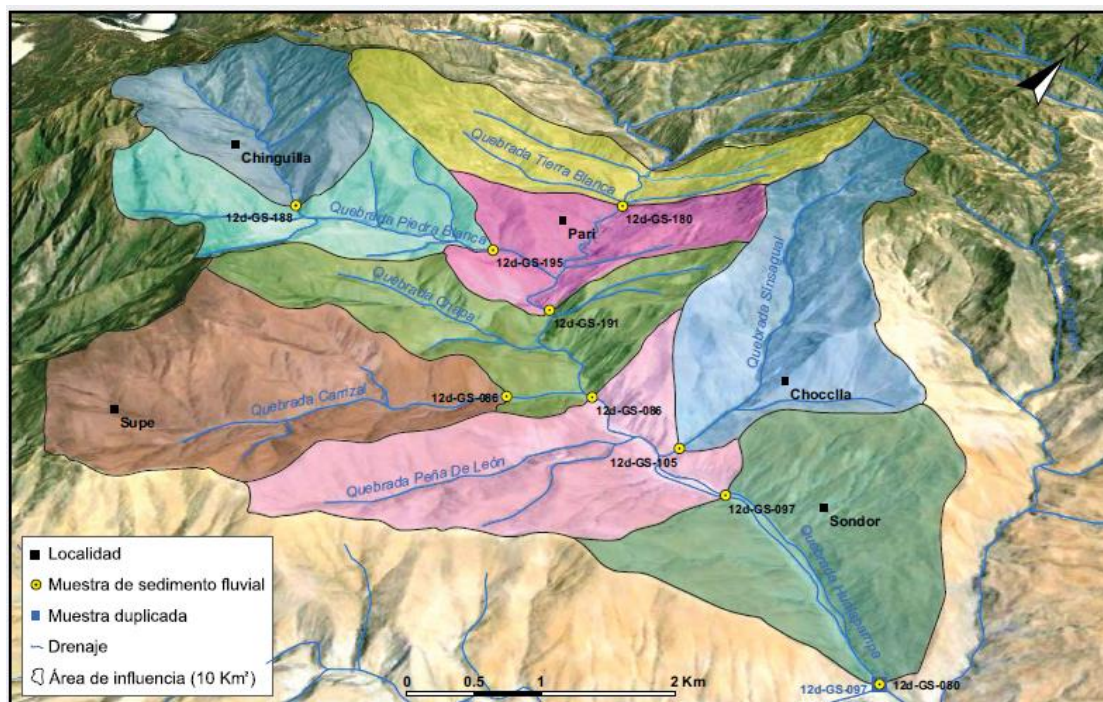
Para el desarrollo de los análisis estadísticos se utilizaron datos geoquímicos de estudios del Instituto Geológico, Minero y Metalúrgico (INGEMMET) y valores de anomalías geoquímicas del Atlas Gequímico del Perú (Vargas *et al.*, 2017).

Estos estudios fueron realizados en franjas transversales al Orógeno Peruano donde se generó información geoquímica de sedimentos activos de quebrada a escala regional. Dicha información ofrece resultados analíticos de 53 elementos, entre los que destacan elementos traza como el Au, Ag, Cu, Mo, Pb, Zn y Cd, además elementos mayores como Al, Fe y Mg. Todos los elementos fueron analizados mediante Espectrometría de Masas con Plasma Acoplado Inductivamente (ICP-MS), a excepción del oro el cual fue analizado usando Ensayo al Fuego (FA, por sus siglas en inglés), este es el proceso estándar para obtener concentraciones de elementos del platino y oro (PGE) (Vargas *et al.*, 2017).

La obtención de sedimento fluvial se realiza siguiendo un diseño de muestreo sistemático en la red hidrográfica, con la finalidad de obtener información geoquímica de la superficie del terreno delimitada por la divisora de aguas como se muestra en la **Figura 3**.

El tipo de muestreo implementado es de tipo sistemático estratificado, es este la determinación de los puntos de muestreo se efectuó a partir de una investigación detallada de la zona y basándose en información geológica-minera disponible (Vargas *et al.*, 2017).

**Figura 3.** Diseño de muestreo sistemático (Tomado de Vargas *et al.*, 2017)

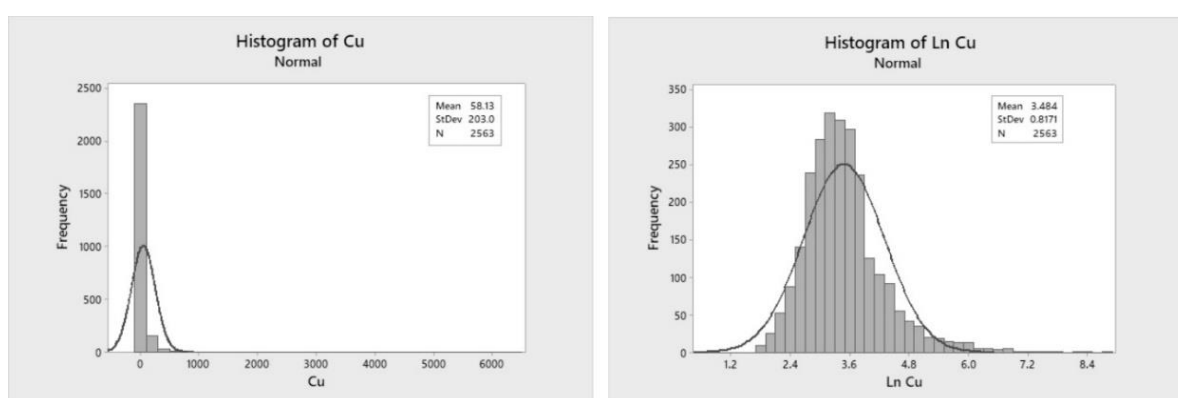


Para los análisis geoquímicos se utilizaron un total de 2563 muestras distribuidas a lo largo de la franja metalogenética XV, estas fueron filtradas y separadas en: elementos mayores, menores y traza; siendo estos últimos los utilizados para los diferentes análisis geoestadísticos en partes por millón (ppm) para facilitar los cálculos.

Los datos geoquímicos en brutos se trataron previamente para realizar un análisis estadístico correcto, este proceso constó de un reemplazo log-normal multiplicativo utilizando el paquete de programas del Software R llamado 'zCompositions', esto se hizo para imputar y tratar los datos con valor 0, ya sea por ceros redondeados, datos faltantes o que estuvieran por debajo de los límites de detección del equipo de medición (Albaladejo & Fernández, 2015). Posterior al tratamiento, los datos se encuentran listos para los análisis estadísticos a realizar utilizando Microsoft Excel 2016.

También, debido al comportamiento No-Normal de los elementos analizados en el estudio se realizó una transformación logarítmica **Figura 4**, a fin de encontrar los valores de fondo y umbrales de alguno de los elementos cuyos valores no fueron encontrados en el Atlas Geoquímico del Perú.

**Figura 4.** *Histograma de Cu, uno de los elementos analizados, a la izquierda el histograma de los datos crudos mostrando distribución log-normal, a la derecha se observa histograma posterior a transformación logarítmica mostrando su nueva distribución normal.*



La información geológica correspondiente a la zona de estudio se adquirió de la plataforma virtual Geocatmin; del Instituto Geológico, Minero y Metalúrgico, INGEMMET. Se descargaron archivos libres en formato Shapefile de Fallas y lineamientos a escala 1:1M, 1:100k y 1:50k, pliegues a escala 1:100K, dominios estructurales, Geología a escala 1:1M y 1:100k, dataciones geocronológicas, Neo-tectónica de falla reciente y ubicación de yacimientos mineros clasificados.

### ***Fase 2.2: Imágenes satelitales***

Para el desarrollo del proyecto se utilizaron 5 imágenes Landsat-8 y 16 imágenes ASTER, estas se tomaron de la página Earth Explorer del Servicio Geológico de los Estados Unidos (USGS) <https://earthexplorer.usgs.gov/>, con una nubosidad menor al 10% y con el número total de bandas, tanto para ASTER como Landsat-8.

Las imágenes obtenidas no se encuentran listas para su procesamiento por lo que es

necesario una etapa previa en la que se minimicen las distorsiones y/o errores que podrían dificultar una clasificación exitosa y para extraer o mejorar la información más importante de una imagen (Khorram *et al.*, 2012; Arozarena & Pastor, 2016). Las principales correcciones son la geométrica, atmosférica y radiométrica (Arozarena & Pastor, 2016).

En el caso de las imágenes de Landsat-8 y ASTER la corrección geométrica no es necesaria, es decir, vienen corregidas de origen.

Para el pre-procesamiento de las imágenes ASTER, se realiza primero la corrección radiométrica para eliminar el efecto del sensor sobre los datos (Pons & Solé-Sugrañes, 1994) y dejarlos en formato de radiancia, posterior a esto se aplica el *layer stacking* o apilado, donde se juntan las 14 bandas con diferentes resoluciones espaciales (VNIR, SWIR, TIR) y se le aplica la corrección atmosférica para eliminar efectos de nubosidad y aplicar una máscara que elimine los valores mínimos de cada banda. Con esto se genera la imagen en formato de reflectancia que es de vital importancia para comparar firmas espectrales.

Finalmente, para poder realizar las diferentes razones y matemáticas de bandas se cambian los tamaños de pixel y se apilan todas a las bandas a 15 metros. El pre-procesamiento de las imágenes Landsat-8 también consiste en las correcciones mencionadas a las imágenes Aster y el apilado de las bandas a implementar en una sola imagen, como podemos ver en la **Figura 5**, el sensor Landsat-8 cuenta con 11 bandas en total, pero en este caso, la banda 8 (pancromática) y 9 utilizada para la detección de nubes de elevada altitud, fueron excluidas desde el inicio ya que la información contenida en estas bandas no era relevante para nuestros propósito.

**Figura 5.** Bandas presentes en el sensor Landsat-8 OLI / TIRS (Tomado de USGS, 2017).

Band #	Band name	Wavelength (µm)	Spatial resolution (m)
1	Coastal aerosol	0.43 - 0.45	30
2	Blue	0.45 - 0.51	30
3	Green	0.53 - 0.59	30
4	Red	0.64 - 0.67	30
5	Near Infrared (NIR)	0.85 - 0.88	30
6	Shortwave Infrared (SWIR) 1	1.57 - 1.65	30
7	Shortwave Infrared (SWIR) 2	2.11 - 2.29	30
8	Panchromatic	0.50 - 0.68	15
9	Cirrus	1.36 - 1.38	30
10	Thermal Infrared (TIRS) 1	10.60 - 11.19	100
11	Thermal Infrared (TIRS) 2	11.50 - 12.51	100

El apilado entregó el primer producto sobre el que se realizaron las posteriores operaciones, a continuación se procedió con una modificación necesaria para las bandas correspondientes al TIRS (10 y 11) las cuales cuentan con una resolución espacial de 100 metros, diferente a las primeras 7 bandas que son de 30 metros de resolución, esta modificación se hizo con el fin de realizar correctamente las operaciones mediante vecinos cercanos en el software ArcGis.

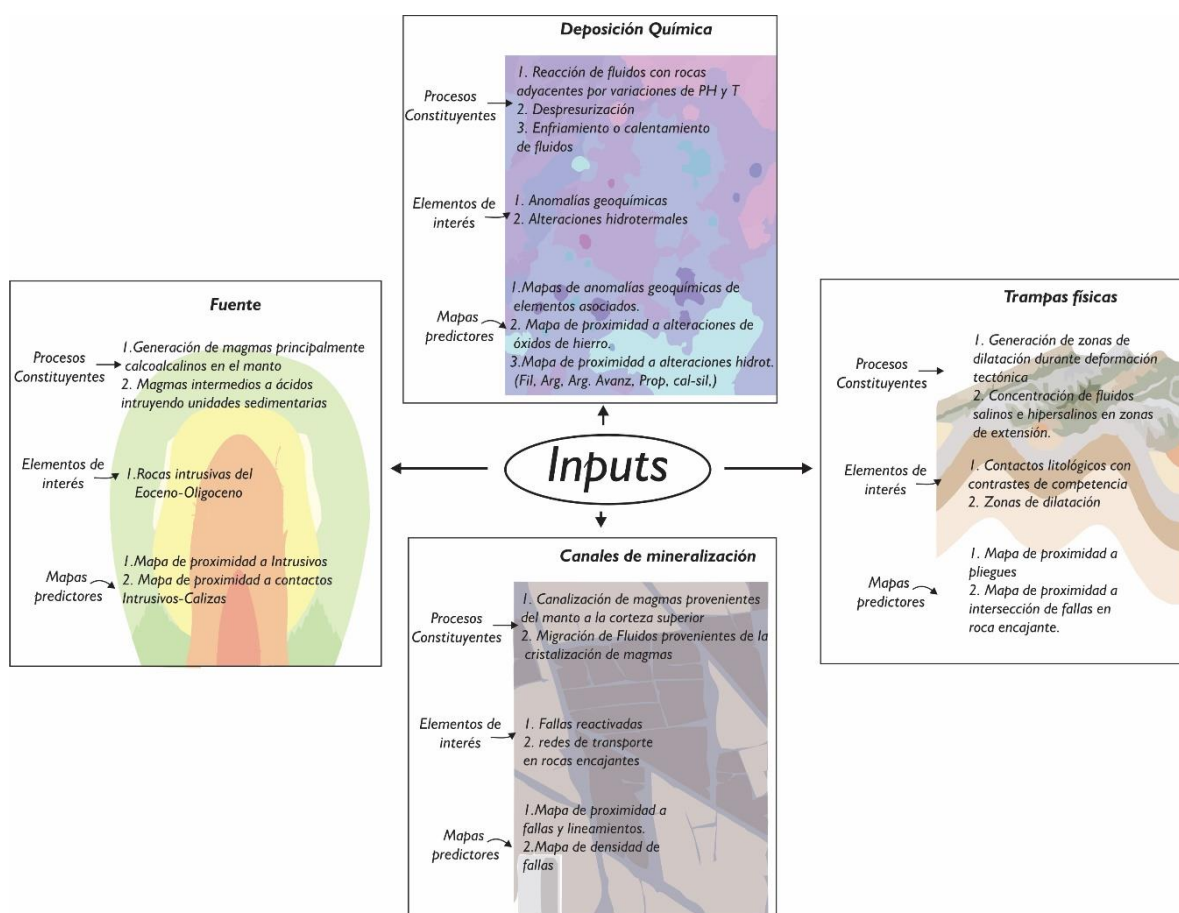
### Fase 2.3: Machine Learning

Para entrenar, generar y validar el modelo de *Machine Learning* utilizando Redes Neuronales Artificiales, se utilizaron como *inputs*: Los mapas generados con las técnicas de procesamiento de imágenes satelitales (Mapa de alteraciones hidrotermales, zonas de óxidos de hierro), geoestadística (mapas de anomalías geoquímicas), mapas de proximidad a fallas-lineamientos, mapa de proximidad a intrusivos, densidad de lineamientos, mapa de distancia a pliegues y mapa de contactos intrusivos-calizas. Los mapas se trabajaron en formato Raster con una resolución espacial de 50 metros.

El tamaño de pixel se selecciona objetivamente con base a la metodología propuesta por (Carranza, 2009), que propone que el tamaño de celda debe garantizar que no haya solapamiento de dos o más yacimientos en una misma celda, este rango de tamaño se genera mediante un cálculo de patrón de puntos realizado en Python. Estos componentes se resumen

en la **Figura 6** y se eligieron en respuesta a una investigación de los procesos de mineralización en los yacimientos de Pórfidos y Skarns en la zona de estudio (Boni *et al.*, 2009; Jones, 1992; Maher, 2010; Perelló, Carlotto, Zárata, *et al.*, 2003; Rivera *et al.*, 2010, 2011; Sillitoe, 2010).

**Figura 6.** Componentes de los sistemas de Pórfidos-Skarn en la zona de estudio, mostrando los procesos constituyentes, elementos de interés y mapas predictores utilizados como inputs en el entrenamiento del algoritmo de Redes Neuronales, esquema (Tomado y modificado de Sun *et al.*, 2019).



Para la extracción de información de las capas utilizadas se implementó la herramienta “*multi values to point*” de ArcMap, donde se juntó toda la data y se extrajo en formato CSV, el cual fue necesario para la escritura del código de programación.

El pre-procesamiento de los datos en formato CSV constó de dos etapas; la primera fue una eliminación de los valores nulos que se encontraban hacia los bordes de la zona por cuestiones de recorte y tamaño de pixel.

Este proceso se realizó en Jupyter Notebook, con las herramientas de la librería “Pandas” de Python; la segunda etapa de pre-procesamiento constó de una normalización de todos los datos de entrada, en esta técnica se cambian los valores de las columnas numéricas de cada set de datos, y se transforman a una escala común (Ba *et al.*, 2016). El rango o escala utilizada fue de (0-1), este proceso se realizó debido a la gran diferencia de rangos entre las capas de entrada, con esto se impide que los inputs con valores muy grandes sesguen el impacto de las de menor rango, generando así una clasificación correcta (Jayalakshmi & Santhakumaran, 2011).

### **Fase 3: Aplicación de las Técnicas**

Los datos pre-procesados se cargaron en los programas correspondientes a las técnicas aplicadas de procesamiento de imágenes satelitales, geoestadística y Machine Learning, es este último se incluyen o los resultados de las dos primeras técnicas.

Para las imágenes satelitales Landsat-8 y ASTER se utilizó el software Arcgis, en el cual se realizan diferentes razones y matemáticas de bandas para identificar las zonas de alteraciones hidrotermales, de óxidos de hierro y también para resaltar lineamientos (Khorram *et al.*, 2012).

Con Microsoft Excel 2016 se generaron los valores máximos, mínimos, de media, mediana, cuartiles, desviación estándar, P-valor y A-valor necesarios para los análisis clúster de estadística univariable y multivariable a utilizar en los diagramas de caja, matrices de Correlación Lineal de Pearson, dendogramas y diagramas de dispersión. Todos estos análisis son necesarios para descubrir zonas de anomalías geoquímicas.

Con los resultados de las técnicas de procesamiento de imágenes satelitales y Geoestadística ya están todos los inputs necesarios para el entrenamiento, validación y aplicación de las redes neuronales artificiales.

En esta etapa se realiza el entrenamiento, validación y generación del código de Machine Learning utilizando Redes Neuronales Artificiales, se aplica el algoritmo de aprendizaje supervisado “*Multi-Layer Perceptron Classifier*” (MLPClassifier) de la librería Scikitlearn en el programa Jupyter de Python, el cual entrena utilizando el método de cálculo de gradiente Backpropagation (Rojas, 1996).

#### **Fase 4: Integración y Comparación de técnicas aplicadas**

En esta fase se compilan y comparan todos los resultados generados con las técnicas de procesamiento de imágenes satelitales, Redes Neuronales y análisis geoestadísticos, se analizan los pros y contras de las técnicas, así como su correlación.

#### **Fase 5: Interpretación de los datos y análisis de resultados**

Esta etapa consta de la interpretación del mapa final, la efectividad de las redes neuronales en la prospectividad minera y la relación o de las diferentes capas de entrada con la ocurrencia de yacimientos de tipo Pórfido-Skarn de la Franja Metalogenética XV.

#### **Fase 6: Redacción del libro de proyecto de grado**

La última etapa la cual se realizará a lo largo de los 6 meses del proyecto consiste en documentar y organizar los resultados obtenidos con las diferentes técnicas durante el desarrollo del proyecto en un informe detallada, es este se especificarán las conclusiones con respecto a la prospección geológica de la franja metalogenética xv.

## **4. Marco Geológico**

### **4.1 Marco Estructural**

La franja metalogenética XV se encuentran de manera generalizada en 4 dominios estructurales dentro del cinturón Andahuaylas-Yauri al sureste de la deflexión de Abancay

(Marocco, 1978). La mayor parte de la franja ubicada en la cordillera occidental y en menor medida en el dominio del borde sur de la Cordillera Oriental y los altiplanos occidental y oriental (Perelló, Carlotto, Zárate, et al., 2003). En estudios más detallados y recientes (Carlotto *et al.*, 2005) define 5 diferentes dominios estructurales en los Andes de sur del Perú: Borde Nor-Este de la Cordillera Occidental que incluye el Bloque Anta, el Altiplano Occidental o Bloque Cusco-Sicuani, el Umbral Cusco-Puno, el Altiplano Oriental o Cuenca Putina y el Borde Sur Oeste de la Cordillera Oriental. En la franja se encuentran 4 de estos dominios; Abaracando la mayor área de la zona se encuentra el dominio del Borde Noreste de la Cordillera Occidental que corresponde a una zona parcialmente plana, con una altitud que varía entre 4000 y 4300 m (Carlotto, 2006). Este terreno se halla cortado por la fuerte incisión de los ríos como el Apurímac que crean profundos valles que alcanzan entre 3000 y 2000 m, aquí las series sedimentarias que son principalmente unidades mesozoicas que tuvieron una sedimentación siliciclástica y carbonatada entre el Jurásico y el cretácico están deformadas por sistemas de pliegues cilíndricos de dirección E-O a ONO-ESE con plano axial ligeramente inclinado al SO (Carlotto, 1998). También se encuentran pliegues que se extienden varios kilómetros que en muchas ocasiones se encuentran desplazados débilmente por fallas inversas (Pecho, 1981).

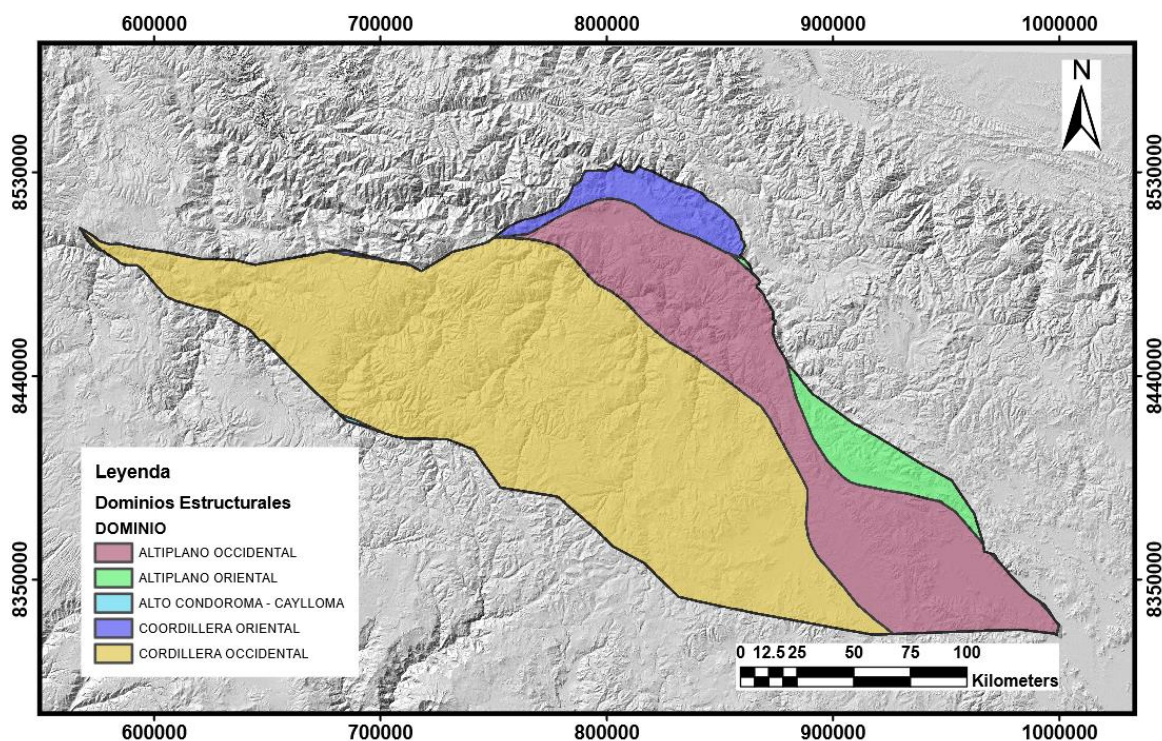
Para este dominio en la zona de estudio los límites están marcados por un sistema de cabalgamientos, inicialmente de dirección E-O (Abancay-Limatambo-Zurite) y posteriormente NO-SE (Huanquite-Accha) los cuales hacen parte del sistema Cusco-Lagunillas-Mañazo (Carlotto, 1998; Carlotto *et al.*, 2002).

La zona noreste del área de estudio se encuentra influenciada principalmente por el dominio estructural del Altiplano Occidental o Bloque Cusco-Sicuani, el cual se encuentra separado, al sur, de la Cordillera Occidental y del Bloque Anta por el cabalgamiento Abancay-Limatambo-Zurite y por la falla Acomayo y al NE debido a antiguos accidentes como la falla Uchuyqosqo de un alto estructural o Umbral Cusco-Puno (Perelló *et al.*, 2003), esta falla

evidencia reactivaciones recientes facilitando el emplazamiento de volcanes de composición shoshonítica del Plio-cuaternario (Perelló *et al.*, 2003).

El tercer dominio estructural presente en la zona estudiada es el del Alto estructural o Umbral Cusco-Puno, definido por Perelló *et al.* (2006) como un estrecho corredor estructural de pocos kilómetros de ancho el cual va desde Sicuani a Urubamba y de Limatambo a Curahuasi, estos dos segmentos se encuentran desplazados por la antigua falla de Patacancha y Tamburco (Carlotto, 1998; Carlotto *et al.*, 2006), En el sector de Sicuani el Umbral Cusco-Puno separa el Altiplano Occidental del Altiplano Oriental o cuenca Putina, en tanto que en Cusco separa el Altiplano Occidental del Borde Sur Oeste de la Cordillera Oriental a través de la falla Urcos-Sicuani-Ayaviri (Carlier *et al.*, 2005).

**Figura 7.** Mapa simplificado de dominios Estructurales en la región de Cusco, Abancay y Sucuani, indicando la ubicación de la zona de estudio (Tomado y modificado de Carlotto, 2006).



## 4.2 Estratigrafía

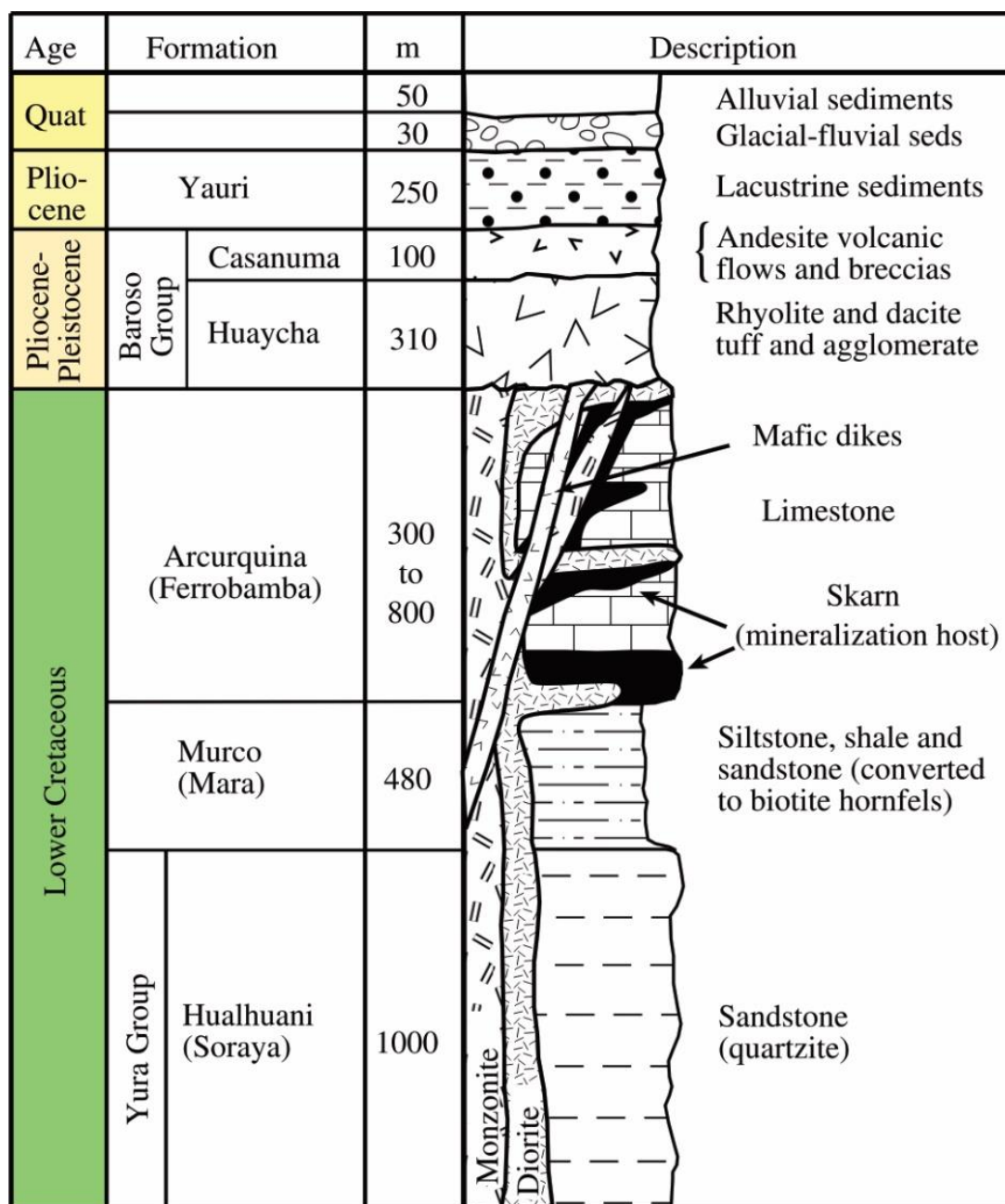
La estratigrafía de la franja consiste en rocas sedimentarias clásticas del Grupo Yura

(Jurásico-Cretácico), la Formación Mara (Aptiano) y calizas Ferrobamba del Albiano Turoniano (Cerpa *et al.*, 2004).

Al norte del sistema de fallas Cusco-Lagunillas-Mañazo, destaca la presencia de rocas sedimentarias que incluyen las capas rojas del Grupo San Jerónimo de edad Eoceno inferior-Oligoceno inferior y las formaciones Punacancha y Paruro, del Oligoceno superior-Mioceno, a su vez, estas unidades se encuentran intruidas por el Batolito Andahuaylas- Yauri de composición calco alcalina y de edad Eoceno-Oligoceno inferior (~48-30 Ma; (Carlotto, 1998).

El batolito incluye una primera etapa de rocas máficas caracterizadas por gabros *cumulat* y dioritas emplazadas entre 48 y 43 Ma, las que están intruidas a su vez (segunda etapa) por pulsos de granodiorita y cuarzo monzodiorita entre ~42 y ~30 Ma (Carlotto, 1998), lo que indica un fuerte levantamiento en el Eoceno medio-Oligoceno inferior. Contemporáneamente a esta segunda etapa, se depositan rocas volcánicas de la Formación Anta (Eoceno medio-Oligoceno inferior) constituidas por más de 1000 m de flujos de lavas de andesitas y flujos piroclásticos de dacitas, intercalados con conglomerados volcano-clásticos (Carlotto, 1998; Victor Carlotto *et al.*, 2005).

**Figura 8.** Columna estratigráfica del distrito Tintaya, mostrando los pulsos de dioritas, monzonitas y diques máficos emplazados en dos etapas. Esta columna se tomó como la tipo (Tomado y modificado de Maher, 2010).



#### 4.2.1 Grupo Yura

Este grupo está constituido por más de 2,200 metros de espesor de rocas sedimentarias marinas (Carlotto *et al.*, 2009). Ha sido dividido en tres formaciones, la más antigua presenta una secuencia de 700 metros de lutitas negras bituminosas con bancos potentes de calizas negras, a la que se le denomina formación Piste.

Descansando sobre la anterior, se encuentran areniscas, lutitas y calizas negras lenticulares con un espesor de 800 metros, a las que se les asigna el nombre de formación Chuquibambilla.

Por último, la parte superior del grupo está constituida por cuarcitas y areniscas cuarzosas con más de 700 metros de espesor, a las que les asigna el nombre de formación Soraya (Instituto Geológico Minero y Metalúrgico, 2010).

#### ***4.2.2 Formación Soraya***

El nombre asignado a esta formación proviene del pueblo de Soraya, ubicado en la parte alta de la margen derecha del cañón del río Chalhuanca; donde la secuencia consiste principalmente en areniscas cuarzosas y cuarcitas que forman prominentes crestas diferenciándolas de otras formaciones. La Formación Soraya está constituida por una secuencia monótona de cuarcitas y areniscas cuarcíticas de grano fino a medio, aunque existen algunas con granos gruesos de cuarzo. El color varía de gris blanquecino en la base a rosado en el tope; presenta estratificación en bancos medianos a gruesos (Instituto Geológico Minero y Metalúrgico, 2010).

#### ***4.2.3 Formación Mara***

Esta formación fue definida por Jenks (1951) al referirse a unos afloramientos rojizos que se encuentran en los alrededores del pueblo de Mara. En los cuadrángulos estudiados de Santo Tomás, Antabamba y Chalhuanca, la Formación Mara consiste de capas de areniscas lutáceas y lutitas de color rojizo; por su amplia extensión constituye un buen nivel guía para la interpretación estratigráfica y estructural.

Al sur de Chuquibambilla afloran pequeños remanentes de esta formación en los núcleos de varios sinclinales. En la Formación Mara se puede distinguir tres miembros: El

inferior se caracteriza por la predominancia de areniscas, el intermedio es lutáceo con algunas intercalaciones de areniscas y conglomerados con clastos de cuarcita, y el superior está constituido por areniscas y lutitas abigarradas y termina hacia el tope, en algunos lugares con calizas amarillentas. El color predominante de esta formación es rojo a marrón rojizo (Instituto Geológico Minero y Metalúrgico, 2010).

#### **4.2.4 Formación Ferrobamba**

La Formación Ferrobamba es una secuencia monótona de calizas negras a gris oscuras, aunque en ciertos niveles presentan bancos calcáreos de color amarillento. Las calizas son masivas, bastante compactas, estratificadas en bancos de 0,30 a 2 metros. En el tope se observan generalmente calizas arenosas de color gris claro con tintes rojizos y en la base, niveles de lutitas carbonosas.

Asimismo, contiene nódulos de chert de forma alargada de más de 0,15 metros de longitud. Gran parte de la Formación Ferrobamba ha sido intruída por cuerpos ígneos de diferente naturaleza, siendo esta la roca huésped de los diferentes yacimientos tipo skarn de cobre, plata, zinc, hierro, etc. En el cuadrángulo de Antabamba, pequeños afloramientos de la Formación Ferrobamba se observan en las partes altas de Mollebamba, Antabamba y Huaquirca. (Instituto Geológico Minero y Metalúrgico, 2010).

## **5. Marco Teórico**

### **5.1 Caracterización Estadística**

La geoestadística es una rama de la estadística aplicada cuya especialización es la modelación y análisis de la variabilidad espacial en ciencias de la tierra. Esta tiene como objeto de estudio el análisis y predicción de fenómenos en espacio y/o tiempo como la ley de metales, concentraciones de un contaminante, etc (Diaz, 2002; Howarth, 1983) En esta sección se

definirá un marco de referencia conceptual que es significativo para el desarrollo de los análisis estadísticos.

### 5.1.1 Test de Normalidad de Anderson-Darling

El test de normalidad de Anderson-Darling se basa en comprobar la hipótesis que  $n$  independientes variables aleatorias distribuidas idénticamente, tienen una función de distribución continua  $F(x)$  (Anderson & Darling, 1930). Esta prueba asigna pesos mayores a la diferencia que hay entre datos empíricos y los teóricos, el estadístico de prueba está dado por la siguiente ecuación:

$$A = n \int_{-\infty}^{\infty} [F_n(x) - \hat{F}(x)]^2 \psi(x) \hat{f}(x) dx \quad (0)$$

Donde  $\hat{F}$  es la distribución que se está probando, para los análisis geo-estadísticos en particular se necesita probar si esta  $F$  teórica es de distribución normal (Olivares, 2017). Esta distribución de 3 integrales se puede resumir en la **Ecuación 1**:

$$A = - \frac{\sum_{i=1}^n (2i-1) (\ln \hat{F}(X_i) + \ln [1 - \hat{F}(X_{n+1-i})])}{n} - n \quad (1)$$

Donde  $n$  es la cantidad de observaciones y  $\hat{F}$  es la distribución normal que se está probando, se hace una sumatoria del recorrido de la función empírica de datos primero ordenándolos de menor a mayor y en la segunda parte las observaciones ordenadas al revés. Para definir si los datos tienen una distribución normal, después de aplicar la fórmula (2) se debe comparar el resultado con el P valor, este valor es el estadístico de ajuste de una función normal, la cual se calcula normalmente para un nivel de confianza del 95% y para un número de datos  $n \geq 5$  (Olivares, 2017; Anderson & Darling, 1930).

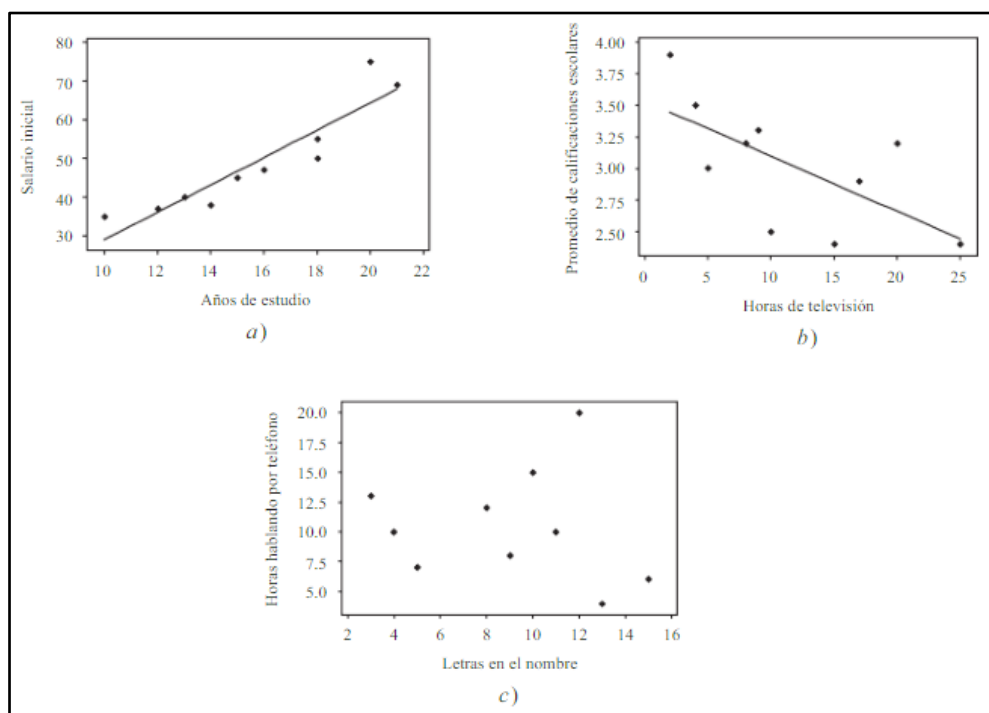
Finalmente al comparar los resultados, se necesita definir si la hipótesis es nula, esta se cumple cuando el P valor  $\geq$  A valor (Test de normalidad), es decir, la función SI cumple con una distribución normal o por el contrario es una hipótesis alterna si A valor  $>$  P valor (Anderson

& Darling, 1930).

### 5.1.2 Diagramas de Dispersión (Scatter Plots)

Los gráficos de dispersión (también llamados diagramas x-y) son una de las técnicas más poderosas y ampliamente utilizadas para la visualización de datos de exploración (Isaaks & Srivastava, 1989; Keim et al., 2010) Los gráficos de dispersión tradicionales representan cada objeto de un conjunto de datos con un punto (o cualquier otra marca), el cual se encuentra en dos dimensiones (X y Y) ortogonales y continuas (Howarth, 1983; Sarikaya & Gleicher, 2018).

**Figura 9.** Diagrama de dispersión tradicional. a) Correlación positiva, b) correlación negativa y c) ninguna correlación (Tomado de Spiegel, 2009)



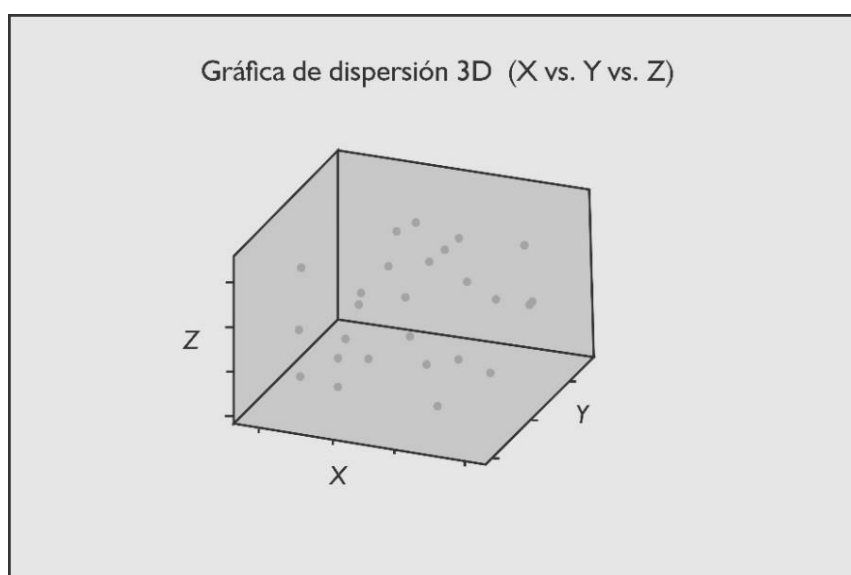
Spiegel (2009) afirma que estos diagramas son una herramienta útil para determinar la correlación o grado de relación de las variables.

Si en un gráfico de dispersión todos los datos tienden a encontrarse cerca de una línea recta se le llama *correlación lineal*, si “Y” tiende a aumentar junto con “X” se conoce como

*correlación positiva o directa* y por el contrario si “Y” tiende a disminuir mientras “X” aumenta se denomina *correlación negativa o inversa*. Por último si no se evidencia ningún tipo de relación entre las variables, se dice que están *descorrelacionadas* (Piringer *et al.*, 2004).

Los diagramas de dispersión 3D añaden un dimensión adicional a la de los convencionales (x, y), con esta nueva dimensión se reducen problemas como la superposición de puntos, también los diagramas 3d ayudan a agrupar y representar un volumen de datos mayor, facilitando una representación eficiente del total de muestras (Piringer *et al.*, 2004).

**Figura 10.** *Diagrama de dispersión 3D.*



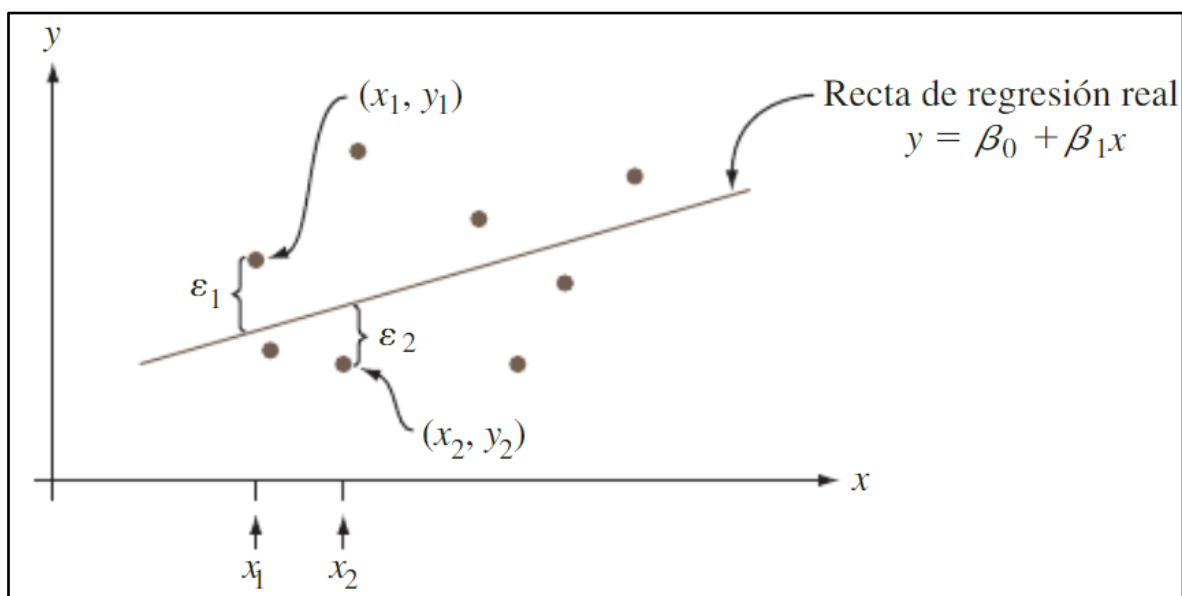
### 5.1.3 Regresión Lineal Simple

Una correspondencia lineal es la relación matemática determinística más simple entre dos variables (x,y) para las cuales  $y=mx+b$  determina una línea recta que tendrá como pendiente  $m$  y una intersección en  $y$  igual a  $b$ . A modo general se tienen dos variables  $x$  y  $y$ , siendo la primera la variable independiente, pronosticadora o explicativa y la segunda será la variable aleatoria, la cual es dependiente o de respuesta (Devore, 2016; Isaaks & Srivastava, 1989). En el modelo de regresión lineal simple se considera la ecuación modelo (**Ecuación 2**) definida como:

$$Y = b + mx + \epsilon \tag{2}$$

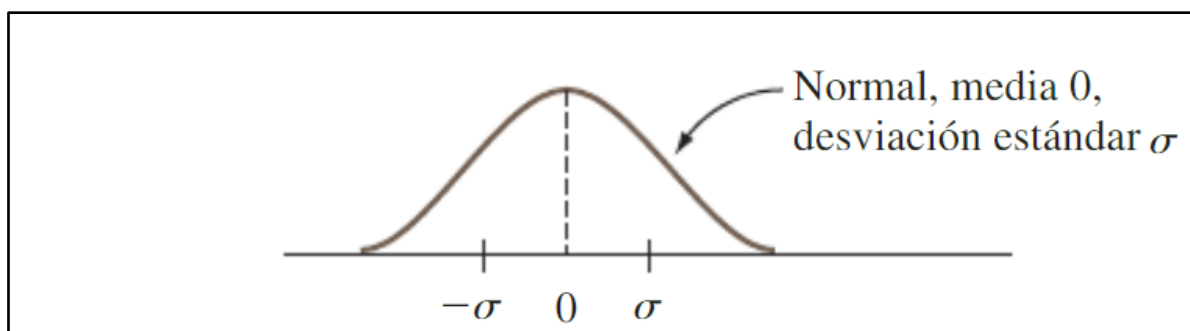
En esta ecuación la variable  $\epsilon$  se conoce como desviación aleatoria o término de error aleatorio, corresponde al error aleatorio el cual permite a  $(x, y)$  quedar fuera de la recta de regresión, por encima de la recta (cuando  $\epsilon > 0$ ) o por debajo (cuando  $\epsilon < 0$ ), por lo tanto si  $\epsilon$  no existe cualquier par observado  $(x, y)$  coincidiría en un punto justo sobre la recta  $y = mx + b$  (Devore, 2016).

**Figura 11.** Puntos correspondientes a observaciones del modelo de regresión lineal simple, se observa la recta de regresión y las distancias a puntos fuera de ella (Tomado de Devore, 2016)



Al realizar el modelo de regresión lineal simple, se debe analizar el parámetro de varianza  $\sigma$ , el cual mide el tamaño de la desviación típica de la recta de regresión, es decir cuan alejados están los datos de la media aritmética, cuando  $\sigma$  es pequeña los puntos estarán más cercanos a la recta y por el contrario si  $\sigma$  tiene valores grandes (Devore, 2016; Howarth, 1983; Mariethoz & Caers, 2015).

**Figura 12.** Distribución de  $\epsilon$ , donde 0 equivale a la media aritmética de los datos (Tomado de Devore, 2016)



### 5.1.4 Regresión Lineal Múltiple

Los análisis de regresión lineal múltiple son una herramienta que sirve para analizar la relación entre un conjunto de variables independientes con una sola variable dependiente, estos análisis están dentro de los análisis estadísticos más utilizados en las ciencias del comportamiento (Aiken *et al.*, 2009). Una forma general de la ecuación de regresión lineal (**Ecuación 3**) escrita para un caso individual  $i$ , esta dada por:

$$\hat{Y}_i = b_0 + b_1X_{i1} + b_2X_{i2} + \dots + b_jX_{ij} + \dots + b_pX_{ip}. \quad (3)$$

Donde  $X_{i1}, X_{i2}, \dots, X_{ip}$  son los valores del caso  $i$  en los predictores  $j= 1, 2, \dots$  hasta  $p$ ,  $b_1, b_2, \dots, b_p$  son los coeficientes de regresión parciales y  $b_0$  es la intercepción de la regresión.

Aiken *et al.* (2009) plantea 3 usos para los análisis de regresión múltiple; el primero uso es para descripción, ya que ayudan a resumir de una manera sencilla la relación de un conjunto de predictores con un criterio en un solo momento en el tiempo. El segundo uso es el de predecir basándose en características previamente medidas y el último es el de pruebas teóricas en la que ideas derivadas de búsquedas anteriores se traducen en hipótesis que son probadas utilizando análisis de regresión múltiple (Aiken *et al.*, 2009; Howarth, 1983)

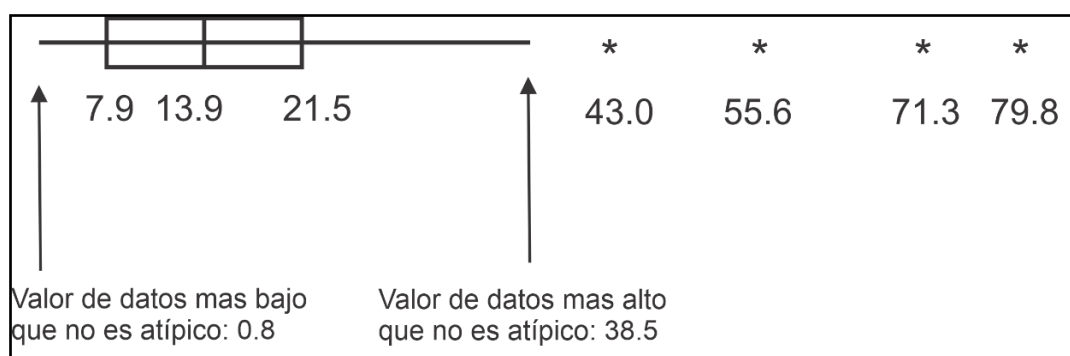
### 5.1.5 Valores Atípicos (Outliers)

Se definen como los valores que al analizar los datos se encuentran muy alejados de la gran mayoría de los demás, dentro de un mismo conjunto de datos. Son significativos porque pueden afectar grandemente los valores de datos estadísticos como la media y desviación estándar que son importantes a la hora de interpretar un conjunto de muestras (Triola, 2018; Isaaks & Srivastava, 1989).

Triola (2018) plantea que para identificar y representar los valores atípicos se suelen utilizar las gráficas de caja modificadas, estas se construyen siguiendo los siguientes pasos:

- i. Definir los cuartiles  $Q_1$ ,  $Q_2$  y  $Q_3$ .
- ii. Determinar el rango intercuartil (RIQ), este se halla  $Q_3 - Q_1$ .
- iii. Multiplicar  $RIQ * 1.5$
- iv. Graficar el diagrama, y los datos atípicos serán los que estén por arriba o por debajo del rango  $Q_1 - Q_3$  en una cantidad superior a  $RIQ * 1.5$ .

**Figura 13.** Diagrama de caja modificado. Mostrando un ejemplo de las velocidades de datos en aeropuertos para erizon en Mbps, los valores atípicos son los que están por fuera del diagrama (Tomado de Triola, 2018)



### 5.1.6 Coeficiente de Correlación Lineal de Pearson

Este término se define en términos de la covarianza de las variables X y Y, esta

covarianza es una medida de la forma en la que las variables independientes varían conjuntamente y esta correlación está dada por la **Ecuación 4** (Diaz, 2002; Fallas, 2012).

$$\rho_{x,y} = \frac{\text{cov}(X,Y)}{\sigma_x \sigma_Y} = \frac{E[(X - \mu_x)(Y - \mu_y)]}{\sigma_x \sigma_Y} \quad (4)$$

Donde X y Y son las variables aleatorias,  $\mu_x$  y  $\mu_y$  son las medias de X y Y y varianzas  $\sigma_x^2$  y  $\sigma_y^2$ , respectivamente (Fallas, 2012; Howarth, 1983; Isaaks & Srivastava, 1989) Para determinar esta correlación entre dos variables es importante entender el estadístico “r”, el cual es el estimador de  $\rho$  y su fórmula (**Ecuación 5**) expresada de la manera más simple es:

$$r = \frac{1}{n-1} \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{s_X} \right) \left( \frac{Y_i - \bar{Y}}{s_Y} \right) \quad (5)$$

Donde  $\left( \frac{X_i - \bar{X}}{s_X} \right) \left( \frac{Y_i - \bar{Y}}{s_Y} \right)$  son los valores estandarizados de X y Y respectivamente,

$\bar{X}$  la media de X

$\bar{Y}$  la media de Y

$s_X$  la desviación estándar de X

$s_Y$  la desviación estándar de Y

Según Fallas (2012) el estadístico “r” es importante porque posee las siguientes características:

- 1) Mide la dirección e intensidad de la correlación lineal, esto lo hace sin despreciar la posibilidad de existencia de una correlación de otro tipo.
- 2) La dirección de la asociación es indicada por el signo del coeficiente de correlación, es decir, un valor de “r” negativo indica una correlación negativa en la que valores grandes de Y se asocian con valores pequeños de X y viceversa. Por el contrario

si es “r” un valor positivo se conoce como relación directa o positiva en la que valores grandes de Y se asocian a valores grandes de X.

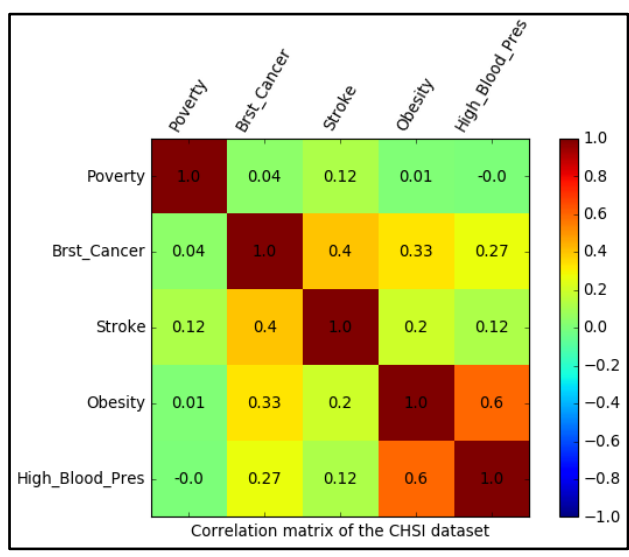
3) El valor de “r” varía entre 1 y -1, entre más cercano a 1 mayor será la asociación, siendo 1 el equivalente a una relación perfecta. Si el valor es 0 indica que no hay correlación lineal (Fallas, 2012; Mariethoz & Caers, 2015).

**5.1.7 Matriz de Correlación Lineal**

Zadeh (2019) la define como una herramienta para mostrar la correlación de dos variables numéricas a fin de tener una visión de su relación, se usa normalmente para bases de datos con muchos atributos. Uno de los métodos más populares para encontrar esta relación es el que usa el Coeficiente de Correlación de Pearson.

La matriz de correlación es simétrica, donde todos los elementos de la diagonal son igual a +1, ya que la relación de un elemento consigo mismo es del 100%. Es importante enfatizar que la matriz da una visión de correlación pero no es una herramienta confiable para estudiar causalidad (Zadeh, 2019;Howarth, 1983; Isaaks & Srivastava, 1989).

**Figura 14.** Matriz de correlación, ejemplo de la base de datos del CHSI (Tomado de Zadeh, 2019)

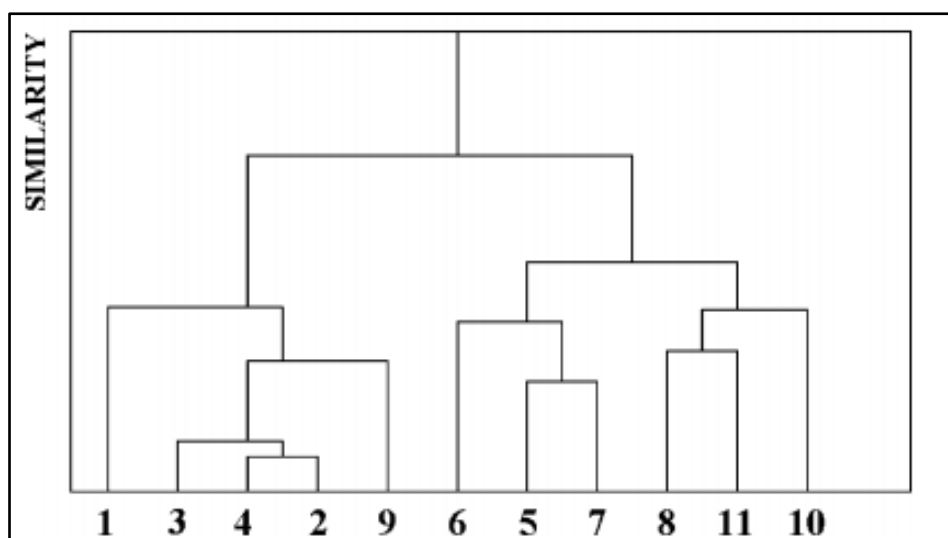


**5.1.8 Dendogramas (Cluster Analysis)**

Las técnicas de agrupamiento son usadas frecuentemente para mostrar e interpretar similitudes entre objetos o variables, las cuales se representan como miembros de una categoría. Los resultados de estos agrupamientos se suelen reportar utilizando los dendogramas (Forina *et al.*, 2002; Howarth, 1983; Isaaks & Srivastava, 1989).

Los dendogramas muestran en el eje Y la similitud de los datos y en el eje X las diferentes observaciones o conjuntos de datos(Forina *et al.*, 2002).

**Figura 15.** Dendograma de un conjunto de datos (Tomado de Forina *et al.*, 2002)



Forina *et al.* (2002) clasifica las técnicas de agrupamiento en 3:

- a) Técnicas visuales: que son basadas generalmente en la observación de componentes principales de los diagramas.
- b) Métodos no jerárquicos: la agrupación no se muestra utilizando los dendogramas
- c) Métodos jerárquicos: existen los métodos aglomerados y divisivos, en esta sección solo se analizará el primero ya que es el que utiliza los dendogramas. Están basados en la similitud que hay entre dos objetos o grupos de objetos y está dado por la **Ecuación 6**:

$$S_{ij} = 1 - \frac{d_{ij}}{d_{max}} \tag{6}$$

Donde  $d_{ij}$  es la distancia entre los dos objetos analizados y  $d_{max}$  es la distancia máxima entre dos objetos en la base de datos. Los dos objetos más alejados tienen una relación de 0.

Los dendogramas para este método se grafican empezando por los datos con menor semejanza los cuales están a la máxima distancia y se va dividiendo en grupos cada vez más relacionados hasta llegar, si es posible, a una similitud del 100% o total (Forina *et al.*, 2002; Isaaks & Srivastava, 1989).

## 5.2 *Machine Learning*

Varios estudios demuestran que los algoritmos de aprendizaje automático (AAP), son más precisos que las técnicas estadísticas como el análisis discriminante o la regresión logística, especialmente cuando el espacio de características es complejo o se espera que los conjuntos de datos de entrada tengan distribuciones estadísticas diferentes (Abedi *et al.*, 2012; Brown *et al.*, 2000; Harris *et al.*, 2003; Zuo *et al.*, 2011).

Los AAP tienen el potencial de identificar y modelar las complejas relaciones no lineales entre las ocurrencias minerales y las características evidenciales (Brown *et al.*, 2000).

Estos métodos pueden manejar una gran cantidad de características que podrían ser importantes en los estudios de prospectividad mineral. Sin embargo, aumentar el número de características probatorias de entrada puede conducir a una mayor complejidad y a un mayor número de parámetros del modelo, por lo que el modelo se vuelve susceptible de sobreajuste debido a la “maldición de la dimensionalidad” (Galiano *et al.*, 2015; Galiano & Chica-rivas, 2012).

Esta se refiere una serie de problemas que surgen cuando los datos tratados tienen una gran cantidad de variables fuera del espacio físico descrito generalmente con solo tres dimensiones y se manifiesta de dos maneras: la distancia media entre datos aumenta con el

número de dimensiones y la variabilidad de la distancia disminuye exponencialmente con el número de dimensiones, lo cual es perjudicial para los métodos de aprendizaje automático basados en distancias como vecinos más cercanos o clustering (Bellman & Kalaba, 1965).

A continuación citamos una serie de trabajos realizados mediante diferentes ramas del machine learning aplicados a la geología (Rodríguez-Galiano *et al.*, 2015).

En las últimas décadas se ha desarrollado una gran cantidad de métodos de clasificación (Hastie *et al.*, 2009). Entre las técnicas más utilizadas se encuentran los árboles de decisión (AD) (Breiman *et al.*, 1984), las redes neuronales artificiales (RNA) (Brown *et al.*, 2000; Porwal *et al.*, 2003; Rigol-Sanchez *et al.*, 2003; Rumer, 1986) y máquinas de vectores de soporte (SVM) (Abedi *et al.*, 2012; Zuo *et al.*, 2011).

### **5.2.1 Redes Neuronales Artificiales**

El enfoque más común para desarrollar una clasificación / regresión no paramétrica y no lineal se basa en las redes neuronales artificiales (RNA). Hay muchos tipos diferentes de RNA.

En esta sección se proporciona una breve descripción de una de las RNA más utilizadas: la red neuronal de propagación feed-forward (Rumelhart *et al.*, 1986).

Ha habido muchos intentos de diseñar redes neuronales organizadas por sí mismas. El objetivo es encontrar una poderosa regla de modificación sináptica que permita que una red neuronal conectada arbitrariamente desarrolle una estructura interna que sea apropiada para un dominio de tarea en particular. La tarea se especifica dando el vector de estado deseado de las unidades de salida para cada vector de estado de las unidades de entrada.

Si las unidades de entrada están conectadas directamente a las unidades de salida, es relativamente fácil encontrar reglas de aprendizaje que ajusten iterativamente las fortalezas

relativas de las conexiones para reducir progresivamente la diferencia entre los vectores de salida reales y deseados. El aprendizaje se vuelve más interesante pero más difícil cuando introducimos unidades ocultas cuyos estados reales o deseados no están especificados por la tarea (Rumelhart *et al.*, 1986). En los perceptrones, hay 'analizadores de características' entre la entrada y la salida que no son verdaderas unidades ocultas porque sus conexiones de entrada se fijan a mano, por lo que sus estados están completamente determinados por el vector de entrada: no aprenden representaciones. El procedimiento debe decidir bajo qué circunstancias las unidades ocultas deben estar activas para ayudar a lograr el comportamiento de entrada-salida deseado. Esto equivale a decidir qué deben representar estas unidades. Demostramos que un procedimiento de propósito general y relativamente simple es lo suficientemente poderoso como para construir representaciones internas apropiadas (Rumelhart *et al.*, 1986).

Al igual que en el cerebro, los elementos básicos de procesamiento de una red neuronal artificial son las neuronas (unidades o nodos). En una red neuronal, las unidades se colocan como capas y están conectadas de tal manera que la información fluye unidireccionalmente, desde las unidades de entrada, a través de la unidad o unidades ubicadas en la capa / capas ocultas, hasta las unidades en la capa de salida. Las unidades de entrada distribuyen la señal a las unidades ocultas de la segunda capa. Una neurona básicamente realiza una regresión lineal seguida de una función no lineal,  $f(\cdot)$ . Las neuronas de diferentes capas están interconectadas con los enlaces correspondientes (pesos) (Rodríguez-Galiano *et al.*, 2015).

El modelo estándar RNA multicapa, cuya neurona  $j$  en la capa  $l + 1$  produce

$$x_j^{l+1} = f(\sum_i w_{ij}^l x_i^l + w_{bj}^l) \quad (7)$$

donde  $w_{ij}^l$  son los pesos que conectan la neurona  $i$  en la capa  $l$  con la neurona  $j$  en la capa  $l + 1$ ,  $w_{bj}^l$  es el término de sesgo de la neurona  $j$  en la capa  $l$ , y  $f$  es una función de

activación logística (Rodríguez-Galiano *et al.*, 2015).

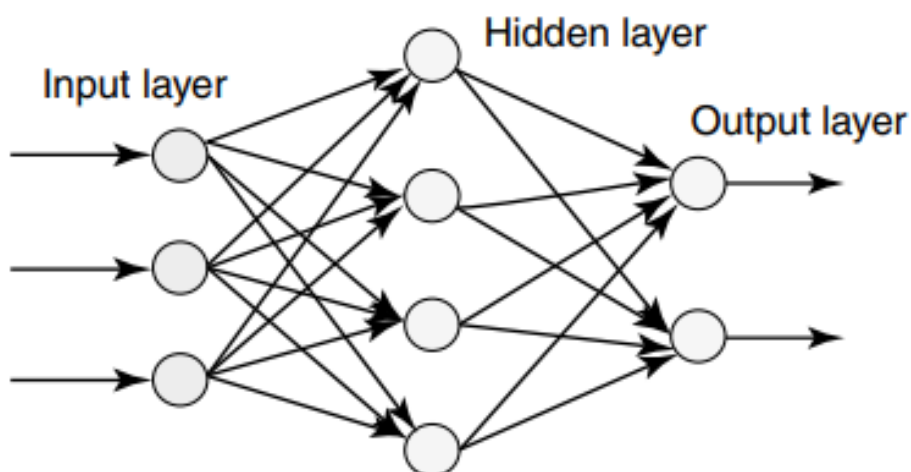
El objetivo del algoritmo es encontrar un conjunto de ponderaciones que garantice que, para cada vector de entrada, el vector resultante de la red sea el mismo o lo suficientemente cercano al vector de salida deseado.

Si hay un conjunto definido y finito de casos de entrada y salida (patrones), el error general en el funcionamiento de la red con un conjunto particular de pesos se puede calcular comparando los vectores de salida reales y deseados para cada patrón, por ejemplo, por el método de mínimos cuadrados. El entrenamiento de una RNA necesita seleccionar una estructura (número de capas ocultas y nodos por capa), la inicialización adecuada de los pesos, la tasa de aprendizaje y los parámetros de regularización para evitar un ajuste excesivo (Rodríguez-Galiano *et al.*, 2015a).

Hay dos categorías principales de arquitecturas de red según el tipo de conexiones entre las neuronas, "redes neuronales de alimentación directa" y "redes neuronales recurrentes". Si no hay "retroalimentación" de las salidas de las neuronas hacia las entradas en toda la red, entonces la red se denomina "red neuronal de alimentación hacia adelante". De lo contrario, si existe tal retroalimentación, es decir, una conexión sináptica desde las salidas hacia las entradas (ya sea sus propias entradas o las entradas de otras neuronas), entonces la red se denomina "red neuronal recurrente" (Sazli, 2006).

Por lo general, las redes neuronales se organizan en forma de "capas". Las redes neuronales de retroalimentación se dividen en dos categorías según el número de capas, ya sea de "una sola capa" o de "múltiples capas" (Sazli, 2006). En la **Figura 16** se ilustra el modelado de una red neuronal multicapa.

**Figura 16.** *Arquitectura de una red neuronal artificial con múltiples capas o layers (Tomado de Abraham, 2005).*



### 5.2.1.1 Algoritmo de Backpropagation

Entre muchos otros algoritmos de aprendizaje, el "algoritmo de backpropagation" es el más popular y el más utilizado para el entrenamiento de redes neuronales de retroalimentación. Es, en esencia, un medio de actualizar los pesos sinápticos de las redes mediante la propagación inversa de un vector de gradiente en el que cada elemento se define como la derivada de una medida de error con respecto a un parámetro. Las señales de error se definen generalmente como la diferencia entre las salidas de red reales y las salidas deseadas. Por lo tanto, se debe disponer de un conjunto de resultados deseados para la formación. Por esa razón, la propagación hacia atrás es una regla de aprendizaje supervisado (Rumelhart *et al.*, 1986; Sazli, 2006).

## 5.3 Imágenes Satelitales

Para hablar de imágenes satelitales es importante tener claro el significado de los Sistemas de Información Geográfica (SIG) y los sensores remotos.

Los SIG son sistemas compuestos por un software, hardware y procedimientos que sirven para capturar, manejar, manipular, analizar, modelar y representar datos georreferenciados, esto con la finalidad de resolver problemas de gestión y planificación en

diferentes áreas de las ciencias (Ortega & Ramos, 2016 ; Ibañez, 2005).

Los sensores remotos son utilizados para el estudio de la superficie terrestre mediante la recopilación y análisis de datos a fin de extraer información de interés sin necesidad de contacto físico alguno, estos sensores son unos de los medios más potentes para adquirir información actualizada en sitios que son de difícil acceso o donde la toma de datos es muy costosa y por lo general estos se encuentran situados en una plataforma como satélites, aviones, etc. (Sholarin & Awange, 2015; Schowengerdt, 2006).

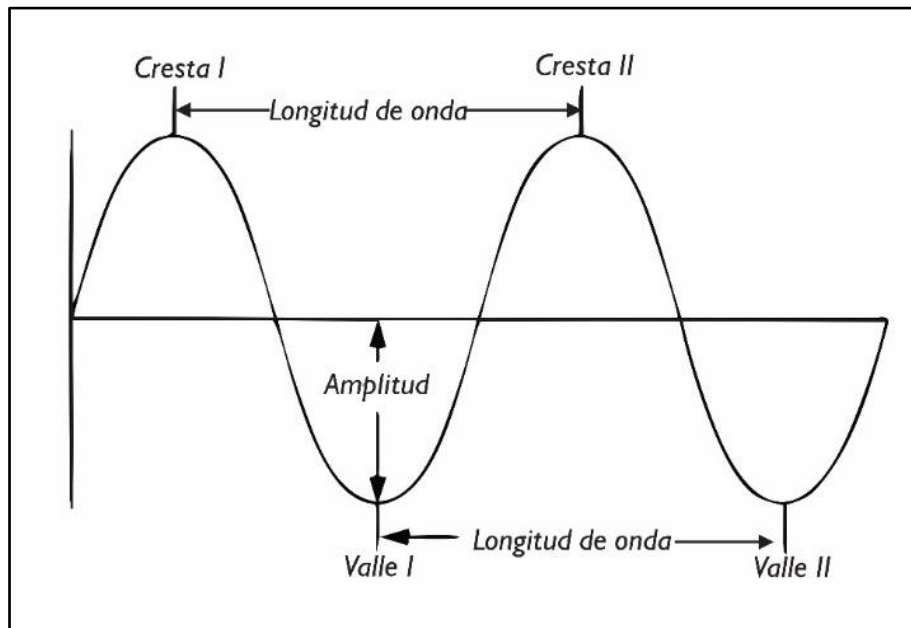
El fundamento principal utilizado en los sensores remotos son los principios de la radiación electromagnética, comprendiendo su interacción con la materia a través de las leyes de radiación (Demirel & Düzgün, 2011; Sholarin & Awange, 2015).

### ***5.3.1 Principios de la Radiación Electromagnética***

La radiación puede ser definida como energía en forma de ondas proveniente de una fuente, todos los objetos tienen cierta cantidad de radiación y estos también tienen la capacidad de reflejarla. Los sensores remotos se basan en la medición de la energía, tanto la emitida como la reflejada de los diferentes objetos o áreas en la superficie terrestre (Demirel & Düzgün, 2011; Sholarin & Awange, 2015).

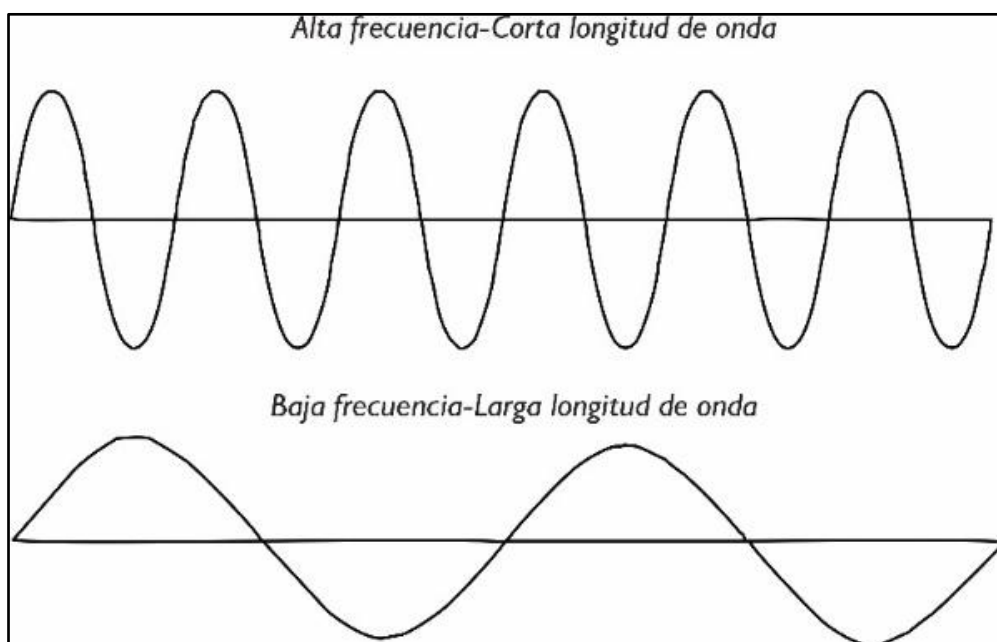
La transferencia de energía de un objeto a otro está gobernado por las ondas electromagnéticas, las cuales tienen 3 características: la primera es la longitud de onda, definida como la distancia entre 2 crestas o valles **Figura 17**, la unidad para longitudes de onda puede estar en diferentes magnitudes como metros, nanómetros o micrómetros, etc (Schowengerdt, 2006).

**Figura 17.** Longitud de onda y amplitud de una onda típica (Modificado de Demirel & Düzgün, 2011)



Otra de las propiedades es la amplitud la cual se define como la altura de la cresta y es una medición de la energía de una onda, las unidades de amplitud normalmente están expresadas en  $\text{watts/m}^2 / \mu\text{m}$ . Por último, la frecuencia es el número de crestas o valles que pasan por un punto fijo de referencia en un cierto periodo de tiempo, la unidad de frecuencia es Hertz, la cual equivale a 1 ciclo/segundo (**Figura 18**) (Demirel & Düzgün, 2011)

**Figura 18.** Frecuencia de ondas (Modificado de Demirel & Düzgün, 2011)



La relación entre la longitud de onda ( $\lambda$ ) y la frecuencia ( $\nu$ ) está dada por:

$$c = \lambda \nu \quad (7)$$

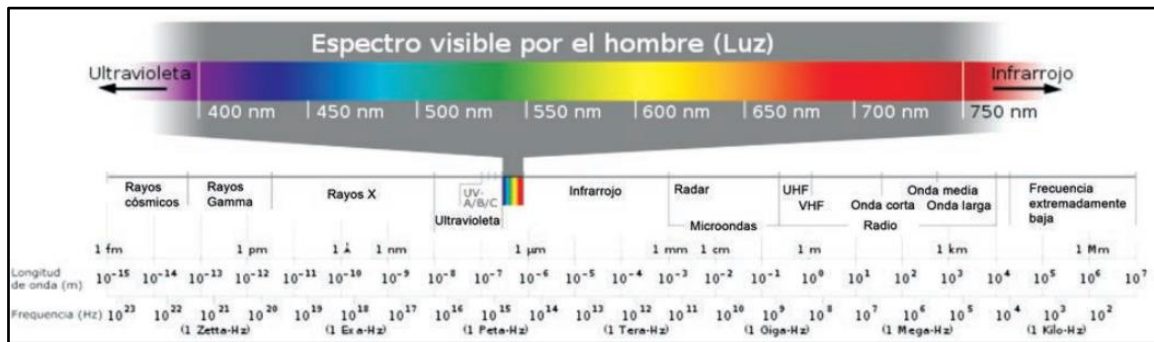
Donde “c” en esta ecuación es la velocidad de la luz, la cual es  $3 \times 10^8$  m/s. Como se observa en la ecuación, la energía electromagnética tiene una velocidad igual a la de la luz y puede ser expresada usando tanto la frecuencia como la longitud de onda, también la longitud de onda y la frecuencia son inversamente proporcionales, implicando que a longitudes de ondas más cortas, la frecuencia es mayor y viceversa. Esta relación viene de los principios de la teoría electromagnética desarrollada por James Clerk Maxwell para explicar cómo se transfiere la energía de un objeto a otro (Demirel & Düzgün, 2011).

### ***5.3.2 Espectro Electromagnético***

El espectro electromagnético se define como el conjunto de todas las frecuencias o número de ciclos de la onda por unidad de tiempo posibles a las que se produce la radiación electromagnética (Ordóñez, 2017). Este espectro formalmente es infinito y continuo con un límite teórico inferior 0 y uno teórico de  $\infty$ .

La clasificación más común del espectro electromagnético establece las siguientes categorías de radiación electromagnética: ondas subradio, ondas radioeléctricas, microondas, rayos T, infrarrojos, luz visible, rayos ultravioleta, rayos X, rayos gamma y rayos cósmicos (**Figura 19**). Dentro de estos espectros, se denomina visible a la región del espectro electromagnético que el ojo humano es capaz de percibir, este se encuentra entre la radiación infrarroja y la radiación ultravioleta (Ordóñez, 2017).

Figura 19. Espectro electromagnético (Tomado de Ordoñez, 2017)

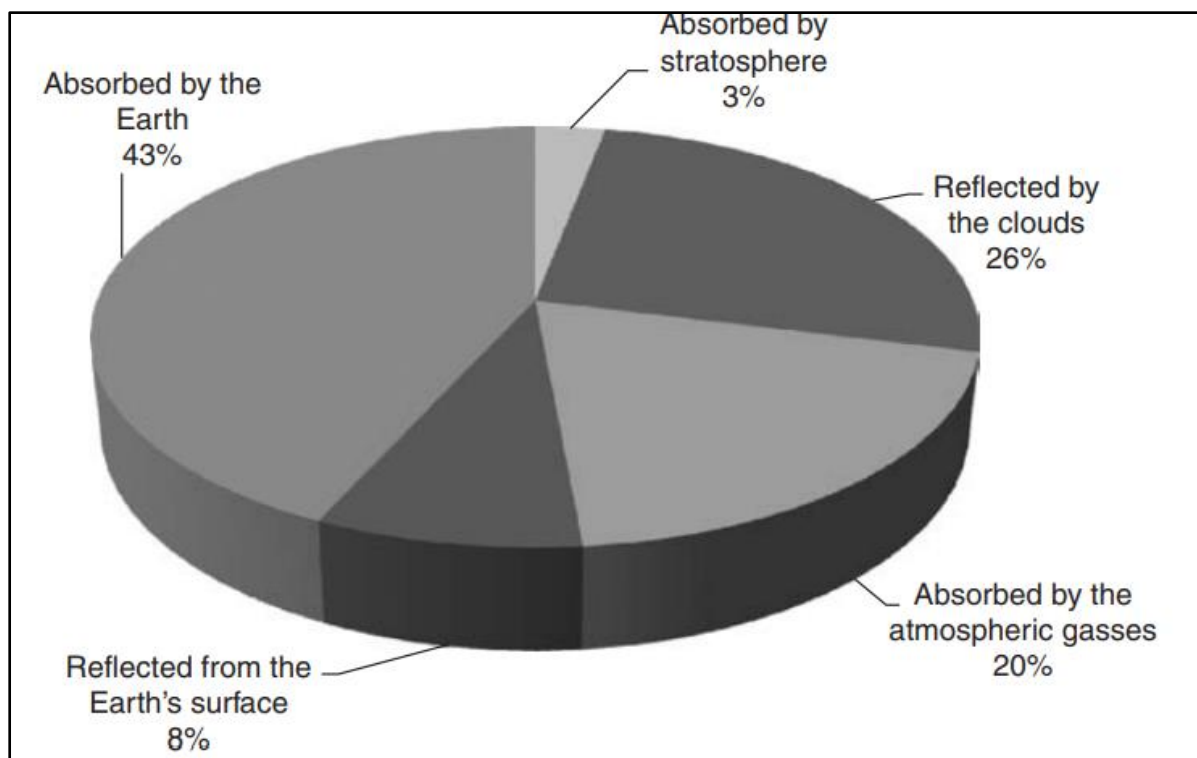


### 5.3.3 Interacción de la Energía Electromagnética con los Objetos en Superficie Terrestre

Cuando la energía electromagnética pasa a través de la atmósfera interactúa con los objetos en la superficie de la tierra. Esta interacción resulta en 3 procesos: reflexión, absorción y/o transmisión (Liang & Wang, 2012).

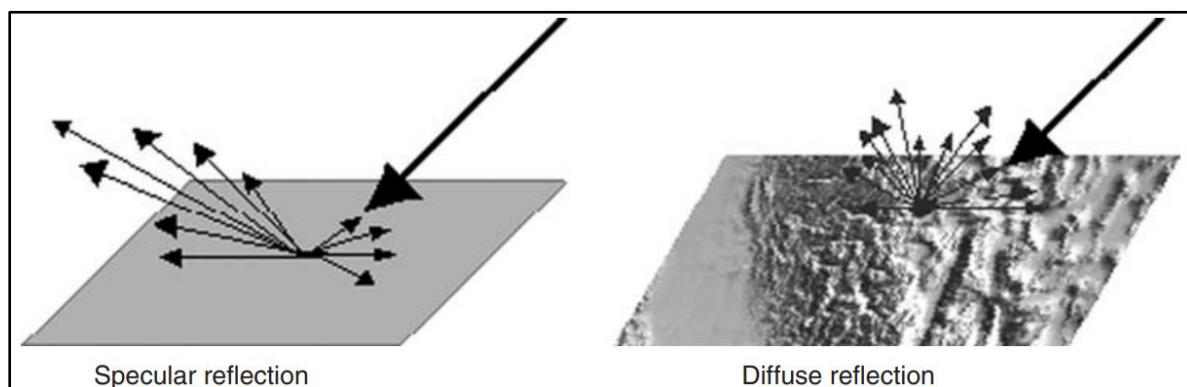
La energía incidente, la cual es la energía interactuando con los objetos en la superficie terrestre es la suma de la energía reflejada, absorbida y transmitida debido a la ley de la conservación de la energía. La absorción es la energía perdida debido a gases en la atmósfera y la transmisión es la energía electromagnética que pasa directamente a través de la ventana atmosférica, la distribución total de la energía solar total entrante se resume en la **Figura 20**.

**Figura 20.** Distribución de la energía solar total entrante con longitudes de onda menores a  $4.5 \mu\text{m}$  (Tomado de Demirel & Düzgün, 2011)



La reflexión está definida como el redireccionamiento de la luz después de entrar en contacto con una superficie opaca, la cantidad de energía reflejada depende de la rugosidad y la longitud de onda de la energía incidente (Schowengerdt, 2006). La rugosidad de la superficie la cual es menor que la longitud de onda resulta en reflexión especular, si la superficie es más suave que la longitud de onda, casi toda la energía incidente es reflejada y el ángulo de incidencia es igual al ángulo de reflexión (**Figura 21**).

Esta ocurre normalmente en superficies las cuales son casi lisas como espejos, metales pulidos y cuerpos de agua de baja energía como lagos. La reflexión que se produce en objetos que tienen una rugosidad superficial mayor que la longitud de onda de la energía incidente se denomina reflexión difusa (Schowengerdt, 2006).

**Figura 21.** Reflexión especular y difusa (Tomado de Demirel & Düzgün, 2011)

La mayoría de los objetos en la tierra tienen ambas reflexiones, la reflexión difusa determina las características espectrales de las superficies de los objetos y la forma en la que se ven los objetos en diferentes longitudes de onda (Schowengerdt, 2006).

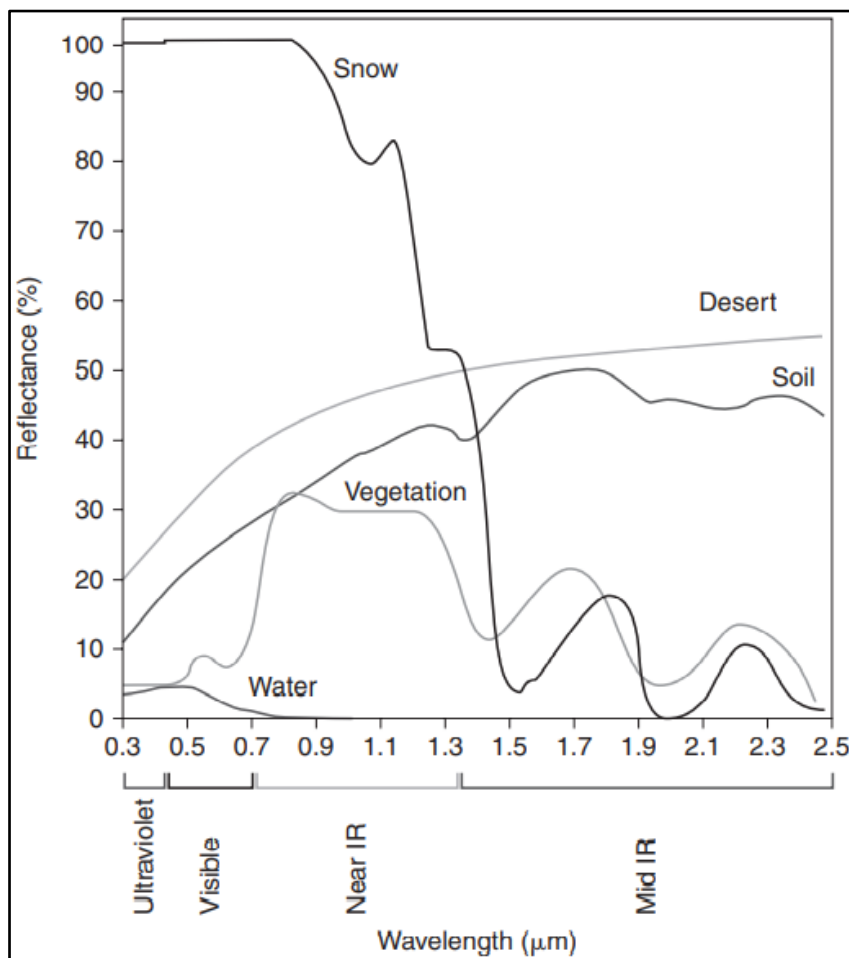
#### 5.3.4 Curvas de Reflectancia Espectral

Las características de reflectancia de las superficies de los objetos  $[\rho(\lambda)]$  para una longitud de onda  $(\lambda)$  dada están medidas por el radio de energía reflejada  $[Er(\lambda)]$  de la superficie del objeto hasta la energía incidente en la superficie del objeto  $[Ei(\lambda)]$  (Demirel & Düzgün, 2011; Liang & Wang, 2012). Estas características medidas definen la reflectancia espectral de una superficie de un objeto, y está dada por la siguiente **Ecuación 9**:

$$\rho(\lambda) = Er(\lambda)/Ei(\lambda) \quad (9)$$

Cuando  $\rho(\lambda)$  se obtiene para varias longitudes de onda también se obtienen las curvas de reflectancia espectral (**Figura 22**). Estas curvas constituyen la firma de reflectancia de las características de la Tierra para que proporcionen valores de longitud de onda adecuados para su uso en aplicaciones de teledetección (Liang & Wang, 2012).

**Figura 22.** Curvas de reflectancia espectral típicas para características comunes en la Tierra (Tomado de Demirel & Düzgün, 2011).



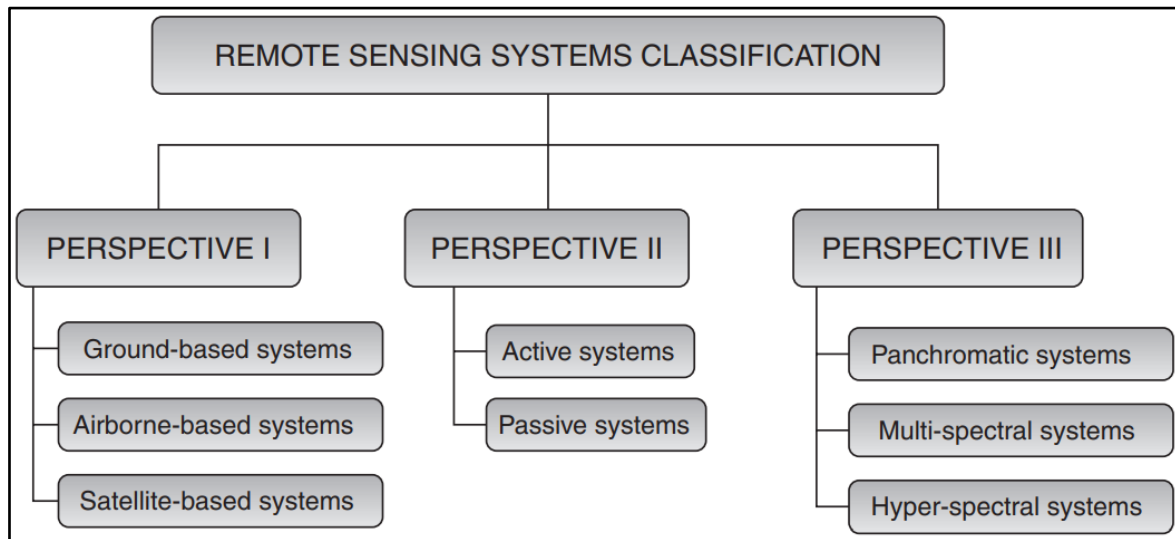
La energía transmitida, emitida, absorbida y reflejada de los objetos en la superficie terrestre como suelos, vegetación, minerales agua etc, en varias longitudes de onda permite la detección de estos utilizando los sensores remotos (Liang & Wang, 2012).

### 5.3.5 Sistemas de Sensores Remotos

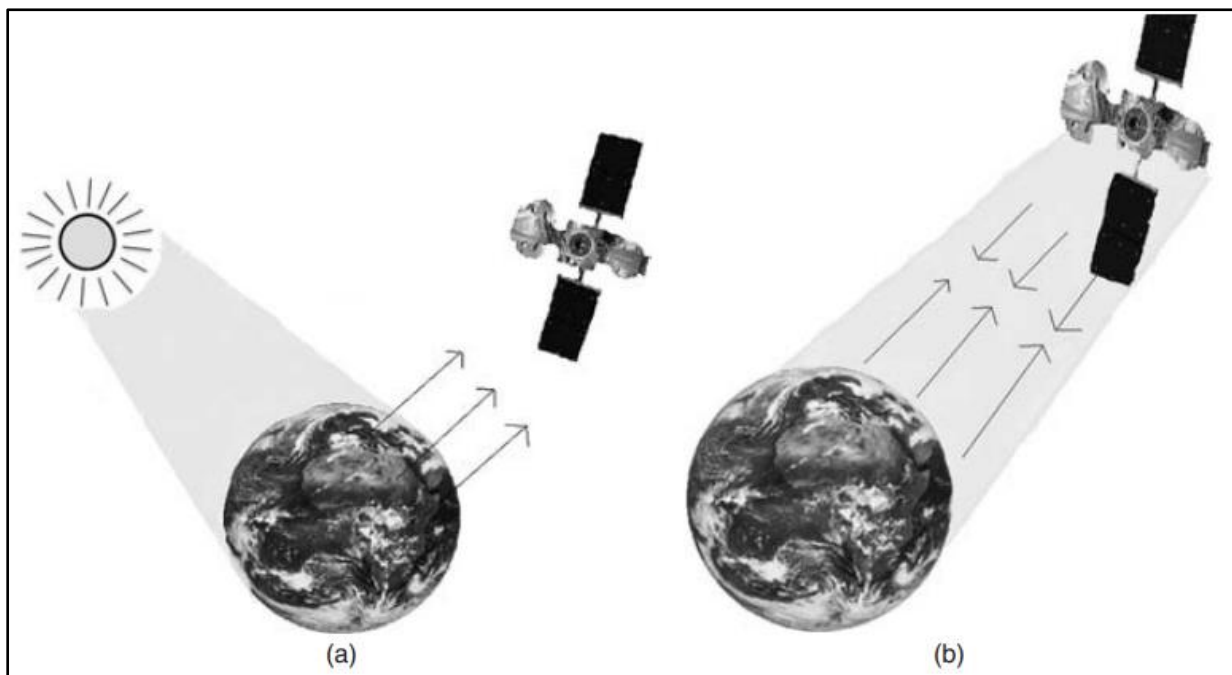
Los sistemas de sensores remotos pueden ser clasificados basados en 3 perspectivas (**Figura 23**) en la primera la plataforma del sensor es el enfoque principal pudiéndose clasificar en sistemas terrestres, aéreos y espaciales. En la perspectiva 2, los sistemas de sensores están agrupados en 2 categorías: pasivos o activos, en la perspectiva 3 los sistemas están clasificados de acuerdo con la habilidad del sensor para registrar reflectancia espectral y resolución espacial; estos sistemas pueden ser pancromáticos, multi-espectrales o hiper-espectrales

(Rodríguez *et al.*, 2015).

**Figura 23.** Clasificación de sistemas de sensores remotos con 3 perspectivas diferentes (Tomado de Demirel & Düzgin, 2011).



La clasificación de los sistemas de sensores remotos en la perspectiva 2 depende de la fuente de energía que irradia la Tierra, los sensores pasivos están diseñados para detectar energía electromagnética que existe naturalmente (**Figura 24**) y la fuente de esta energía es el Sol. La energía del sol es reflejada o absorbida y posteriormente se vuelve a emitir, esto mismo ocurre con las longitudes de onda de infrarrojo térmico (Rodríguez *et al.*, 2015). Por el contrario en los sistemas remotos activos generan su propia fuente de energía, en estos la energía es enviada directamente al objetivo y recibida por sensores para medir la interacción del objeto y la energía (Rodríguez *et al.*, 2015).

**Figura 24.** Sensores pasivos (a) y sensores activos (b) (Tomado de Demirel & Düzgün, 2011).

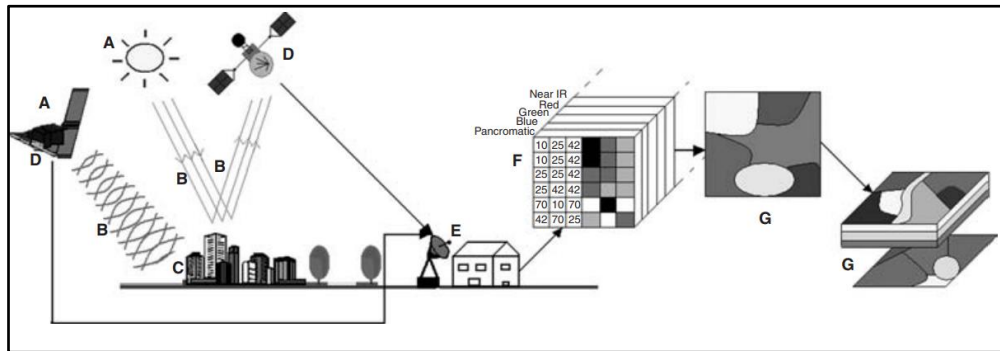
En la perspectiva 3, los sensores remotos están clasificados basado en la resolución espacial y espectral de los sensores, el primer tipo de sensores son llamados pancromáticos/ de alta resolución, donde los sensores electromagnéticos registran niveles de energía que varían entre  $0.5\text{--}0.9\mu\text{m}$  y la resolución espacial, la cual se define como el tamaño más pequeño del objeto que puede ser detectado por el sensor (Rodríguez *et al.*, 2015).

Los sensores multi-espectrales recopilan la información en regiones de longitud de onda separadas llamadas bandas. Por último los sensores hiper-espectrales tiene la capacidad de detectar información de diferentes espectros de una manera detallada al proporcionar la mayor cantidad de bandas con un rango espectral angosto (Arozarena & Pastor, 2016).

Independientemente de las perspectivas de clasificación de los sensores remotos, estos cuentan con siete elementos principales (**Figura 25**), estos son: (A) una fuente de energía, (B) radiación y la atmósfera, (C) interacción con el objetivo, (D) registro de la energía en el sensor, (E) transmisión, recepción y procesamiento de datos, (F) interpretación y análisis, y (G) aplicación, la cual se realiza para revelar nueva información sobre el objeto de interés en la

superficie a partir de la información extraída de las imágenes (Demirel & Düzgün, 2011).

**Figura 25.** Componentes de los sensores remotos (Tomado de Demirel & Düzgün, 2011)

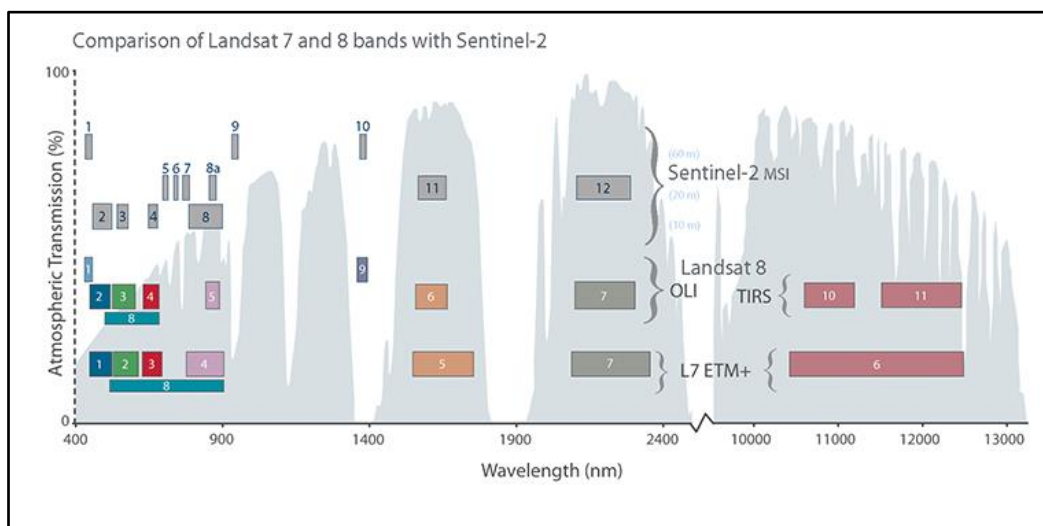


### 5.3.6 Bandas Espectrales

Los sensores remotos trabajan midiendo los datos en múltiples regiones del espectro electromagnético, a este rango de longitudes de onda medidas por un sensor se conoce como banda espectral. La visión humana es un sistema capaz de detectar 3 de estas bandas o longitudes de ondas: rojo, verde y azul (RGB) (Humboldt, 2019).

Este modelo de color RGB está basado en la síntesis aditiva para representar un color mediante la mezcla por adición de los 3 colores de luz primario (Humboldt, 2019).

**Figura 26.** Comparación de las bandas de los distintos sensores que se encuentran a lo largo de las diferentes longitudes de onda (Tomado de USGS, 2014).



Desde el punto de vista de la teledetección se destacan unas series de estas bandas espectrales:

- Espectro visible (0,4 a 0,7  $\mu\text{m}$ ): se trata de la única radiación electromagnética que percibe la vista humana, en estas se distinguen 3 bandas elementales; la azul (0,4 a 0,5  $\mu\text{m}$ ); verde (0,5 a 0,6  $\mu\text{m}$ ) y rojo (0,6 a 0,7  $\mu\text{m}$ ).
- Infrarrojo próximo (0,7 a 1,3  $\mu\text{m}$ ): denominado también infrarrojo reflejado y fotográfico.
- Infrarrojo medio (1,3 a 8  $\mu\text{m}$ ): donde ocurren procesos de emisión de la superficie terrestre y de reflexión de la luz solar.
- Infrarrojo lejano o térmico (8 a 14  $\mu\text{m}$ ): incluye la porción emisiva del espectro terrestre.
- Micro-ondas (> 1mm): es un tipo de energía transparente a la cubierta nubosa (Chuvieco, 1990)

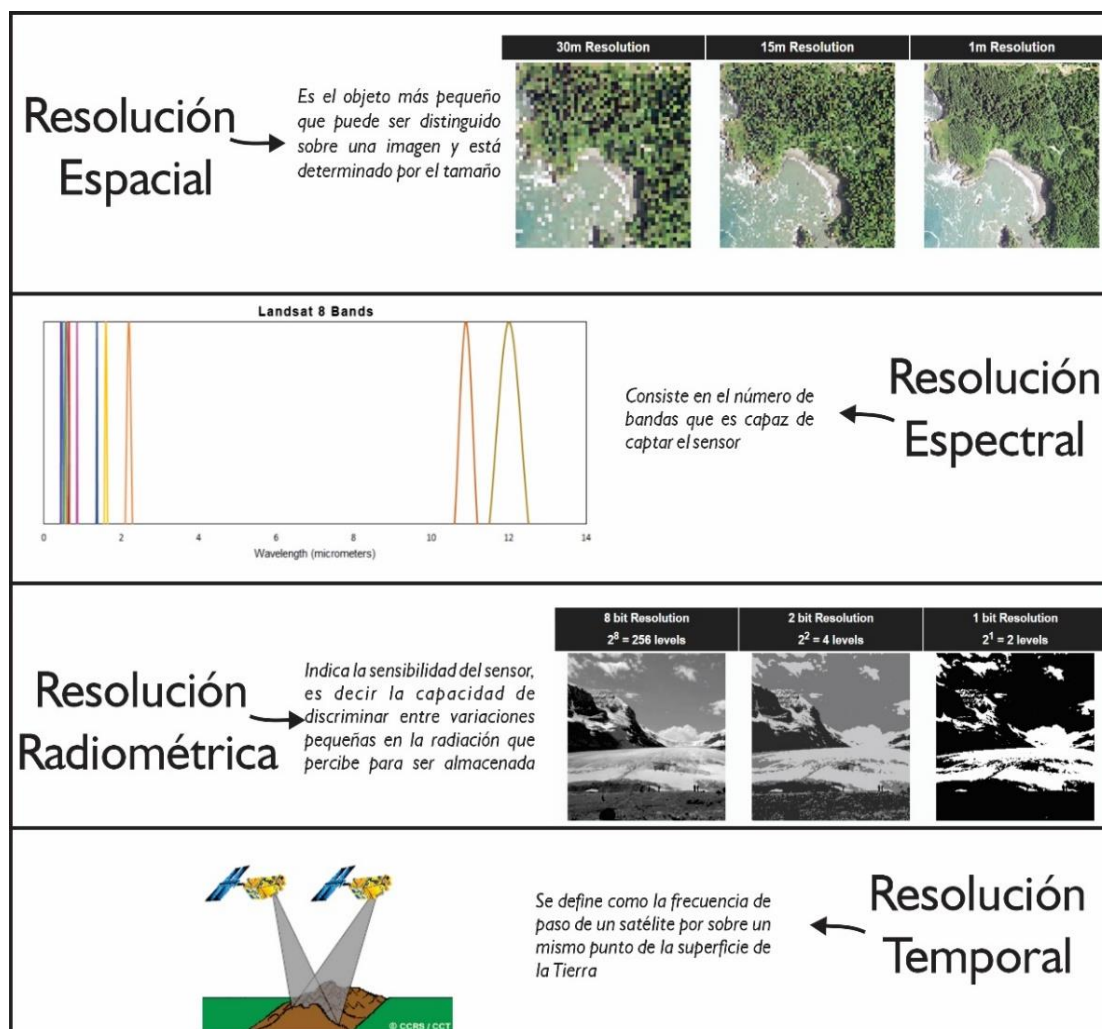
### **5.3.7 Tipos de Resolución de un Sistema Sensor**

La resolución de un sistema sensor es la habilidad para registrar, discriminar y dar información al detalle, este concepto de resolución implica por lo menos 4 manifestaciones: espacial, espectral, radiométrica y temporal (Chuvieco, 1990).

La *resolución espacial* es el objeto más pequeño que puede ser distinguido sobre una imagen y está determinado por el tamaño del pixel, medido en metros sobre el terreno (Chuvieco, 1990), la *resolución espectral* consiste en el número de bandas que es capaz de captar el sensor, la *resolución radiométrica* indica la sensibilidad del sensor o la capacidad de discriminar entre variaciones pequeñas en la radiación que percibe para ser almacenada y procesada (Chuvieco, 1990). Por último la *resolución temporal* se define como la frecuencia de paso de un satélite por sobre un mismo punto de la superficie de la Tierra

(Humboldt, 2019).

**Figura 27.** Tipos de resolución que se presentan en una imagen (Modificado de Humboldt, 2019).



### 5.3.8 Combinación de Bandas

Los sensores remotos pueden ser utilizados para aplicaciones sencillas como la delimitación de cuerpos de aguas superficiales, para esto solo bastaría un proceso simple como la visualización de una banda de infrarrojos cercana o media, pero para trabajos más complejos como un mapeo mineral, las combinaciones algebraicas de ciertas bandas a través de la división, resta, suma o producto pueden ser de gran ayuda y generar una información mejorada (Khorram *et al.*, 2012).

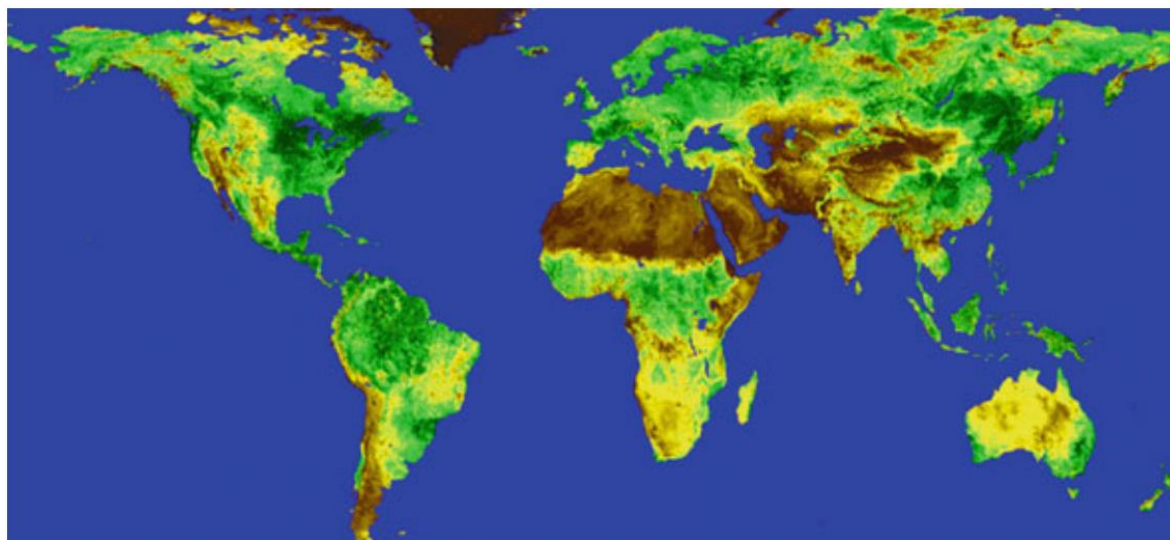
Por ejemplo, una de las combinaciones más utilizadas es la del índice de vegetación de diferencia normalizada o NDVI por sus siglas en inglés, este índice está definido por la

**Ecuación 10:**

$$NDVI = \frac{B_2 - B_1}{B_2 + B_1} \quad (10)$$

Donde  $B_2$  representa los valores de brillo (los valores de píxeles) de una banda de infrarrojo de una imagen y  $B_1$  representa los valores correspondientes en la banda roja de la imagen (Khorram *et al.*, 2012). Estas imágenes pueden ser a blanco y negro o a color (**Figura 28**).

**Figura 28.** Ejemplo de una imagen NDVI representando la cubierta vegetal global en varios tonos de verde (Tomado de Khorram *et al.*, 2012).



### 5.3.9 Teledetección Aplicada a la Minería

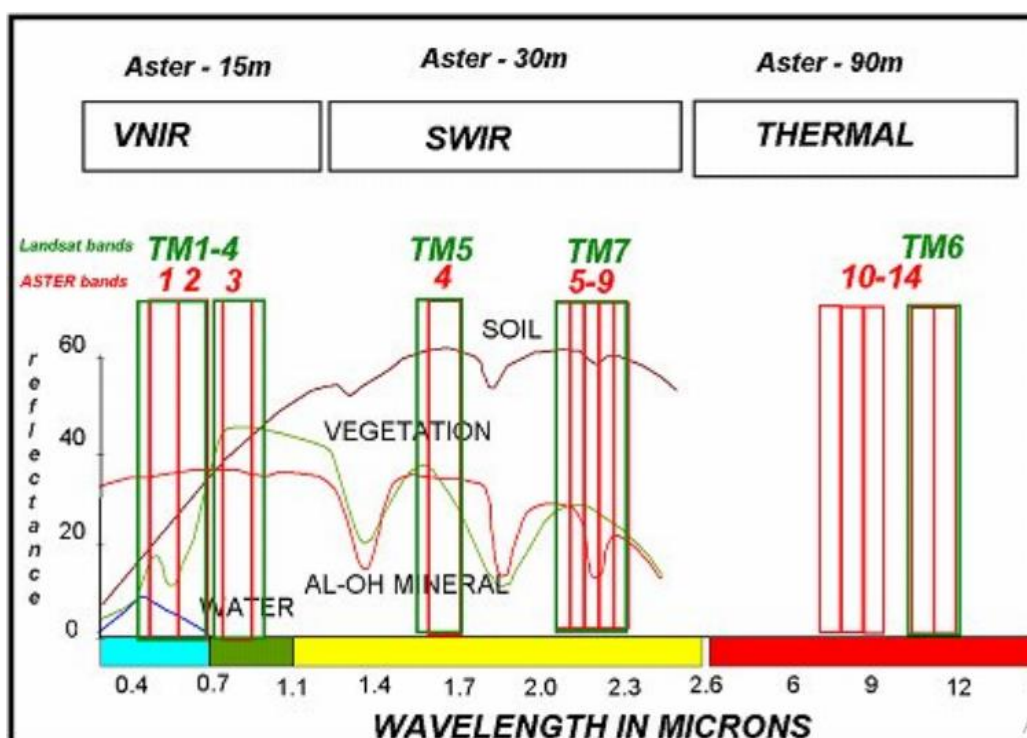
Las imágenes satelitales son una herramienta importante en la exploración minera ya que minimizan costos en cuanto al trabajo de campo. Con la ayuda de las imágenes satelitales de sensores multispectrales (capaces de detectar entre 3 y 20 bandas) y sensores hiperspectrales (capaces de detectar hasta cientos de bandas) se puede, entre otras cosas:

discriminar entre diferentes tipos de litología, detectar grandes rasgos lineales, elaborar mapas geomorfológicos e identificar minerales como la epidota, illita, alunita, caolinita, clorita, óxidos, entre otros (Vargas, 2010). Las imágenes más utilizadas en la exploración minera son: LANDSAT, ASTER, ALI, QUICKBIRD, SPOT, HYPERION, entre otras (Vargas, 2010). Para el desarrollo del proyecto se utilizarán las imágenes ASTER, las cuales presentan características espectrales y geométricas diseñadas especialmente para aplicaciones geológicas como la identificación de áreas de alteración hidrotermal (Vargas, 2010).

### 5.3.10 ASTER (*Advanced Spaceborne Thermal Emission and Reflection Radiometer*)

Es un sensor multi-espectral a bordo de uno de los satélites del sistema de observación terrestre de la NASA, Terra, el cual fue lanzado en 1999 (Kalinowski & Oliver, 2004). Estos sensores miden la radiación electromagnética emitida y reflejada desde la atmósfera y superficie terrestre en 14 bandas, las cuales están subdivididas en 3 grupos: VNIR, SWIR y TIR (Figura 29).

**Figura 29.** Distribución de bandas ASTER y Landsat (Tomado de Kalinowski & Oliver, 2004).



El primer grupo es la radiación visible e infrarroja cercana (VNIR) que cuenta con las bandas (1-3) con una resolución espacial de 15m, el segundo grupo que conforman las bandas (4-9) que graban la radiación infrarroja de onda corta (SWIR) con una resolución espacial de 30m y por último de la banda 10 a la 14 que registran la radiación térmica infrarroja (TIR) a una resolución espacial de 90m (Kalinowski & Oliver, 2004).

**5.3.11 Cruces de Bandas en ASTER**

Las relaciones y combinaciones de bandas más comunes en minería y geología en general con imágenes ASTER son:

**Figura 30.** *Compilación de cruces de bandas comunes con imágenes ASTER, se utilizan para resaltar aspectos de interés geológico utilizados en minería (Tomado de Kalinowski & Oliver, 2004).*

Feature	Band or Ratio	Comments	Reference	Features	Red	Green	Blue	Reference
<b>Iron</b>				Vegetation and visible bands**	3, 3/2, or NDVI	2	1	
Ferric iron, Fe <sup>3+</sup>	2/1		Rowan; CSIRO	AlOH minerals/advanced argillic alteration***	5/6 (phen)	7/6 (musc)	7/5 (kaol)	Hewson (CSIRO)
Ferrous iron, Fe <sup>2+</sup>	5/3 + 1/2		Rowan	Clay, amphibole, laterite	(5x7)/6 <sup>2</sup> (clay)	6/8 (amph)	4/5 (lat)	Bierwith
Laterite	4/5		Bierwith	Gossan, alteration, host rock	4/2 (goss)	4/5 (alt)	5/6 (host)	Volesky
Gossan	4/2		Volesky	Gossan, alteration, host rock	6 (goss)	2 (alt)	1 (host)	
Ferrous silicates (biot, chl, amph)	5/4	Fe oxide Cu-Au alteration	CSIRO	Decorellation (envi)	13	12	10	Bierwith
Ferric oxides	4/3	Can be ambiguous*	CSIRO	Silica, carbonate, basic degree index	(11x11)/10/12 (silica)	13/14 (carb)	12/13 (basic)	Bierwith
<b>Carbonates / Mafic Minerals</b>				Silica, carbonate	(11x11)/(10x12)	13/14	12/13	Nimoyima
Carbonate / chlorite / epidote	(7+9)/8		Rowan	Silica	11/10	11/12	13/10	CSIRO
Epidote / chlorite / amphibole	(6+9)/(7+8)	Endoskarn	CSIRO	Discrimination for mapping	4/1	3/1	12/14	Abdelsalam
Amphibole / MgOH	(6+9)/8	Can be either MgOH or carbonate*	Hewson	Discrimination in sulphide rich areas	12	5	3	
Amphibole	6/8		Bierwith	Discrimination	4/7	4/1	(2/3) x (4/3)	Sultan
Dolomite	(6+8)/7		Rowan, USGS	Discrimination	4/7	4/3	2/1	Abrams (USGS)
Carbonate	13/14	Exoskarn (cal/dolom)	Bierwith, Nimoyima, CSIRO	Silica, Fe <sup>2+</sup>	14/12	(1/2) + (5/3)	MNF Band 1	Rowan (USGS)
<b>Silicates</b>				Enhanced structural features	7	4	2	Rowan (USGS)
Sericite / muscovite / illite / smectite	(5+7)/6	Phyllic alteration	Rowan (USGS); Hewson (CSIRO)					
Alunite / kaolinite / pyrophyllite	(4+6)/5		Rowan (USGS)					
Phengitic	5/6		Hewson					
Muscovite	7/6		Hewson					
Kaolinite	7/5	Approximate only*	Hewson					
Clay	(5x7)/6 <sup>2</sup>		Bierwith					
Alteration	4/5		Volesky					
Host rock	5/6		Volesky					

**6.3.12 Pre-Procesamiento de Imágenes**

Las imágenes obtenidas de los sensores remotos no se encuentran listas para su procesamiento por lo que es necesario una etapa previa en la que se minimicen las distorsiones y/o errores que podrían dificultar una clasificación exitosa y para extraer o mejorar la

información más importante de una imagen (Khorram *et al.*, 2012; Arozarena & Pastor, 2016).

Las principales correcciones son la geométrica, atmosférica y radiométrica (**Figura 31**) (Arozarena & Pastor, 2016).

**Figura 31.** Principales correcciones en el pre-procesamiento de imágenes (Modificado de Khorram *et al.*, 2012).

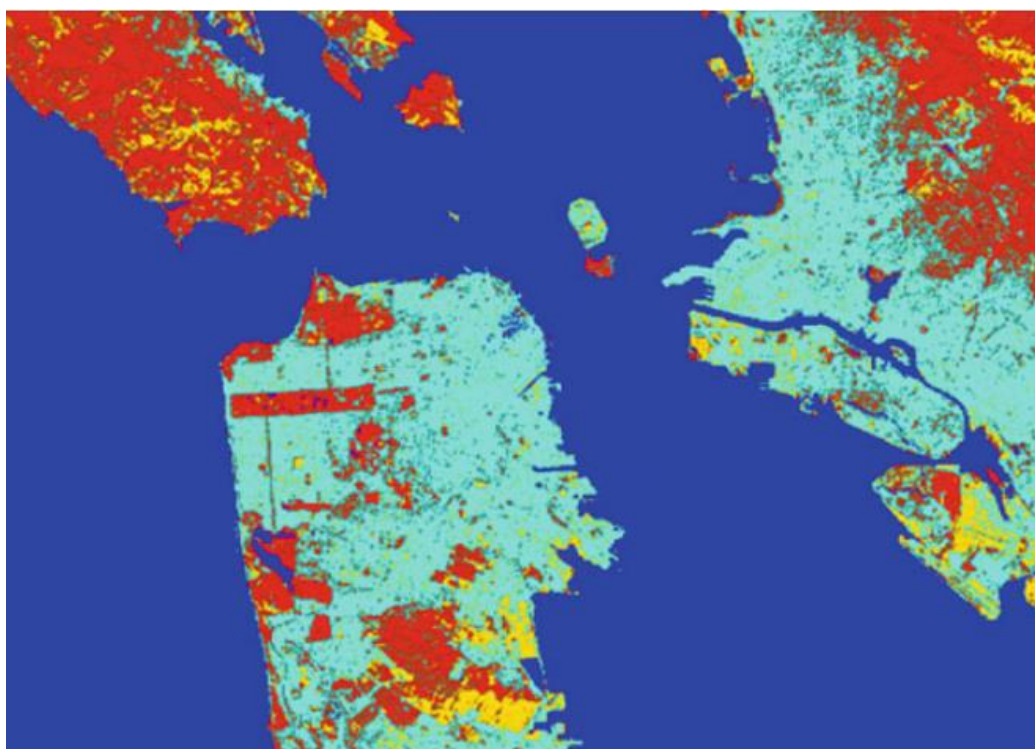
CORRECCIÓN GEOMÉTRICA	CORRECCIÓN RADIOMÉTRICA	CORRECCIÓN ATMOSFÉRICA
<p>Las imágenes capturadas con sensores remotos presentan 2 tipos de distorsiones geométricas: las <b>Sistemáticas</b> que se deben al movimiento de la imagen por el desplazamiento del vehículo sobre cual está montado el sensor, también por variaciones en la velocidad de escaneo del espejo, distorsiones panorámicas y distorsiones por la curvatura de la tierra. Las <b>No-Sistemáticas</b> que se dan por variaciones en la altitud o mal funcionamiento del sensor (Khorram <i>et al.</i>, 2012)</p>  <p>(Tomado de Fis, 2015)</p>	<p>Las distorsiones radiométricas varían dependiendo del sensor, pero normalmente se realizan correcciones para disminuir los efectos de dispersión o absorción causados por presencia de partículas en la atmósfera, los diferentes ángulos de incidencia de la luz solar y la distancia Sol-Tierra (FIS, 2015)</p>  <p>(Tomado de Khorram <i>et al.</i>, 2012)</p>	<p>Las correcciones atmosféricas se realizan debido a las partículas suspendidas en la atmósfera y las nubes que pueden cambiar los datos en el momento de su adquisición, debido a esto en la mayoría de las aplicaciones terrestres los analistas trabajan en días de baja nubosidad o con múltiples escenas (Khorram <i>et al.</i>, 2012)</p>  <p>(Tomado de Satimaging, 2015)</p>

### 6.3.13 Procesamiento de Imágenes Satelitales

Las imágenes de satélite que ya han sido pre-procesadas están listas para su procesamiento, lo que significa que están listas para una clasificación (Khorram *et al.*, 2012). Durante esta clasificación, cada pixel en una imagen es asignado a una categoría en particular de interés con la finalidad de convertir los datos espectrales originales en un mapa temático que los usuarios lo puedan entender y extraer la información más importante y significativa (Arozarena & Pastor, 2016). Las técnicas de clasificación de imágenes más utilizadas se basan en la lógica estricta, en la que los pixeles de la imagen se le asignan definitivamente a una sola categoría en el esquema de clasificación seleccionado (Hagen, 2003). Los sistemas de

clasificación basados en la lógica rígida involucran una de las 2 técnicas de clasificación: supervisada y no supervisada (Khorram *et al.*, 2012). Aunque el objetivo de estas técnicas es el mismo, siguen una estrategia conceptual fundamentalmente diferente, en resumen, la *clasificación supervisada* se basa en el intérprete humano de la imagen el cual identifica sitios de entrenamiento que representa cada clase del sistema de clasificación seleccionado al cual se le aplica un procedimiento de análisis para comprobar cada pixel de la imagen con los perfiles estadísticos y finalmente asignar los pixeles a la clase que mayor coincidencia (Khorram *et al.*, 2012; Arozarena & Pastor, 2016). La clasificación no supervisada por el contrario no tiene participación de un analista sino que la mayoría del trabajo se ejecuta a través de un proceso informático iterativo el cual asigna los pixeles de una imagen a grupos específicos que representan un grupo natural de pixeles espectralmente similares (**Figura 32**) (Khorram *et al.*, 2012).

**Figura 32.** *Clasificación no supervisada del área de la Bahía de San Francisco (Tomado de Khorram et al., 2012).*



## 6. Resultados

### 6.1 Análisis Estadísticos

Los datos pre-procesados se trataron en Excel y se extrajo la información necesaria para los diferentes cálculos de estadística univariable y multivariable. Estos parámetros se calcularon para 45 elementos como se muestra en la **Figura 33**.

**Figura 33.** Resumen de parámetros estadísticos de datos geoquímicos procesados de sedimentos activos en la Franja Metalogenética XV. Para cada elemento se generaron datos estadísticos de valor mínimo y máximo (Maximum and minimum), cuartil 1 y 3 (1st Quartile-3rd), mediana (Median), Desviación estándar (StDev) y el P-Valor, A-Valor generados mediante el Test de Anderson-Darling.

N. Samples: 2563	Au (ppm)	B (ppm)	Ba (ppm)	Cu (ppm)	V (ppm)	Ti (ppm)	Sr (ppm)	Li (ppm)	Cr (ppm)
A-Squared (Anderson-Darling)	814.58	966.23	363.32	666.37	123.62	53.58	272.59	260.95	106.03
P-Value (Anderson-Darling)	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005
Mean	0.03165	0.1124	78.97	58.13	101.75	974.66	75.64	4.4679	58.05
Standard Deviation	0.21146	1.1834	177.84	202.95	83.95	729.93	97.72	6.7102	50.809
Minimum	0.0001	0.00015	0.00017	2	5	0.012	2.2	0.045	2
1st Quartile	0.004	0.0002	0.001	19.3	47	500	30.4	0.05	20
Median	0.009	0.0016	0.006	29.4	79	800	52.6	0.061	43
3rd Quartile	0.01	0.03	110	45.3	128	1300	87.5	8	81
Maximum	5.33	22	3023	6408	793	7600	1600	73	427
N. Samples: 2563	Co (ppm)	Ce (ppm)	Cd (ppm)	Bi (ppm)	Be (ppm)	As (ppm)	Ag (ppm)	Zr (ppm)	Zn (ppm)
A-Squared (Anderson-Darling)	93.32	257.24	448.33	731.81	155.67	583.99	659.27	210.83	406.29
P-Value (Anderson-Darling)	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005
Mean	10.343	13.491	0.8337	2.954	0.58818	17.91	0.2103	6.5893	87.2
Standard Deviation	6.101	18.506	1.4376	14.574	0.3216	59.35	0.7447	8.0466	130.72
Minimum	1	0.005	0.001	0.05	0.003	0.9	0.005	0.0001	7.1
1st Quartile	6.9	0.005	0.19	0.19	0.4	3	0.09	1.6	40.2
Median	9	0.0001	0.9	4	0.45	7	0.1	3.4	59.1
3rd Quartile	12.3	25.6	1	4	0.7	15	0.15	8.5	90.4
Maximum	119.1	163.3	29	500	3.2	2175	26.2	74.8	2833
N. Samples: 2563	Sb (ppm)	Rb (ppm)	Pb (ppm)	Ni (ppm)	Nb (ppm)	Mo (ppm)	Lu (ppm)	La (ppm)	In (ppm)
A-Squared (Anderson-Darling)	652.35	258.95	639.68	139.68	268.72	579.41	297.19	64.42	334.03
P-Value (Anderson-Darling)	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005
Mean	3.513	6.38	30.09	14.241	0.34592	3.066	0.035632	14.083	0.010462
Standard Deviation	11.952	9.639	99.09	10.179	0.53285	9.496	0.054903	8.387	0.020786
Minimum	0.07	0.0005	1	0.005	0.001	0.13	0.0005	0.4	0.0005
1st Quartile	0.72	0.008	10.4	8.3	0.004	1	0.008	8.9	0.008
Median	4	0.015	17	12	0.015	1.47	0.015	12.6	0.015
3rd Quartile	4	10.8	27.9	17	0.61	3	0.07	17.6	0.02
Maximum	506	104.6	2643	135.6	5.17	318.24	0.67	104	0.35
N. Samples: 2563	Hg (ppm)	Hf (ppm)	Ge (ppm)	Ga (ppm)	Cs (ppm)	Y (ppm)	W (ppm)	U (ppm)	Tl (ppm)
A-Squared (Anderson-Darling)	828.12	436.19	386.4	262.88	418.75	39.13	181.87	293.51	401.74
P-Value (Anderson-Darling)	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005
Mean	0.1293	0.035966	0.0392	2.1459	1.6192	7.4831	7.543	0.5307	0.0963
Standard Deviation	0.8875	0.079411	0.065297	2.7844	4.0122	3.2641	8.7371	0.9342	0.24387
Minimum	0.0005	0.0001	0.0005	0.005	0.001	1.1	0.006	0.005	0.0002
1st Quartile	0.016	0.0005	0.002	0.008	0.005	5.32	0.3	0.04	0.005
Median	0.09	0.0005	0.05	0.2	0.3	7	9	0.05	0.04
3rd Quartile	0.1	0.05	0.1	4.4	1.88	9.02	9	0.8	0.12
Maximum	27	0.9	0.8	15.8	75.1	64.8	96	11.6	6.04
N. Samples: 2563	Th (ppm)	Te (ppm)	Sh (ppm)	Ta (ppm)	Se (ppm)	Sr (ppm)	P (ppm)	Na (ppm)	Mn (ppm)
A-Squared (Anderson-Darling)	304.95	586.71	338.92	918.95	540.43	66.8	251.33	313.11	141.25
P-Value (Anderson-Darling)	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005
Mean	2.0312	0.0235	5.531	0.00638	0.22532	3.2619	402.63	460.8	630.5
Standard Deviation	3.8171	0.08488	4.744	0.07699	0.60648	1.7948	618.55	709	453.2
Minimum	0.0001	0.0001	0.0005	0.0001	0.0001	0.0005	0.01	0.02	1
1st Quartile	0.0013	0.0005	0.7	0.004	0.0005	2.1	0.07	200	388
Median	0.002	0.003	9	0.0042	0.003	2.9	0.14	300	543
3rd Quartile	3.1	0.15	9	0.01	0.15	4	726	600	760
Maximum	54.6	2.24	125.9	2.12	8	16	9691	27400	10000

Al comparar todos los P-valor (P-value) para cada uno de los elementos, con el valor de Anderson-Darling (A-Squared) (**Figura 33**), se observó que A-Squared >>> P-value, esto indica que la función no cumple con una distribución normal y la hipótesis  $H_0$  es nula, por el contrario esta es una hipótesis alterna y desarrolla una distribución normal logarítmica (Log-normal)(Anderson & Darling, 1930).

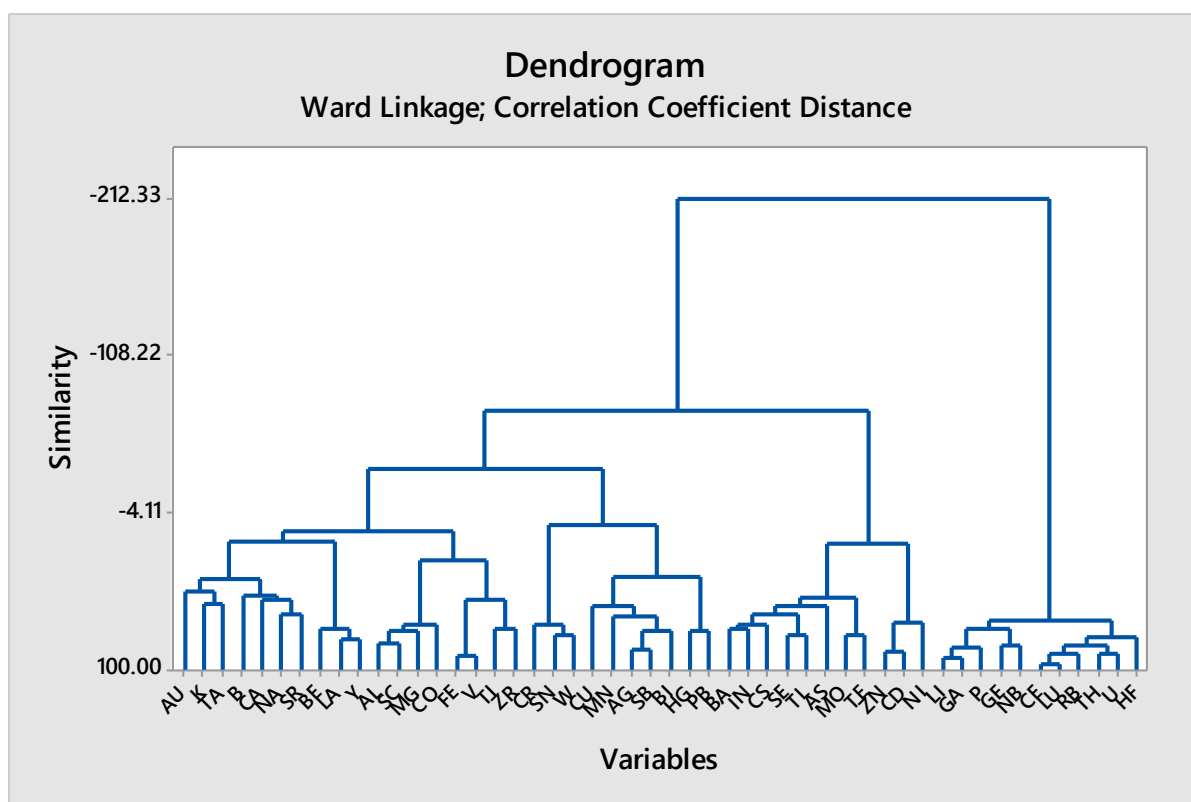
### **6.1.1 Cluster Analysis-Estadística Multivariable**

Para determinar las relaciones entre los distintos elementos se realizó primeramente una segregación entre los elementos mayores con los menores y traza. Posteriormente, con los elementos traza se llevaron a cabo diagramas de cajas (Boxplots), Dendogramas y matrices de correlación lineal de Pearson.

### **6.1.2 Dendogramas**

Después de exámenes cuidadosos de las combinaciones disponibles de métodos de similitud, el método de varianza mínima de Ward generó los agrupamientos mejor representados visualmente y permitió establecer y graficar jerárquicamente las relaciones entre los elementos, con esto se identificaron asociaciones geoquímicas afines a mineralizaciones, principalmente los relacionados con yacimientos tipo Pórfido, relacionados a intrusivos y Skarn (Au, Ag, Cu, Zn, Mo)(Perelló *et al.*, 2003). Se generaron 2 dendogramas, el primero **Figura 34** se realizó con todos los elementos para identificar y filtrar los que presentaran mayor asociación; Individualmente los elementos que presentaron mayor asociación fueron: (Ce-Lu-Rb), (Li-Ga-P), (U-Th-Nb), (Fe-V), (Ag-Sb), (Ti-Zr), (Zn-Cd).

**Figura 34.** Dendrograma de todos los elementos menores y traza (a excepción del Fe) en ppm estudiados.

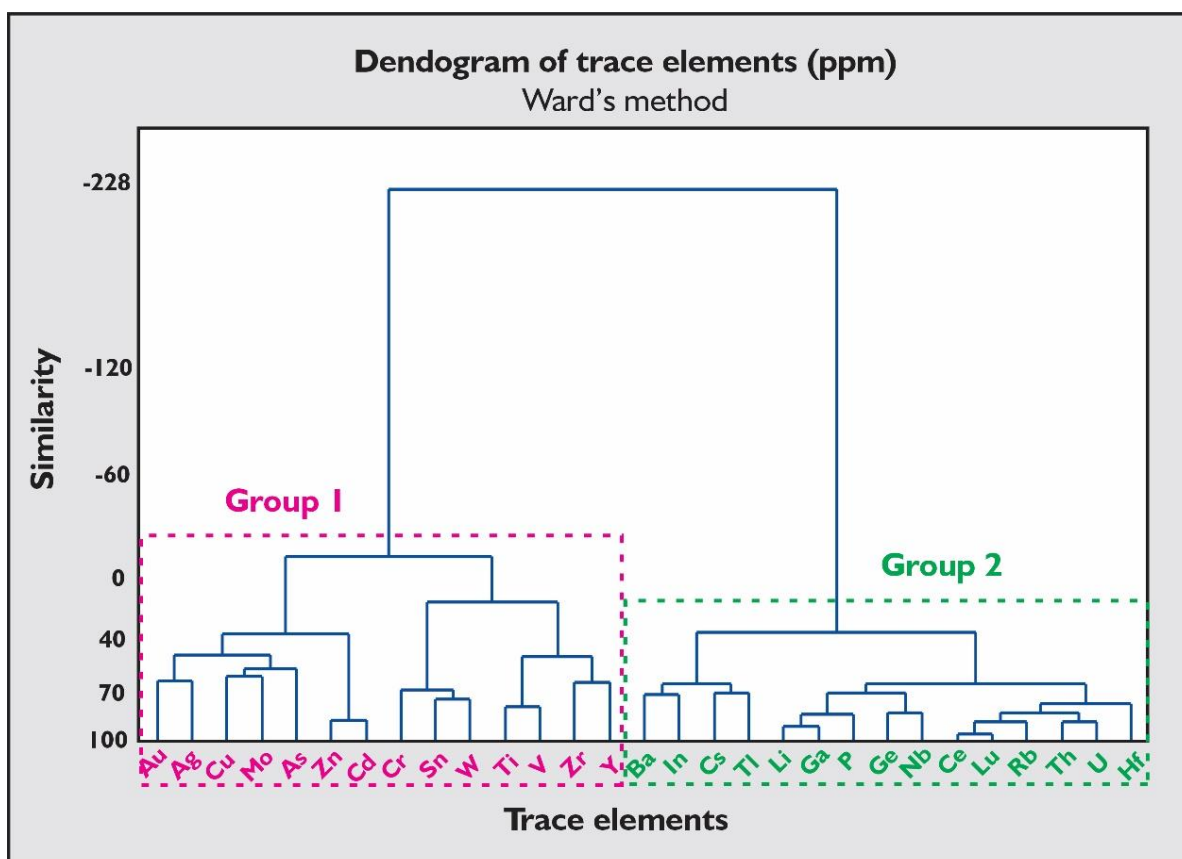


Para el segundo dendrograma se realizó un filtrado de los elementos de menor interés en relación con los tipos de yacimientos de la zona, así como los que presentaban menor relación en general (**Figura 35**). Se identificaron 2 grupos geoquímicos: (1) el Grupo 1 compuesto por Au, Ag, Cu, Mo, As, Zn, Cd, Cr, Sn, W, Ti, V, Zr, Y, y el Grupo 2 compuesto por Ba, In, Cs, Tl, Li, Ga, P, Ge, Nb, Ce, Lu, Rb, U, Th, Hf). Una alta similitud indica alta codependencia entre los componentes. Los elementos de mayor interés son los que pertenecen al grupo 1, principalmente los asociados con los elementos formadores de mineralizaciones (Au-Ag-Cu-Mo-As-Zn-Cd), y relacionados con los depósitos tipo Pórfido, Skarn y Epitermales del área de estudio (Acosta *et al.*, 2009).

También se tuvieron en cuenta los elementos indicadores o “pathfinders” (Ag-Mo-Pb-Zn-Tl-Sb) que nos son formadores de mineralizaciones de interés pero que están asociados a

ellos y que se usaron para detectar y rastrear yacimientos (Jones, 1992).

**Figura 35.** Dendrograma filtrado de elementos estudiados en ppm. Se observan 2 grupos de elementos con distribución similar (Grupo 1: Au, Ag, Cu, Mo, As, Zn, Cd, Cr, Sn, W, Ti, V, Zr, Y; Grupo 2: Ba, In, Cs, Tl, Li, Ga, P, Ge, Nb, Ce, Lu, Rb, U, Th, Hf).



### 6.1.3 Matriz de Correlación Lineal de Pearson

Mediante una matriz de correlación lineal se cuantificó la relación geoquímica entre elementos y se verificaron los resultados de los análisis mediante dendogramas. Los coeficientes de correlación de Pearson ilustran de mejor forma las relaciones individuales entre los elementos filtrados mediante el segundo dendograma.

Se observa una clara relación entre los elementos del grupo 1, asociados a las mineralizaciones de interés, también se observa que el Cesio es el elemento más correlacionado, con coeficientes de correlación de más del 70% con el Uranio (**Figura 36**), Th,

Rb, Nb, Ga y Lu con el que presenta un 92% de correlación.

No se observa ninguna codependencia entre elementos del grupo 1 con elementos del grupo 2, por el contrario son inversamente proporcionales.

**Figura 36.** Matriz de correlación lineal de Pearson entre elementos geoquímicos estudiados.

Los intervalos resaltados son: alto (>0.7), media (0.4-0.7), baja (-4-(+4)), media inversa ((-0.4)-(-0.7)) y alta inversa (<-0.7).

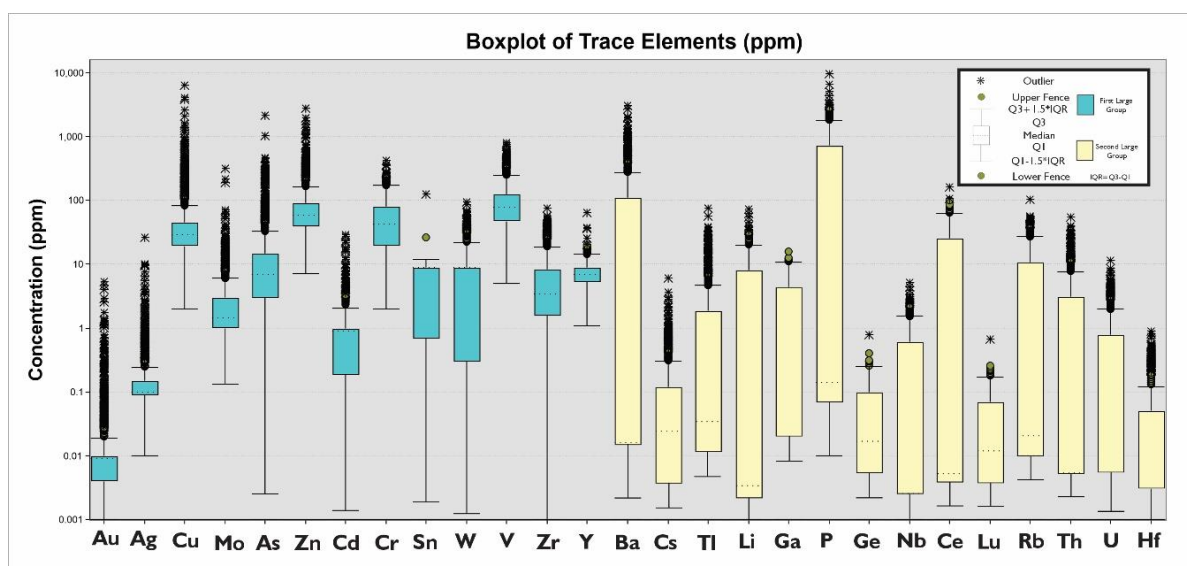
CORRELATION BETWEEN ELEMENTS OF THE STUDY AREA																															
	Au	Ba	Cr	Cu	Li	Na	P	Ti	V	Zn	Zr	Ag	As	Cd	Ce	Cs	Ga	Ge	Hf	In	Lu	Mo	Nb	Rb	Sh	Th	TL	U	W	Y	
Au	1.00																														
Ba	-0.02	1.00																													
Cr	0.11	-0.32	1.00																												
Cu	0.06	0.06	-0.05	1.00																											
Li	-0.07	0.42	-0.46	0.03	1.00																										
Na	-0.01	-0.11	0.06	-0.02	-0.13	1.00																									
P	-0.06	0.39	-0.43	0.05	0.67	-0.14	1.00																								
Ti	0.00	-0.10	0.05	-0.04	0.04	0.21	0.03	1.00																							
V	0.06	-0.14	0.08	0.00	-0.11	0.30	-0.02	0.61	1.00																						
Zn	0.09	0.16	-0.05	0.06	0.05	-0.05	0.04	-0.04	0.01	1.00																					
Zr	0.00	-0.16	0.10	-0.05	-0.24	0.12	-0.27	0.47	0.18	-0.01	1.00																				
Ag	0.29	0.04	0.06	0.26	-0.06	-0.01	-0.05	-0.01	0.00	0.25	0.00	1.00																			
As	0.13	0.20	-0.05	0.10	0.05	-0.07	0.05	-0.16	-0.10	0.26	-0.08	0.23	1.00																		
Cd	0.04	-0.01	0.09	0.06	-0.15	-0.02	-0.14	-0.13	-0.02	0.75	0.02	0.16	0.22	1.00																	
Ce	-0.07	0.50	-0.49	0.02	0.71	-0.18	0.76	0.02	-0.12	0.05	-0.19	-0.06	0.08	-0.16	1.00																
Cs	-0.04	0.37	-0.28	0.03	0.47	-0.09	0.40	-0.04	-0.08	0.11	-0.09	-0.01	0.15	-0.02	0.54	1.00															
Ga	-0.08	0.43	-0.50	0.06	0.82	-0.11	0.77	0.16	0.03	0.02	-0.29	-0.07	0.03	-0.22	0.79	0.42	1.00														
Ge	-0.06	0.23	-0.33	0.03	0.50	-0.08	0.55	0.18	0.04	-0.02	-0.19	-0.05	-0.02	-0.19	0.67	0.32	0.71	1.00													
Hf	-0.05	0.28	-0.31	-0.01	0.48	-0.06	0.38	0.18	-0.02	-0.04	0.10	-0.06	-0.03	-0.17	0.69	0.41	0.52	0.50	1.00												
In	-0.03	0.45	-0.34	0.31	0.45	-0.12	0.40	-0.03	-0.08	0.16	-0.17	0.03	0.26	0.04	0.54	0.33	0.54	0.53	0.37	1.00											
Lu	-0.06	0.45	-0.49	0.03	0.74	-0.12	0.76	0.06	-0.12	0.09	-0.21	-0.06	0.10	-0.10	0.92	0.53	0.83	0.66	0.66	0.55	1.00										
Mo	0.06	0.28	-0.02	0.25	-0.01	-0.05	0.01	-0.11	-0.07	0.20	-0.09	0.07	0.22	0.25	0.05	0.06	0.01	0.02	-0.03	0.23	0.06	1.00									
Nb	-0.06	0.29	-0.41	0.00	0.63	-0.12	0.52	0.19	-0.10	-0.03	-0.23	-0.06	-0.02	-0.21	0.71	0.28	0.74	0.64	0.49	0.44	0.72	0.01	1.00								
Rb	-0.07	0.48	-0.46	0.03	0.70	-0.14	0.65	0.08	-0.15	0.04	-0.16	-0.06	0.08	-0.16	0.82	0.54	0.76	0.56	0.65	0.45	0.81	0.02	0.66	1.00							
Sh	0.07	-0.43	0.49	-0.05	-0.64	0.15	-0.64	0.06	0.13	-0.05	0.32	0.04	-0.08	0.13	-0.68	-0.38	-0.56	-0.37	-0.46	-0.69	-0.07	-0.61	-0.62	1.00							
Th	-0.05	0.34	-0.35	0.01	0.51	-0.12	0.56	0.13	-0.02	0.01	-0.07	-0.05	0.04	-0.14	0.76	0.40	0.61	0.56	0.69	0.41	0.80	0.03	0.60	0.70	-0.46	1.00					
TL	-0.02	0.38	-0.29	0.02	0.36	-0.12	0.32	-0.12	-0.13	0.39	-0.12	0.02	0.29	0.37	0.47	0.40	0.34	0.25	0.30	0.36	0.48	0.14	0.27	0.45	-0.37	0.34	1.00				
U	-0.05	0.32	-0.38	0.00	0.58	-0.14	0.60	0.10	-0.08	0.01	-0.14	-0.05	0.02	-0.14	0.74	0.38	0.65	0.54	0.57	0.41	0.75	0.01	0.65	0.70	-0.50	0.78	0.33	1.00			
W	0.15	-0.34	0.44	0.08	-0.52	0.14	-0.50	0.19	0.22	0.04	0.35	0.21	-0.01	0.13	-0.57	-0.30	-0.59	-0.41	-0.35	-0.28	-0.55	-0.02	-0.49	-0.52	0.56	-0.41	-0.31	-0.43	1.00		
Y	-0.03	0.09	-0.18	0.00	0.22	-0.04	0.36	0.23	0.11	0.17	0.31	-0.01	0.06	0.09	0.37	0.17	0.21	0.18	0.31	0.16	0.43	-0.03	0.14	0.29	-0.07	0.35	0.19	0.28	-0.08	1.00	

### 6.1.4 Diagramas de cajas (Boxplots)

El diagrama de caja (Figura 37) resume de manera visual las concentraciones de los elementos de interés, a la izquierda en azul aparecen los elementos de mayor interés económico pertenecientes al Grupo 1, y a la derecha los elementos del grupo 2.

En el grupo 1 el elemento que presenta las menores concentraciones es el Au y el de mayor concentración el V, el promedio de concentraciones es de 75.29 ppm. El grupo 2 el elemento de menor concentración es el Hf y el de mayor concentración el P, el promedio de concentraciones es de 38.54 ppm.

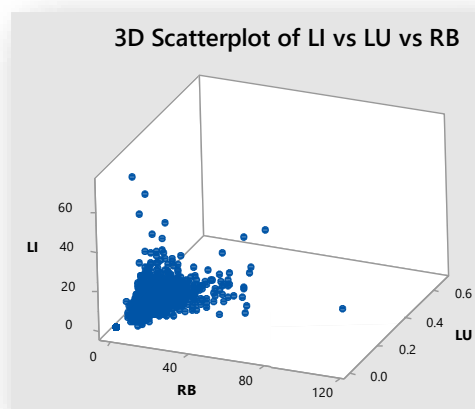
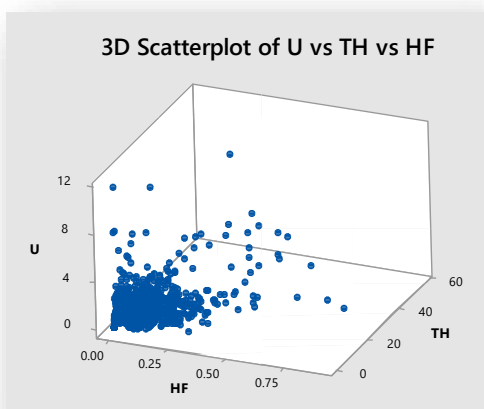
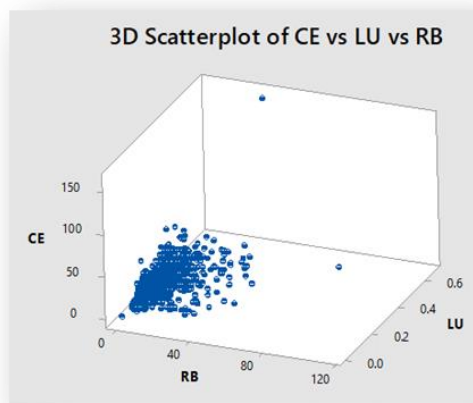
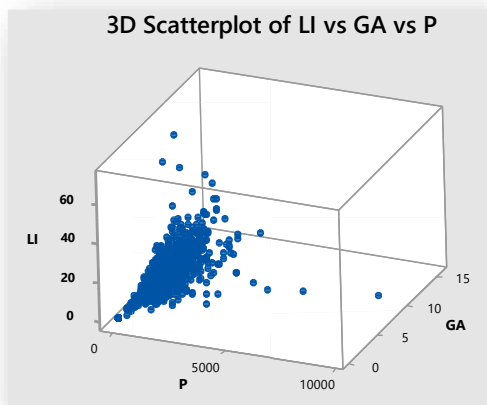
**Figura 37.** Comparación de diagramas de cajas de los diferentes elementos de interés en ppm, organizados por grupos de correlación geoquímica. Grupo 1 (azul) y Grupo 2 (amarillo).



### 6.1.5 Diagramas de Dispersión

Las mejores relaciones entre 2 elementos se graficaron y analizaron mediante diagramas de dispersión 2d y modelos de regresión lineal simple (**Figura 38**) para las relaciones entre 3 elementos se utilizaron gráficas de dispersión 3d, las cuales añadieron una dimensión adicional que redujo los problemas de superposición de puntos y permitió la representación eficiente del total de muestras (Piringer *et al.*, 2004).

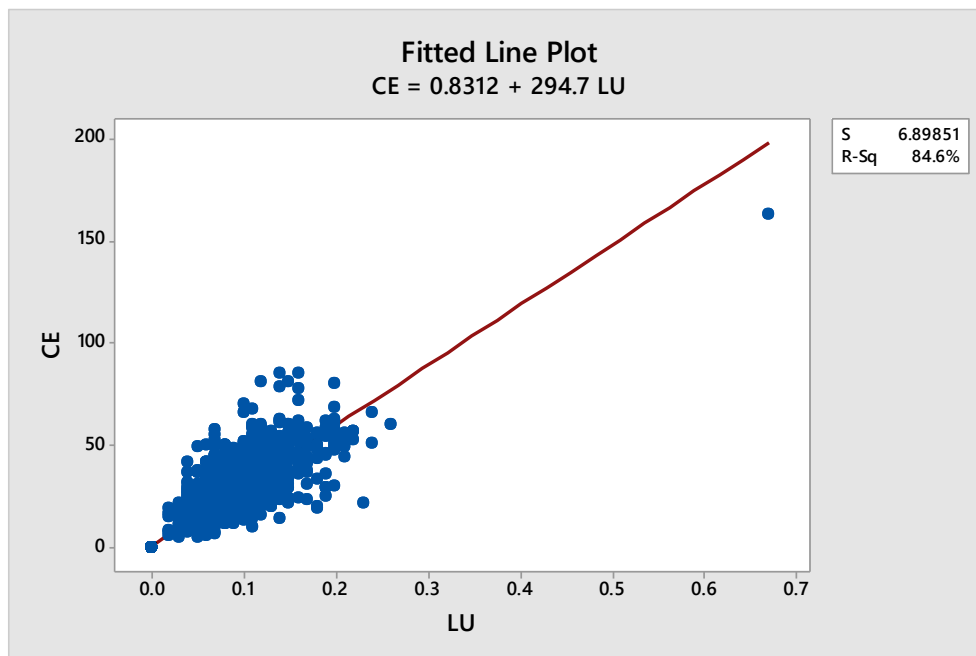
**Figura 38.** Diagramas de dispersión de tríos de elementos mejor correlacionados. En orden descendente las mejores correlaciones fueron (Li-Ga-P) > (Ce-Lu-Rb) > (U-Th-Hf) > (Li-Lu-Rb).



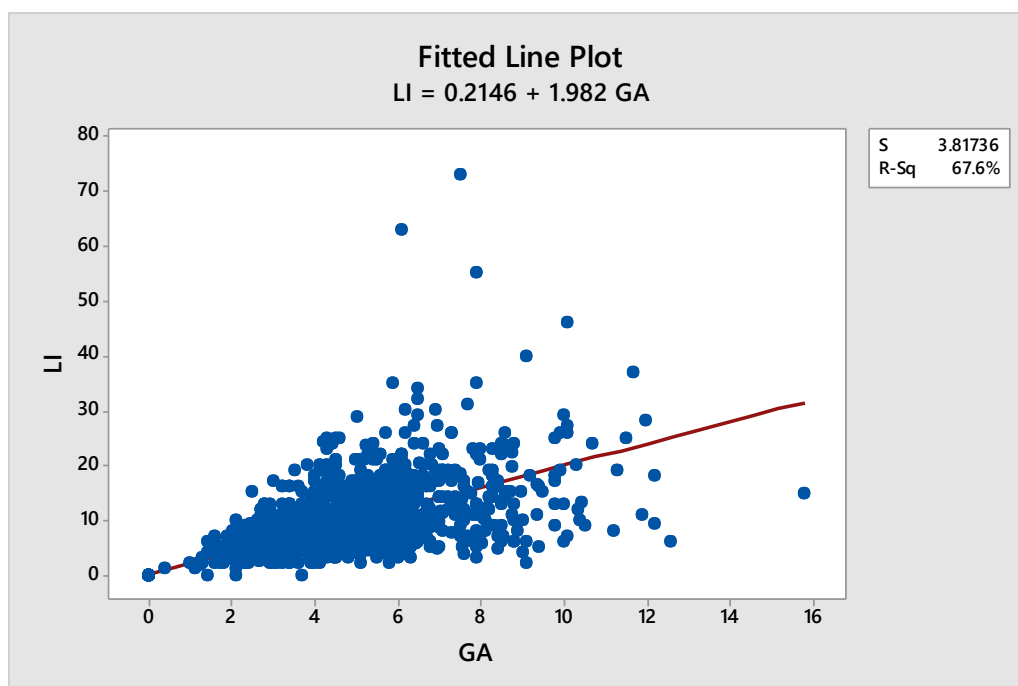
### 6.1.6 Diagramas de Dispersión 2D y Regesión Lineal

Se resumieron las mejores relaciones entre elementos mediante gráficos de dispersión (Figura 39-42), mostrando el coeficiente de determinación, desviación estándar y ecuación de regresión lineal de la forma  $y=mx + b$  donde el elemento en el eje x se muestra de manera dependiente. Estos diagramas se realizaron para (Zn-Cd), (Li-Ga), (Ce-Lu), (U-Th).

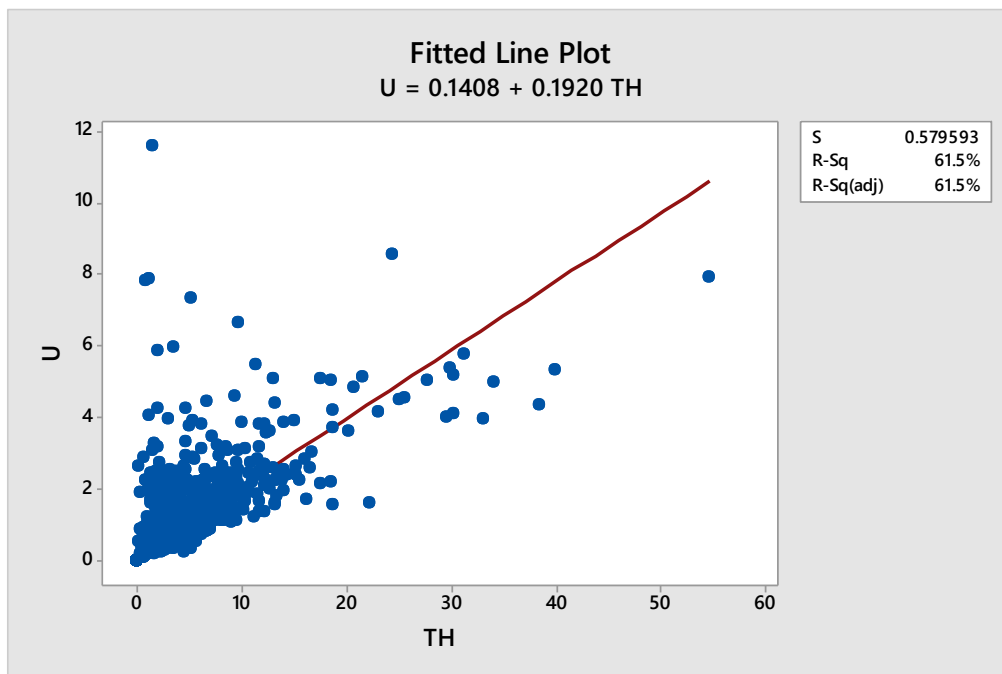
**Figura 39.** Gráfico de dispersión mostrando ecuación de regresión lineal y correlación entre Ce y Lu. Se observa un coeficiente de determinación del 84.6% y desviación estándar de 6.89. Estos dos elementos son los que mejor se correlacionan de todos.



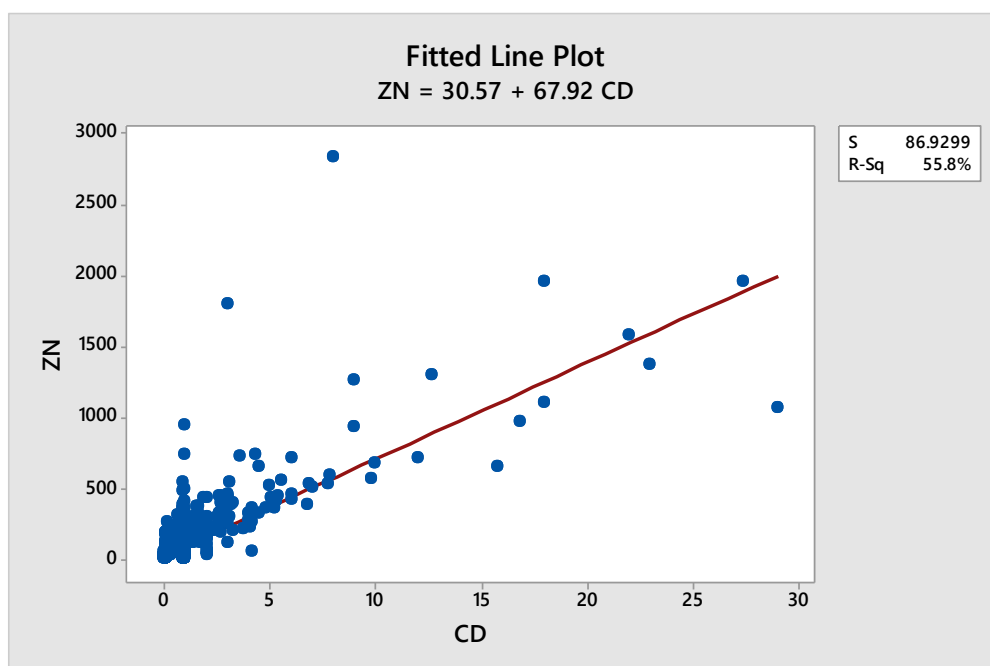
**Figura 40.** Gráfico de dispersión mostrando ecuación de regresión lineal y correlación entre Li y Ga. Se observa un coeficiente de determinación de 67.6% y desviación estándar de 3.81.



**Figura 41.** Gráfico de dispersión mostrando ecuación de regresión lineal y correlación entre el U y Th. Se observa un coeficiente de determinación de 61.5 % y desviación estándar de 0.579.



**Figura 42.** Gráfico de dispersión mostrando ecuación de regresión lineal y correlación entre el Zn y Cd. Se observa un coeficiente de determinación de 55.8% y desviación estándar de 86.92.



La correlación entre el Zn y Cd evidencian una típica asociación calcófila, su afinidad geoquímica se explica ya que ambos constituyen normalmente sulfuros y óxidos (*Vargas et al.*, 2017) los cuales están muy presentes en la mineralizaciones y zonas de alteraciones de óxidos de hierro (Hematita, limonita, jarosita) en la zona de estudio (*Carlotto et al.*, 2009).

En los análisis cluster se observa el Grupo 1 compuesto por Au, Ag, Cu, Mo, As, Zn, Sn, W, mostrando una importante relación entre los elementos formadores de depósitos (Cu, Au, Ag, Mo, Zn) con los *Pathfinders* (Cd, Sn, As, W). Esta tendencia también se puede explicar con el origen del magma calcoalcalino (granitos tipo I, emplazado a profundidad) propuesto por (*Rivera & Condori*, 2010) con cierta contaminación de la superficie (granitos tipo S) del Batolito de Abancay que ocupa gran parte de la zona de estudio (*Rivera*, 2009). Los granitos tipo I (son los que componen principalmente el cuerpo batolítico de Andahuaylas-Yauri, tienen una alta fugacidad lo cual explicaría las altas concentraciones elementos del Grupo 1. La asociación de elementos del grupo 2 (principalmente Ce, Lu, Rb, Th, U, Hf) se pueden explicar en base a lo propuesto por *Rivera & Condori*, (2010) que aunque clasifica al batolito con un dominio principalmente metalumínico típico de magmas relacionados a pórfidos de Cu, Au, Mo, en este se encuentran algunas facies peralumínicas de posible origen por contaminación cortical.

Por otra parte en base a lo propuesto por *Perelló et al.*, (2003) estos dos cluster se pueden explicar cómo dos diferentes pulsos magmáticos desarrollados durante el Eoceno-Oligoceno, esto explicaría las alcalinidades diferentes por 2 posibles grados de contaminación cortical.

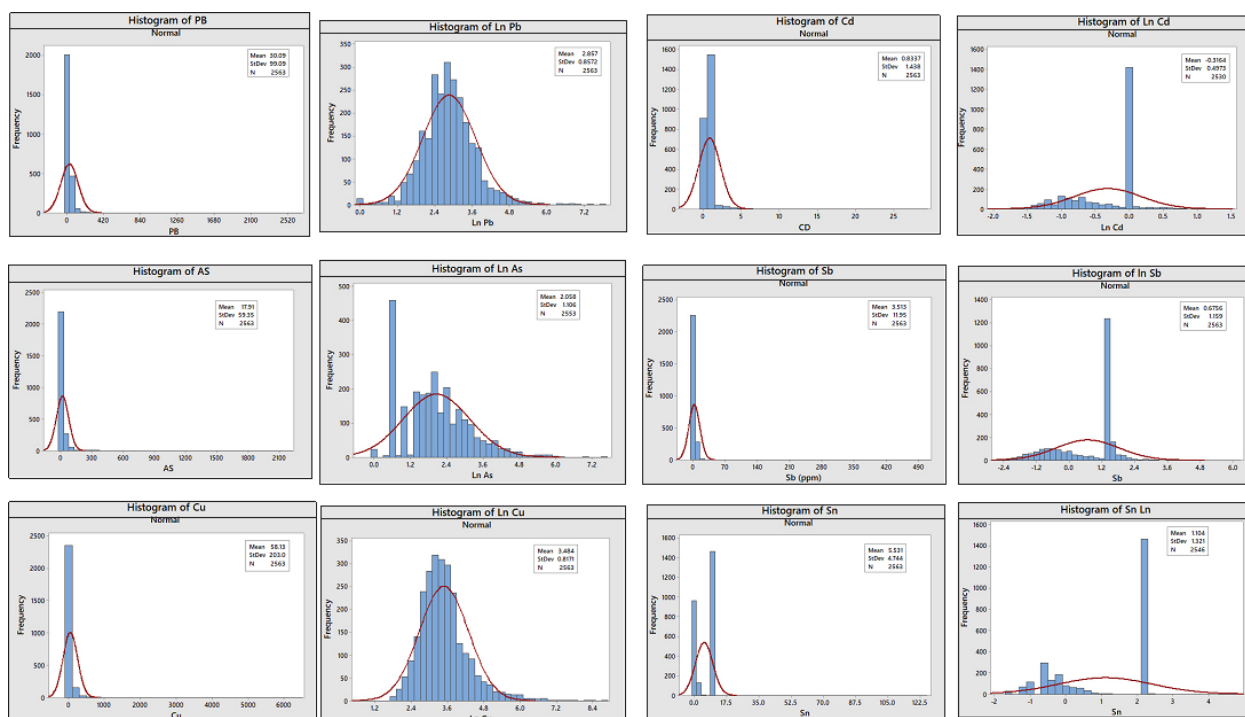
### **6.1.7 Interpolación - IDW**

Después de realizar los diferentes análisis estadísticos de correlación lineal, dendogramas, matrices de correlación lineal, se procedió a realizar una interpolación ponderada

por distancia inversa (IDW-Inverse distance Weighted), esta se usó para Au, Cu, Fe, Ag, Mo, As, Zn, Cd, Pb y Sb. Estos fueron los elementos escogidos por pertenecer a principales asociaciones, ser elementos principales y/o por ser considerados *pathfinders*.

Para la determinación de los valores de fondo y umbrales de anomalías se utilizaron los valores del Atlas Geoquímico del Perú realizado para cada dominio geológico regional, en este, el valor de fondo en distribuciones normales o log-normales se consideró igual a la mediana, mientras que el umbral geoquímico lo definieron como la suma del valor de fondo más 2 veces la desviación estándar, y para aquellas distribuciones que no tendían al patrón gaussiano definieron el valor de fondo por el percentil 50 y el umbral por el percentil 95 (Vargas *et al.*, 2017). Para la determinación de las anomalías de los elementos no analizados por el Atlas se realizó el mismo procedimiento, realizando una transformación logarítmica debido al comportamiento de los datos (**Figura 43**).

**Figura 43.** Histogramas de los elementos no encontrados en el Atlas Geoquímico del Perú y algunos de los incluidos pero realizados nuevamente para corroborar los límites de valores de fondo y umbrales. A la izquierda se encuentran los elementos mostrando distribución logarítmica, a la derecha nuevos histogramas de elementos normalizados. Esta transformación es de vital importancia para el cálculo de los valores de fondo y umbrales de los elementos y para la unificación posterior mediante redes neuronales.



Al tener todos los valores de fondo y umbrales para los elementos de interés se realizó la interpolación mediante IDW en ArcMap como se muestra en la **Figura 44**. En esta se evidencia la distribución en ppm del As, Fe, Ag, Cd, Cr, Cu, Au, Mo, Ni, Pb, Sb, Zn.

**Figura 44.** Mapas de Anomalías Geoquímicas en ppm de elementos analizados. Aunque todos los elementos presentan anomalías las mayores son de Fe, Cu, Au, Cr, Pb y Zn. Casi todos los elementos presentan valores anómalos en la zona suroeste de la franja metalogenética XV. Los elementos con menores anomalías fueron Cd y Sb.

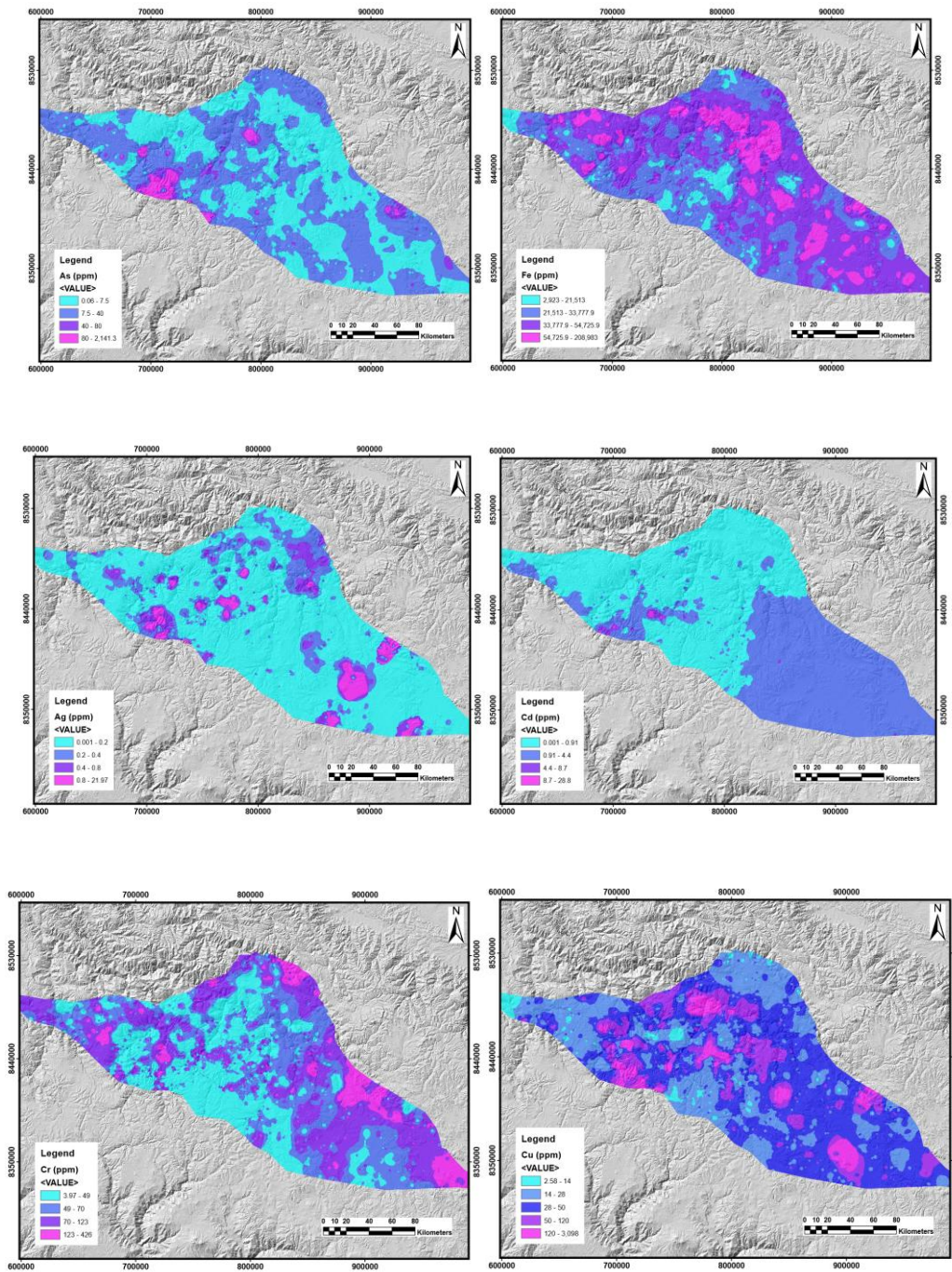
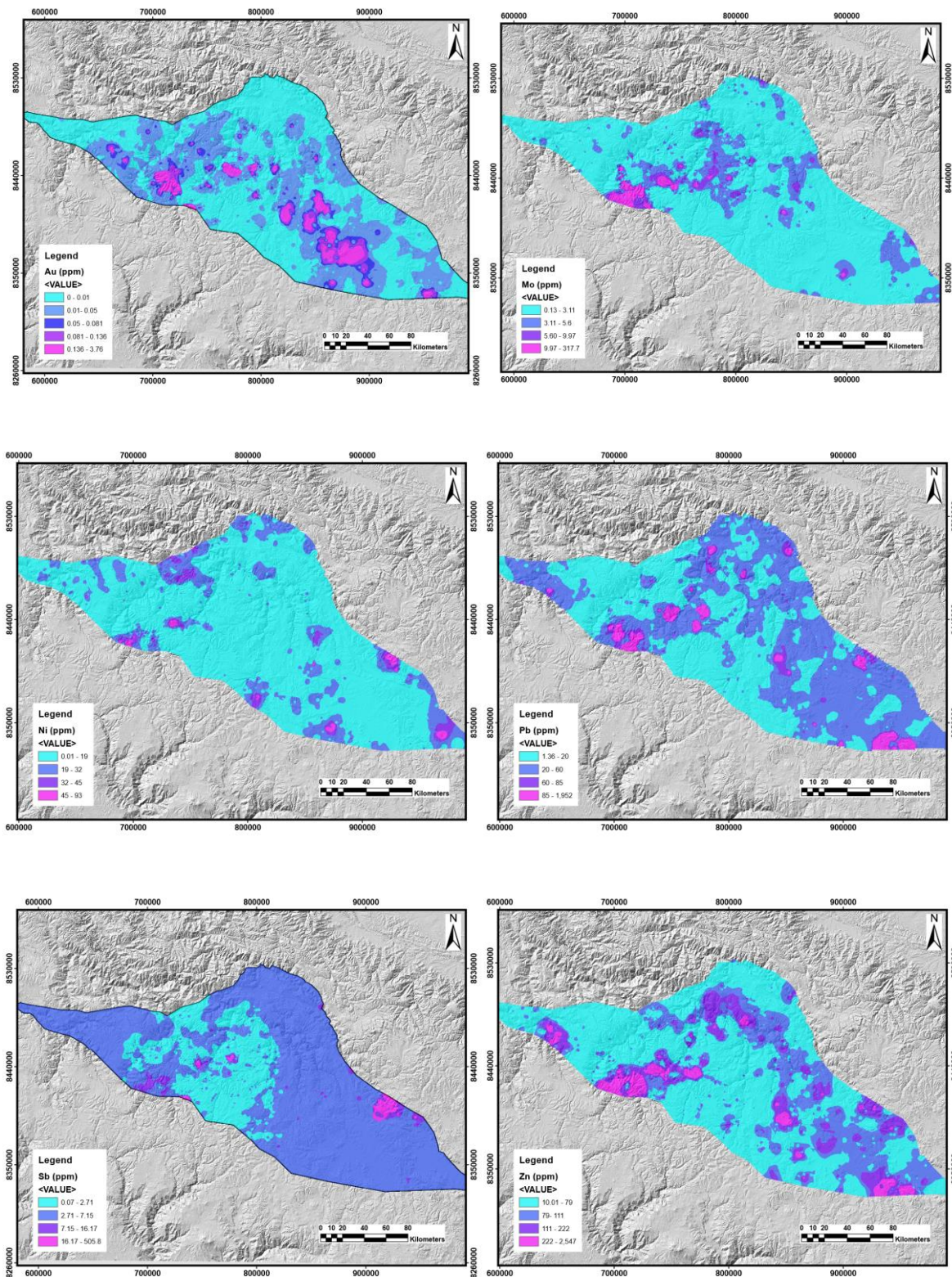


Figura 45. Segundo Bloque de anomalías geoquímicas de elementos analizados en ppm.



As: En la Franja se presentan pequeñas zonas de anomalías que se presentan como halos de dispersión anómalos de 2km aproximadamente, esta se circunscribe al dominio regional de

las rocas sedimentarias marinas y continentales del Mesozoico (Colque, 2016). Hacia el sector suroeste se presenta una gran zona de anomalías relacionadas con las tonalitas de la Unidad Progreso-Plutón Oyoní y las calizas de las formaciones Socosani y las areniscas cuarzosas de la formación Hualhuani. Finalmente, en la zona noroeste se evidencia un halo de anomalía de aproximadamente 15km relacionado nuevamente con las areniscas cuarzosas de la Formación Hualhuani y las calizas de la Formación Arcurquina. Estas anomalías son de gran importancia ya que muchas ocurrencias de Cu, Pb y Zn como las encontradas en la zona de estudio presentan al arsénico como *pathfinder* o elemento indicador (Hale, 1981).

**Cu:** Las anomalías de cobre encontradas corresponden a las descritas por (Vargas *et al.*, 2017). La anomalía de Atabamba relacionada con sedimentos provenientes de las unidades sedimentarias marinas y continentales del Mesozoico (Formación Hualhuani), la anomalía Chinchilpay con valores superiores a 767 ppm, asociada a las rocas sedimentarias marinas y continentales del Mesozoico (Formación Cachíos) y la Anomalía Lorocachi con concentraciones superiores a 2644ppm encontradas en las unidades sedimentarias del paleozoico (Grupo Mitú). Otras anomalías más pequeñas se localizan al noreste de la franja, cerca al Plutón Cotabambas y los flujos andesíticos del Grupo Tacaza. Al analizar las ubicaciones de las anomalías estas evidencian una gran relación con los yacimientos de tipo Pórfido-Skarn de Cu, encontrándose dentro o cerca a los halos anómalos.

**Cr, Zn:** las anomalías de cromo y zinc se encontraron relacionadas espacialmente hacia el sector suroeste, presentan una gran zona de anomalías relacionadas con las tonalitas de la Unidad Progreso-Plutón Oyoní y las calizas de las formaciones Socosani y las areniscas cuarzosas de la formación Hualhuani. Otro sector compartido se encuentra hacia el sureste circunscrito a las areniscas de la Formaciones Muñañi y Calasaya. Las anomalías de cromo van desde 281ppm (Pallcasullo) hasta 428ppm (Huncalaya) (Vargas *et al.*, 2017).

**Mo:** las anomalías de molibdeno se encontraron relacionadas a los intrusivos del paleógeno-Neógeno como las tonalitas de la unidad Progreso, también se presentan en la Formación Socosani y las calizas del grupo Yura.

**Au:** Los valores anómalos alcanzan 3.76 ppm y se encuentran asociados a las unidades intrusivas del Paleógeno-Neógeno (Plutón Charamuray, Plutón Oyoní) hacia el suroeste de la zona, también se encontraron anomalías en la zona sureste en los flujos andesíticos del Grupo Tacaza. Espacialmente se encuentran relacionados con vetas de oro y yacimientos de tipo pórfido identificados. Las anomalías se presentan una dirección preferencial en sentido NW donde se presentan importantes yacimientos como el Tintaya, Katanga, las Bambas y Cotabambas.

**Sb, Cd, Ag:** Se encontraron pocas anomalías de Cd y Sb, estas compartían valores anómalos en la zona sureste circunscritas a la formación Ananea donde hay ocurrencias de vetas de Au con Sb y Ag como acompañante (Acosta *et al.*, 2009). Las anomalías de plata se encontraron relacionadas con los intrusivos del Paleógeno-Neógeno y hacia la zona centro con las limolitas del Grupo Maure, los valores alcanzan 21.97 ppm y muestran relación espacial con los yacimientos de tipo pórfido como el de Los Chancas y vetas de Au.

Al analizar las zonas de anomalías con los yacimientos ya identificados los elementos que mejor coinciden son Cu, Au, Pb y Mo en respuesta a las mineralizaciones tipo Pórfido y Skarn, las que menos coincidieron fueron Cd y Cr.

## 6.2 Imágenes Satelitales

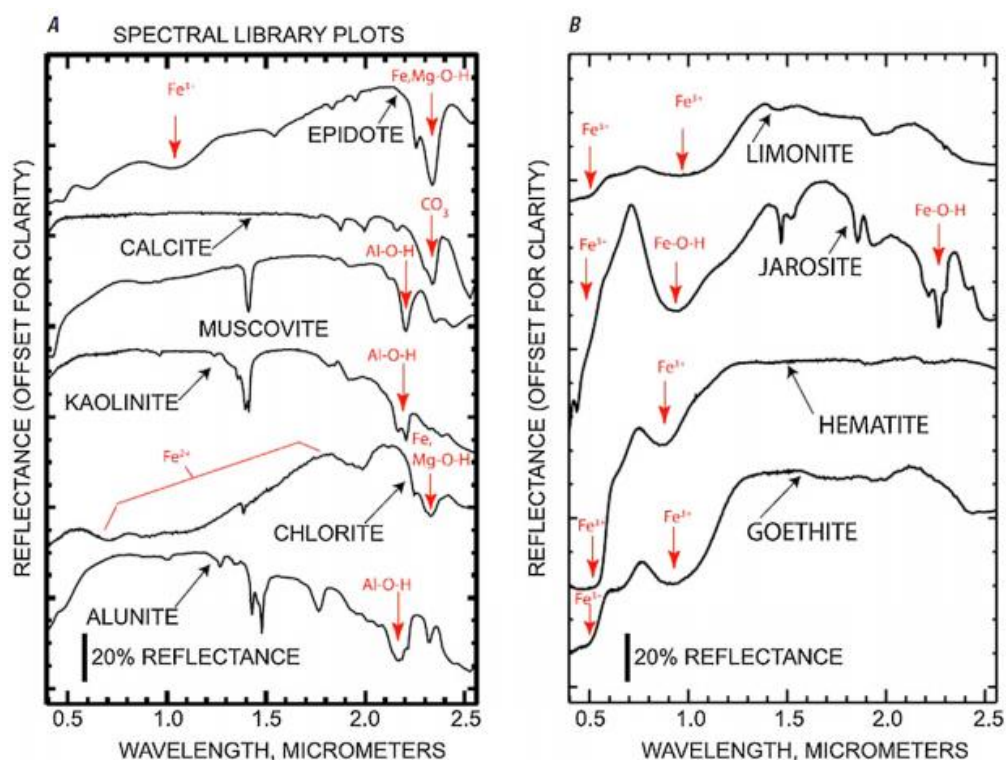
### 6.2.1 Landsat-8

Se crearon varias imágenes de combinación de colores rojo-verde-azul (RGB) y relaciones de bandas basadas en espectros de laboratorio, se utilizaron imágenes de diferentes

combinaciones de colores para resaltar rocas alteradas hidrotermalmente y las unidades litológicas a escala regional. La relación de bandas es una técnica en la que el valor numérico digital de una banda se divide por el de otra banda. Las relaciones de banda son útiles para resaltar ciertas características o materiales que no se pueden ver en las bandas sin procesar (Pour & Hashim, 2015a).

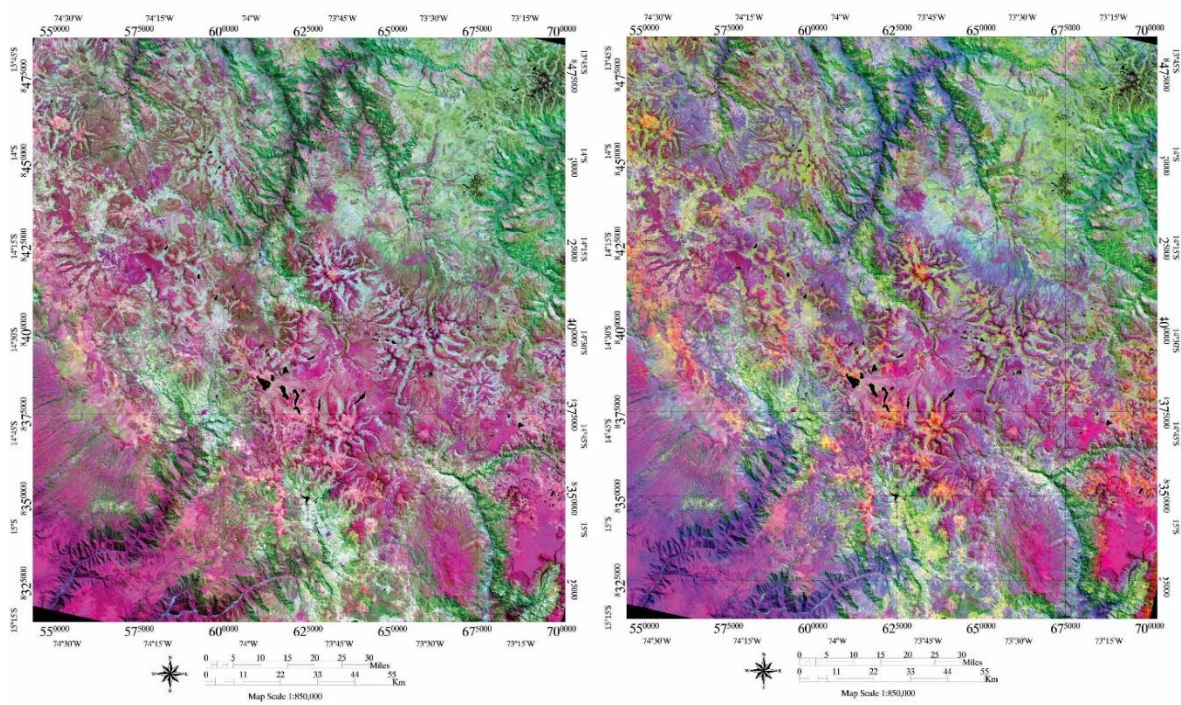
La relación 4/2 es útil para mapear óxidos de hierro (hematita, goethita, jarosita, etc.) porque tienen absorción en la región azul (banda 2), donde tienen una alta reflectancia en la región roja (banda 4). La relación 6/7 se utilizó por su capacidad para mapear caolinita, montmorillonita (minerales de arcilla), micas, talco-carbonatos, etc., todas estas características tienen una alta reflectancia en la banda 6 y una baja reflectancia en la banda 7 de la imagen Landsat-8 (Hunt & Ashley, 1979).

**Figura 46.** A, librería espectral de los minerales epidota, calcita, moscovita, caolinita, clorita y alunita. B, librería espectral de los minerales limonita, jarosita, hematita y goethita (Clark et al., 1997).



A continuación se presentarán los resultados generados con la matemática de bandas en las imágenes de la Franja Metalogenética XV del Perú, haciendo especial énfasis en algunas áreas de interés, por ejemplo en zonas de explotación minera conocidas, esto con el fin de validar los resultados obtenidos y presentar con mayor detalle características que se aprecian en imágenes con una mayor escala como la **Figura 47**. En la **Figura 48** encontramos la comparación entre imágenes con combinación de bandas 4/2 en la franja del rojo y 6/7 para el verde pero variando en el azul la banda 5 del infrarrojo cercano (NIR) y la banda 10 del infrarrojo térmico (TIRS).

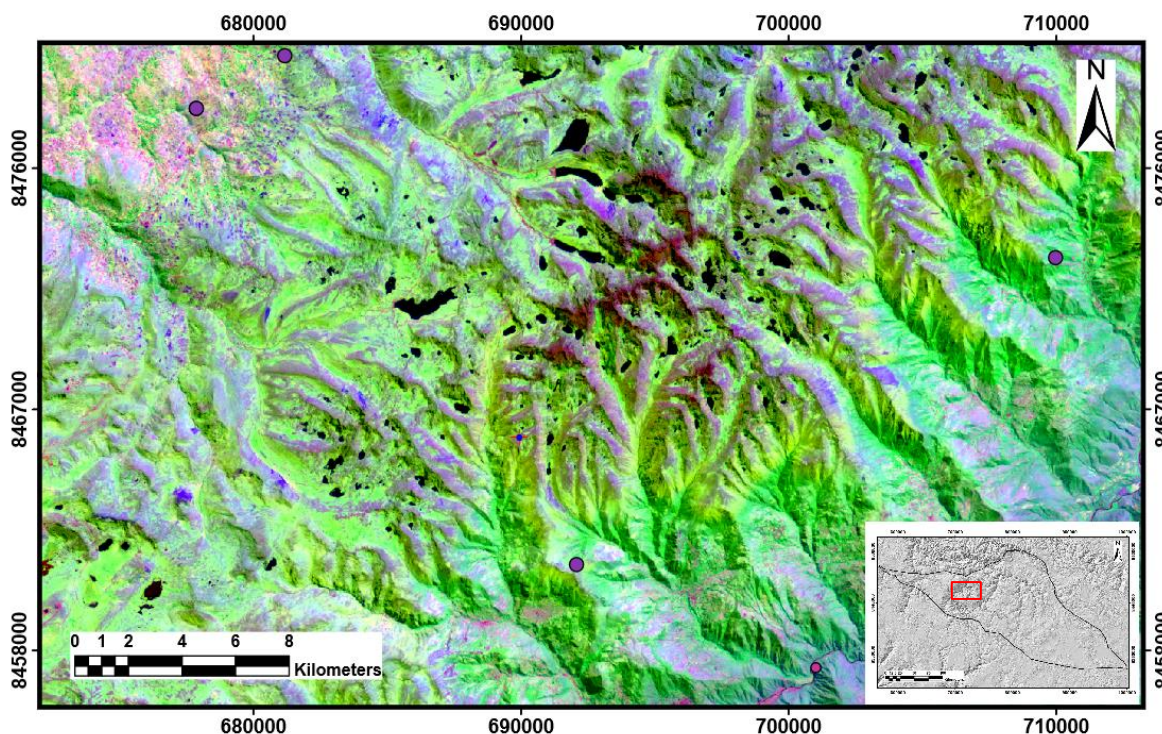
**Figura 47.** Imagen Landsat-8 LC80050702019218LGN00 con combinaciones de bandas RGB 4/2,6/7,5 y RGB 4/2,6/7,10 respectivamente, escala 1:800000.



Los óxidos de hierro y minerales sulfatos comúnmente muestran baja reflectancia cerca al ultravioleta o azul y reflectancia alta cerca al rojo (Rencz, 1999), esto genera un efecto “oxidado” en las imágenes con composición de color natural aunque estas no fueron incluidas por su poco aporte. Las zonas alteradas se encuentran en el dominio del Batolito de Abcancay

o Batolito de Andahuaylas Yauri.

**Figura 48.** Acercamiento de la imagen Landsat-8 LC80050702019218LGN00 mostrando en detalle depósitos tipo skarn.

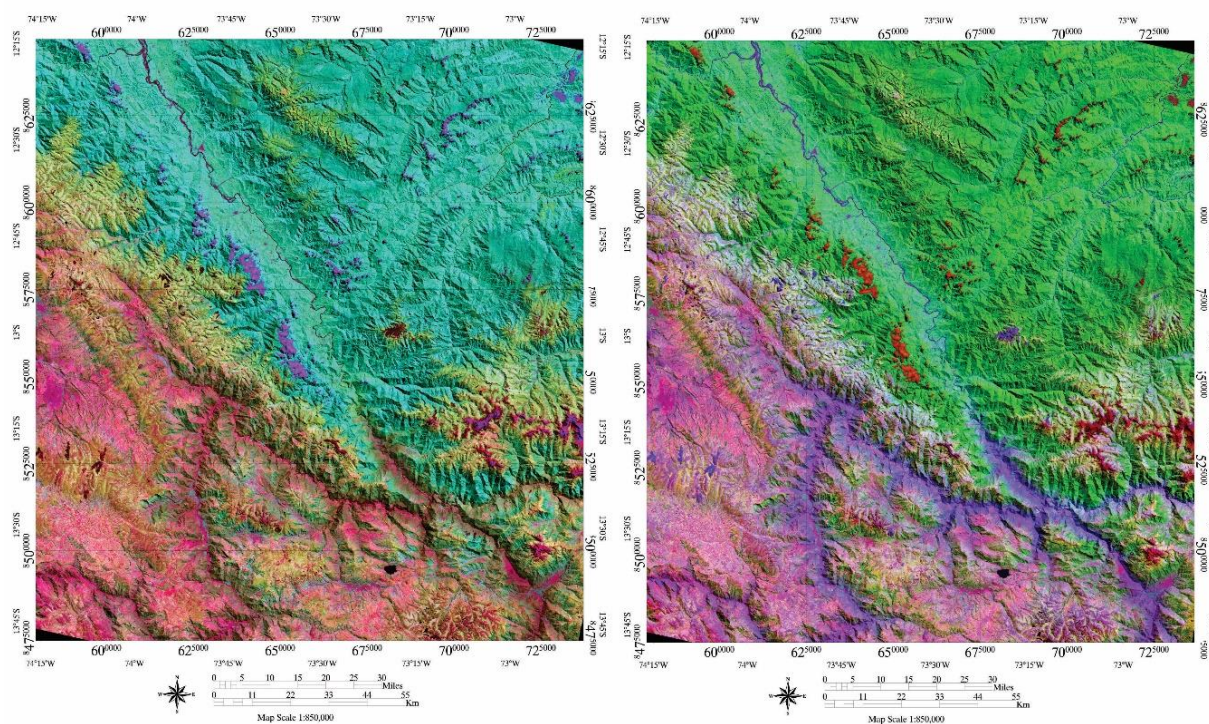


En la **figura 48** se observa como resaltan en variados tonos de morado, purpura y fucsia zonas de alteración y como ellas coinciden con algunos yacimientos de tipo skarn (señalados con un círculo morado) ubicados al sur y noroeste de la imagen. Los minerales de óxido e hidróxido de hierro como limonita, jarosita y hematita tienden a tener características de absorción espectral en el infrarrojo visible a medio de 0,4 a 1,1  $\mu\text{m}$  del espectro electromagnético (Pour & Hashim, 2015b). La relación 6/7 se utilizó por su capacidad para mapear caolinita, montmorillonita (minerales de arcilla), micas, talco-carbonatos, etc., todas estas características tienen una alta reflectancia en la banda 6 y una baja reflectancia en la banda 7 de la imagen Landsat-8 (Rowan *et al.*, 1977).

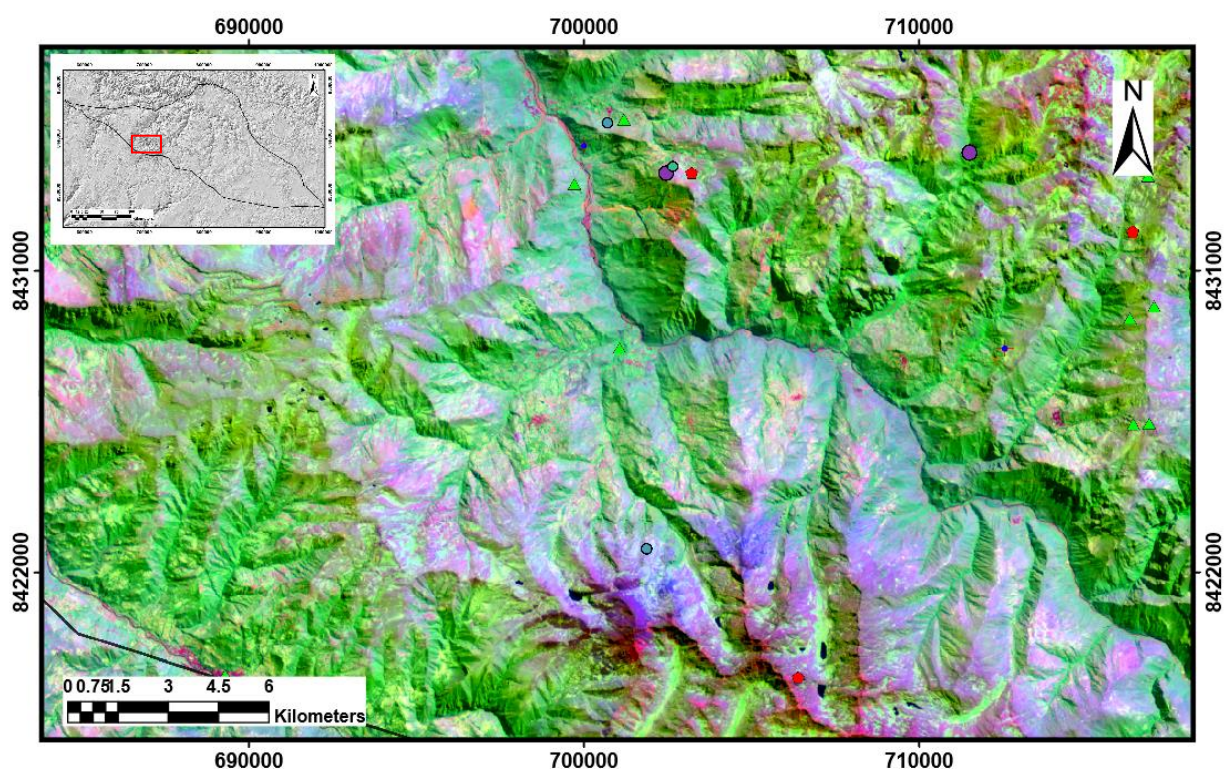
Se pudo apreciar claramente esta diferenciación en la composición del terreno de

estudio en la **Figura 49** donde existió una clara respuesta a la combinación de bandas 4/2 hacia la parte sur oeste mostrando colores asociados al rojo, como el morado intenso en la imagen de la derecha (RGB=4/2,6/7,10) y el fucsia brillante en el mapa a la izquierda (RGB=4/2,6/7,5), por otro lado al noreste se ven zonas verdes y azules respondiendo a la operación 6/7.

**Figura 49.** Imagen Landsat-8 LC80050692019218LGN00 con combinaciones de bandas RGB 4/2,6/7,5 y RGB 4/2,6/7,10 respectivamente.



**Figura 50.** Acercamiento de la imagen Landsat-8 LC80050702019218LGN00 mostrando en detalle depósitos tipo skarn (Círculos morados), pórfidos de Cu-M (rombos rojos), vetas (triángulos verdes) y epitermales (círculos fucsias) de intermedia sulfuración. Se utilizó la misma combinación en falso color 4/2, 6/7, 10.

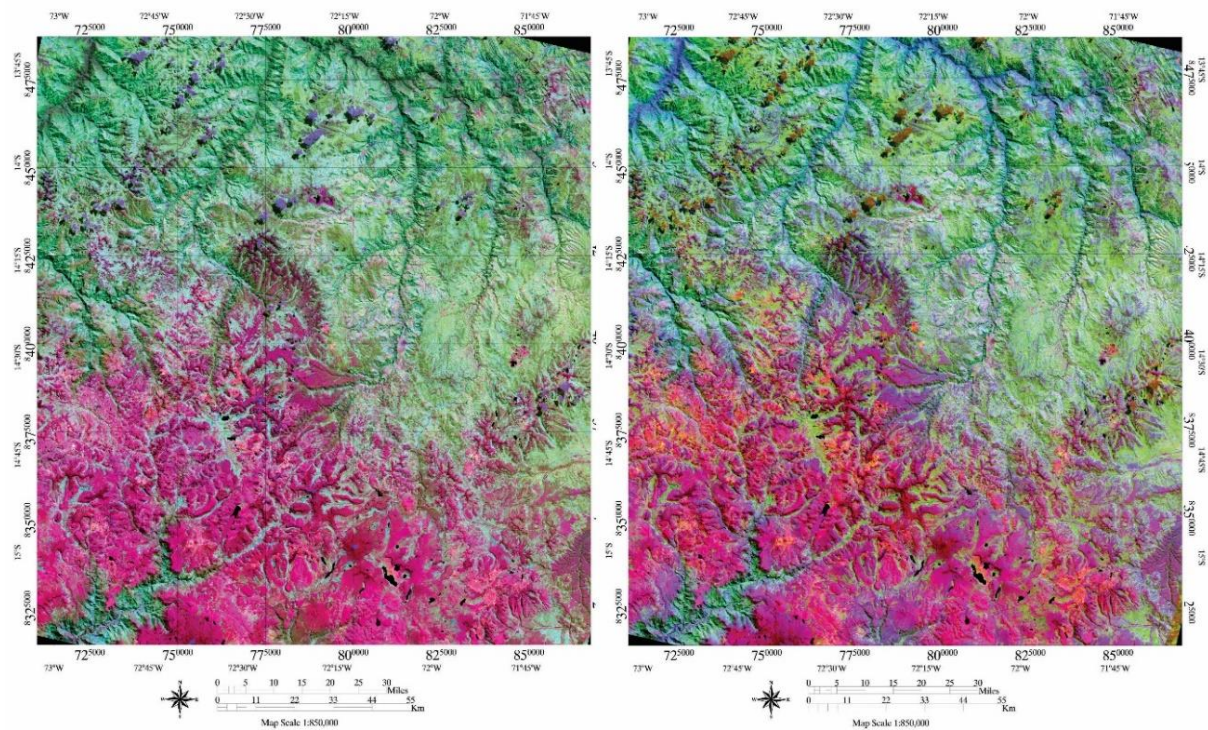


El objetivo del mapeo espectral es separar píxeles con firmas minerales distintivas de otros con información mezclada, como lo podrían ser, por ejemplo, datos de vegetación. Según (Sabins, 1999) es razonable pensar que la vegetación que crece sobre depósitos minerales, tendrá una reflectancia espectral diferente a aquella que no, pero en el caso de la presente zona de estudio, la vegetación no fue un obstáculo mayor, ya que la gran mayoría del área es totalmente desértica, lo que la hizo ideal para el mapeo.

En la **Figura 51**, se observa de nuevo colores típicos de los minerales óxidos de hierro asociados con yacimientos de pórfidos de Cu-Mo y skarns con numerosas vetas que parecen estar más relacionadas con lineamientos estructurales norte-sur.

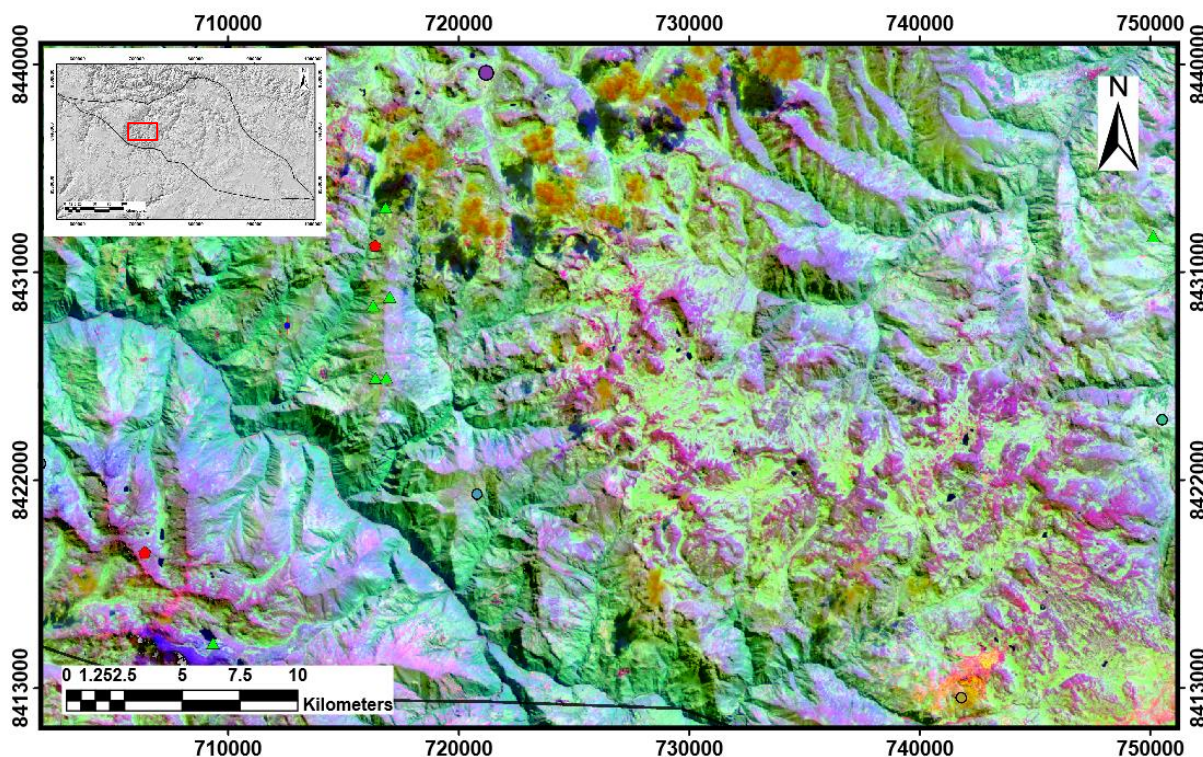
La imagen LC80040702020150LGN00 fue la que se ubicó en la parte más central del área de estudio, por ende, en ella fueron encontradas numerosas zonas de interés sobre las que se hizo énfasis y se presentan a continuación.

**Figura 51.** Imagen Landsat-8 LC80040702020150LGN00 con combinaciones de bandas RGB 4/2,6/7,5 y RGB 4/2,6/7,10 respectivamente. Se observan claramente 2 tonalidades predominantes y como con la banda 10 se resaltan de mejor manera zonas ricas en óxidos de Fe.



En la **Figura 52** se presentaron nuevamente vetas alineadas en sentido norte-sur, como siguiendo la cresta de la cadena montañosa en la parte superior de la imagen, pórfidos de Cu-Mo, epitermales indiferenciados y de alta sulfuración.

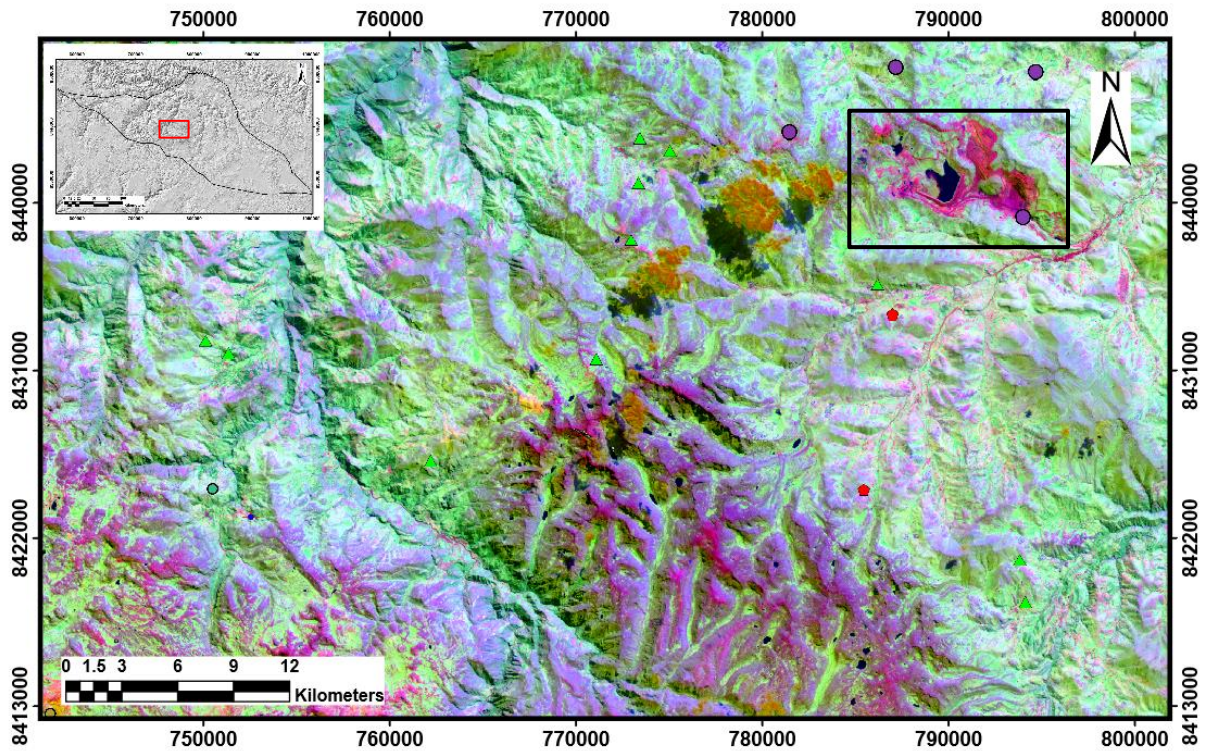
**Figura 52.** Acercamiento de la imagen Landsat-8 LC80050702019218LGN00 mostrando en detalle depósitos clasificados como vetas, pórfidos de Cu-Mo, epitermales indiferenciados y de intermedia sulfuración.



Encerrado en el rectángulo negro de la **Figura 53** se puede identificar el depósito pórfido-skarn de Las Bambas, el cual se encuentra 72 km al suroeste de la ciudad de Cusco, las coordenadas centrales son 787074E /8447490N (WGS 84, zona UTM 18s), comprendida entre las altitudes 3700 a 4650 m.s.n.m. Este depósito se divide en tres sectores: Ferrobamba, Chalcobamba y Sulfobamba, distribuidos a lo largo de 15 km (Acosta *et al.*, 2019).

El depósito está localizado en rocas sedimentarias de composición carbonatada del Cretáceo Inferior (Fm. Arcurquina), las cuales fueron intruídas por diferentes pórfidos de composición monzonítica del Eoceno-Oligoceno (Acosta *et al.*, 2019).

**Figura 53.** Acercamiento de la imagen Landsat-8 LC80050702019218LGN00 mostrando en detalle depósitos clasificados como vetas, pórfidos de Cu-Mo y skarns, la mina conocida como las Bambas se muestra en el rectángulo negro. La combinación de bandas fue la misma que en las anteriores imágenes.



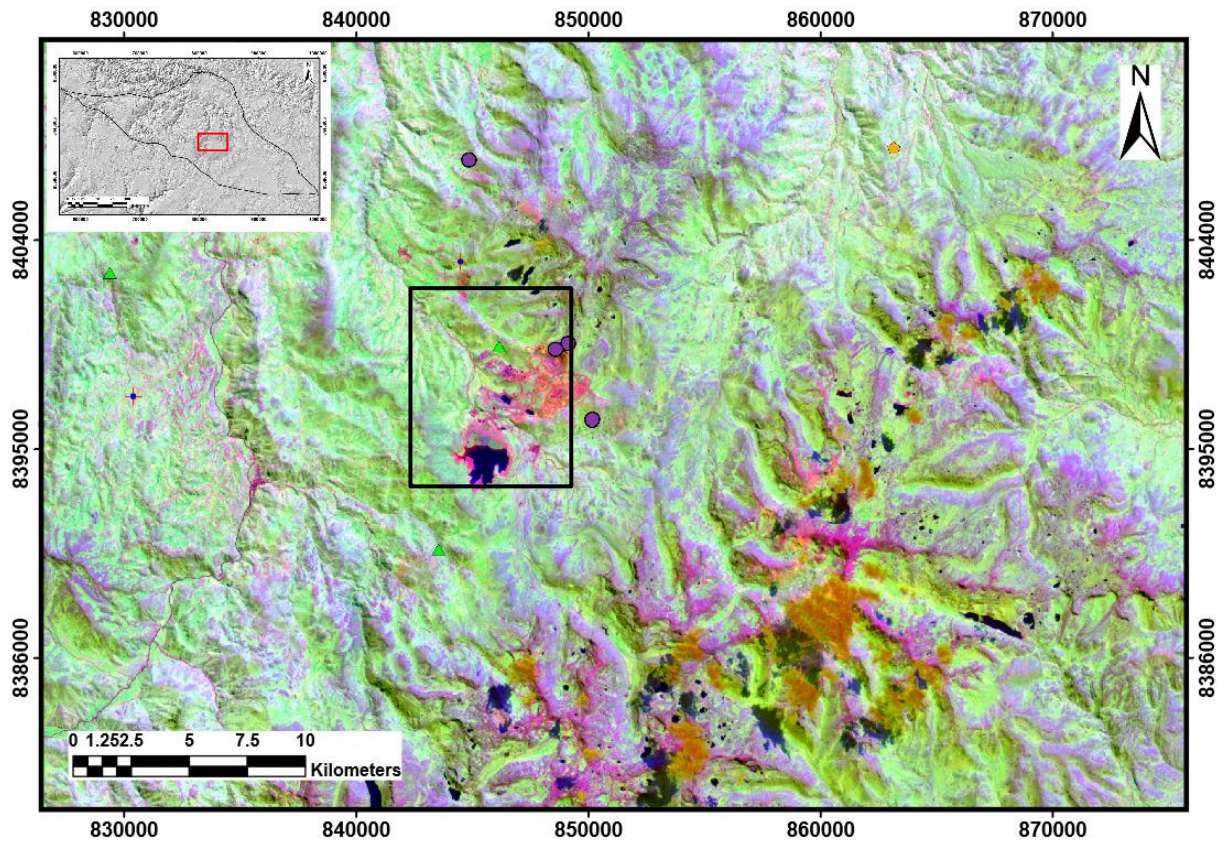
Además de las zonaciones de óxidos de hierro presentes en la zona, la mina Las Bambas comparte estas mismas características de reflectancia en sus píxeles, lo que fue usado como validación para las tonalidades previamente generadas en laboratorio, es decir, confirma que las relaciones de banda utilizadas hasta el momento, guardan grandes similitudes con yacimientos confirmados.

**Figura 54.** Vista panorámica hacia el noroeste del depósito Ferrobamba (Tomado de Sanchez *et al.*, 2019).



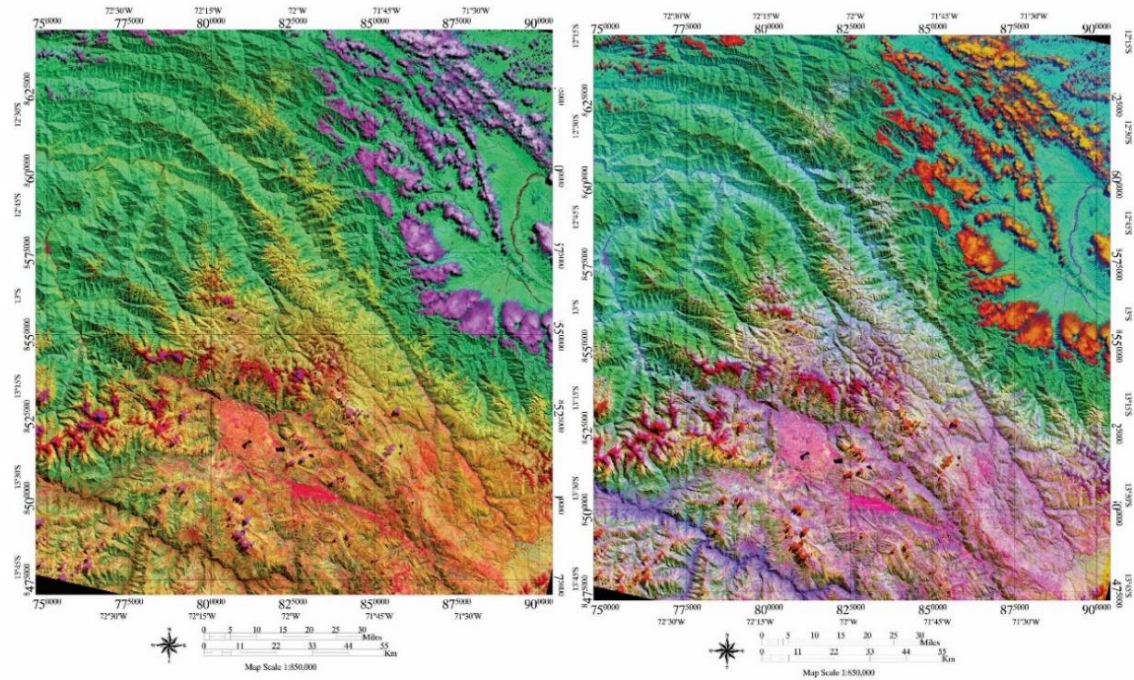
La mina Constanza **Figura 54** se ubica a 100 km al sur de la ciudad de Cusco, las coordenadas centrales son 850190E / 839627N (WGS84, UTM zona 18S) a una altitud de 4000 a 4500 m.s.n.m. Metalogénicamente forma parte de la franja de pórfidos-skarn de Cu-Mo-Au del Mioceno relacionadas al magmatismo del Batolito AndaHuaylas-Yauri (Acosta *et al.*, 2019). La litología está formada por múltiples fases de monzonita y pórfidos monzoníticos, los cuales intruyen a las secuencias sedimentarias del Cretáceo, compuestas por las formaciones Murco (areniscas y lutitas) y Arcurquina (calizas). Estas formaciones hospedan la mineralización económica del depósito.

**Figura 55.** Acercamiento de la imagen Landsat-8 LC80040702020150LGN00 mostrando en detalle yacimientos de skarns (Círculos morados) y vetas (triángulos amarillos). La mina conocida como Constanca se encuentra delimitada por el rectángulo negro.

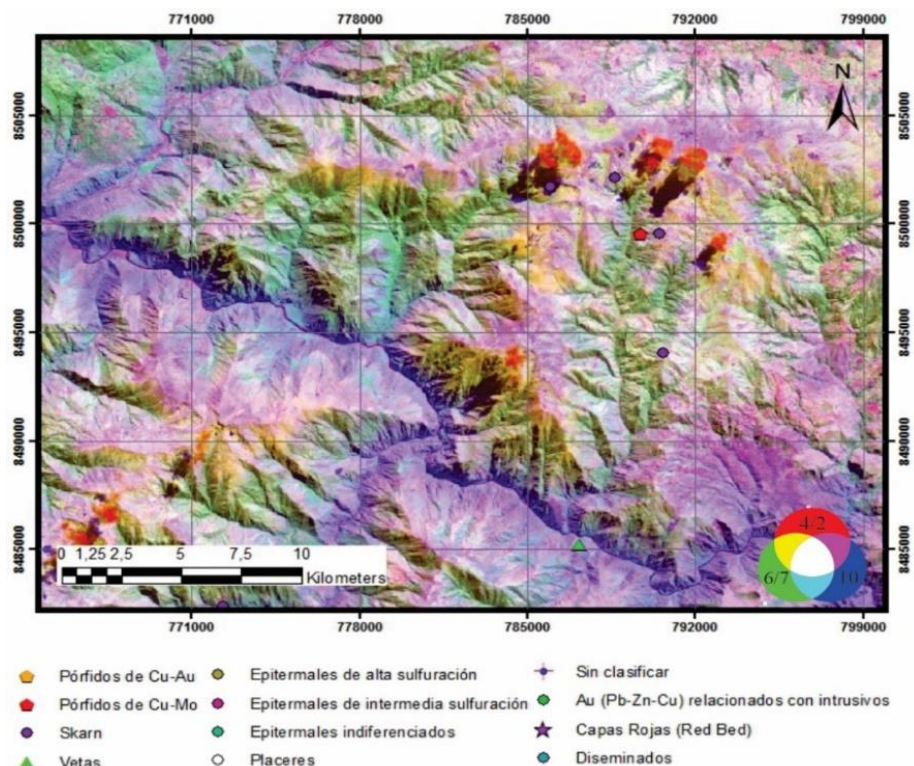


La penúltima imagen tratada muestra una gran zona con colores vibrantes al sur, naranjas rojizos para la combinación de bandas RGB 4/2,6/7,5 y los típicos fucsias y lilas intensos en RGB 4/2,6/7,10. La presencia de skarns es una constante en el lugar, debido a la continua interacción de intrusivos con calizas, acá tampoco fue la excepción ya que se observa en la **Figura 55** por lo menos cinco yacimientos confirmados de este tipo.

**Figura 56.** Imagen Landsat-8 LC80040692020150LGN00 con combinaciones de bandas RGB 4/2,6/7,5 y RGB 4/2,6/7,10 respectivamente.

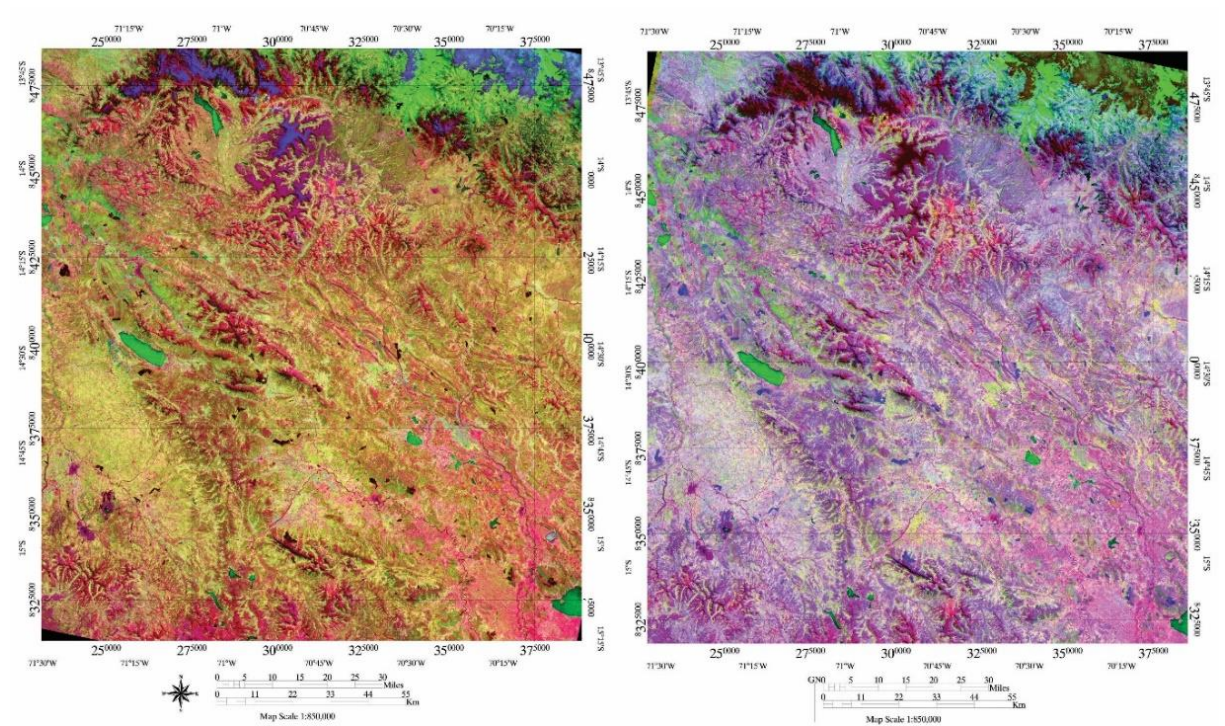


**Figura 57.** Acercamiento de la imagen Landsat-8 LC80040692020150LGN00 mostrando en detalle depósitos clasificados como vetas, pórfidos de Cu-Mo y skarns.



Finalmente en la parte más sureste de la franja se encuentran los distritos mineros de Antapaccay y Tintaya (**Figura 57**) los cuales se encuentran relativamente cerca el uno del otro. El Yacimiento Minero Antapaccay se localiza a 10 Kms. al Sur Oeste de la Mina Tintaya de la comunidad Alto Huarca, del Distrito y Provincia de Espinar, Departamento de Cusco al Sureste de Perú., a una altitud promedio 3800 m.s.n.m. La Mina Antapaccay es un yacimiento de tipo pórfido – skarn de Cu que ha sido explorada en diferentes periodos hasta la actualidad (Ccapa, 2019). Estratigráficamente consiste de la secuencia sedimentaria cretácica, correlacionable con la parte superior del Grupo Yura, durante las deformaciones andinas y ampliamente intruidas por la tectónica andina e, intruidas por stocks, sills y diques del Batolito Andahuaylas – Yauri, y finalmente cubierta por depósitos lacustrinos, volcánicos del Cenozoico y depósitos Cuaternarios (Ccapa, 2019).

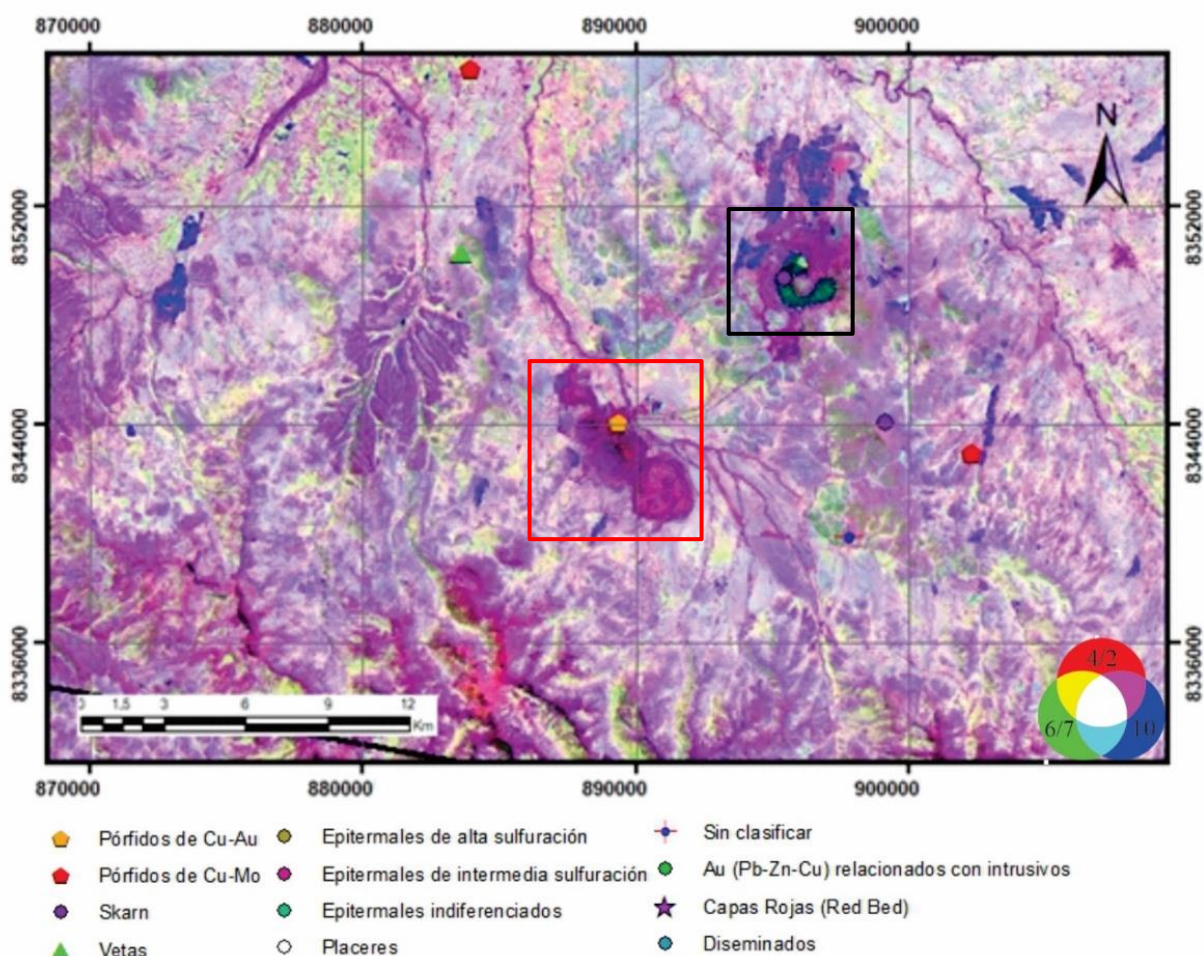
**Figura 58.** Imagen Landsat-8 LC80030702020239LGN00 con combinaciones de bandas RGB 4/2,6/7,5 y RGB 4/2,6/7,10 respectivamente.



El distrito minero de Tintaya (**Figura 59**) está ubicado en la terminación sur de la franja

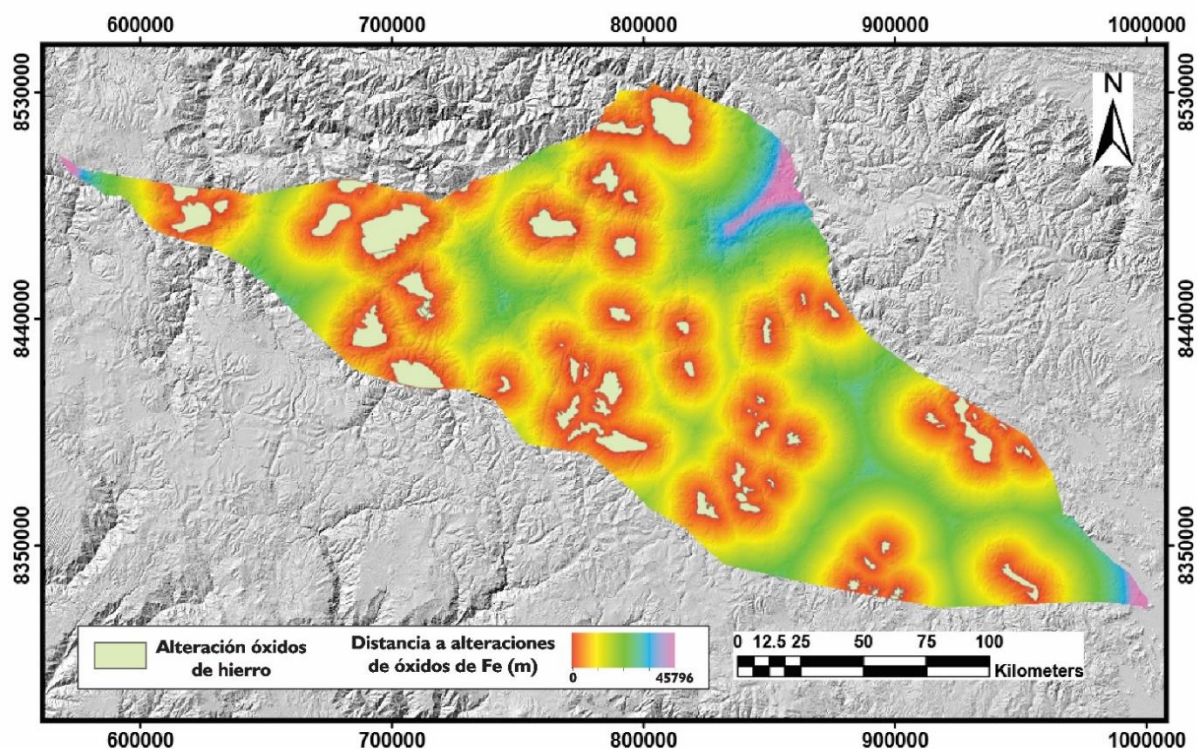
metalogenética de Andahuaylas-Yauri. La mineralización en este cinturón está asociada espacial y temporalmente al Batolito Andahuaylas-Yauri, de composición calco-alcalina y de edad Eoceno medio–Oligoceno inferior (~45-30 Ma) (Perelló *et al.*, 2003).

**Figura 59.** Acercamiento de la imagen Landsat-8 LC80030702020239LGN00 mostrando en detalle yacimientos de skarns, los distritos mineros de Antapaccay (delimitado por el rectángulo rojo) y Tintaya (delimitado por el rectángulo negro).



Las zonas encontradas con alteraciones de óxidos de hierro se muestran en la **Figura 60**, estas se digitalizaron en formato shape para facilitar la aplicación del modelo de RNA.

**Figura 60.** Mapa de alteraciones de óxidos de hierro. Se delimitaron las zonas en respuesta al procesamiento de las imágenes satelitales. El mapa muestra la distancia desde cualquier punto a una zona de alteración de óxidos de hierro. Las zonas más alejadas se encuentran al noreste en el dominio de la Cordillera Oriental.



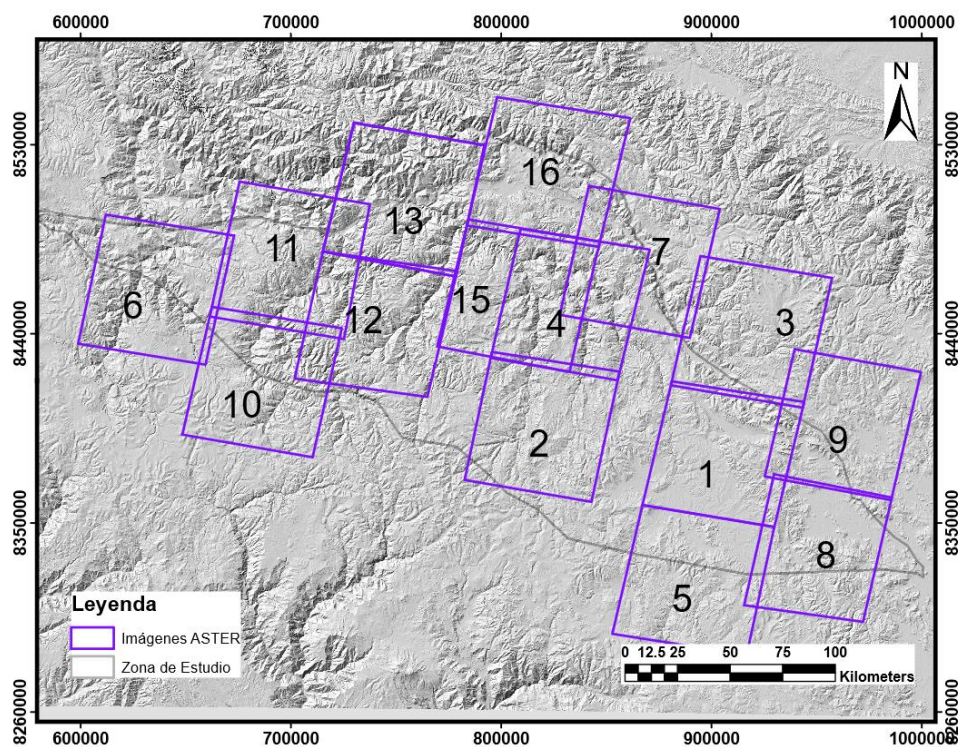
### 6.1.2 Procesamiento de imágenes ASTER

El análisis con imágenes ASTER se realizó posterior a los resultados de anomalías geoquímicas, imágenes Landsat-8 y de características geológicas, donde se preestablecieron zonas de interés. Se utilizaron un total de 16 imágenes (**Figura 61**) con la menor cantidad posible de nubes con los códigos:

- (1) AST\_L1T\_00303172008151024\_20150523095234\_89351,
- (2) AST\_L1T\_00303062004151107\_20150503130413\_26311
- (3) AST\_L1T\_00303172008151015\_20150523095234\_89355
- (4) AST\_L1T\_00303062004151058\_20150503130413\_26315

- (5) AST\_L1T\_00303172008151033\_20150523095237\_57590
- (6) AST\_L1T\_00304202006151608\_20150514013223\_104437
- (7) AST\_L1T\_00304222006150346\_20150514021534\_30069
- (8) AST\_L1T\_00305242006150354\_20150514130535\_70214
- (9) AST\_L1T\_00305242006150345\_20150514130531\_70089
- (10) AST\_L1T\_00306162006151012\_20150514211837\_67905
- (11) AST\_L1T\_00306162006151003\_20150514211830\_122086
- (12) AST\_L1T\_00306262007151646\_20150520014054\_38880
- (13) AST\_L1T\_00306262007151637\_20150520014054\_38887
- (14) AST\_L1T\_00306262007151637\_20150520014054\_38887
- (15) AST\_L1T\_00311102007151004\_20150522004936\_95790
- (16) AST\_L1T\_00311102007150955\_20150522004926\_95565

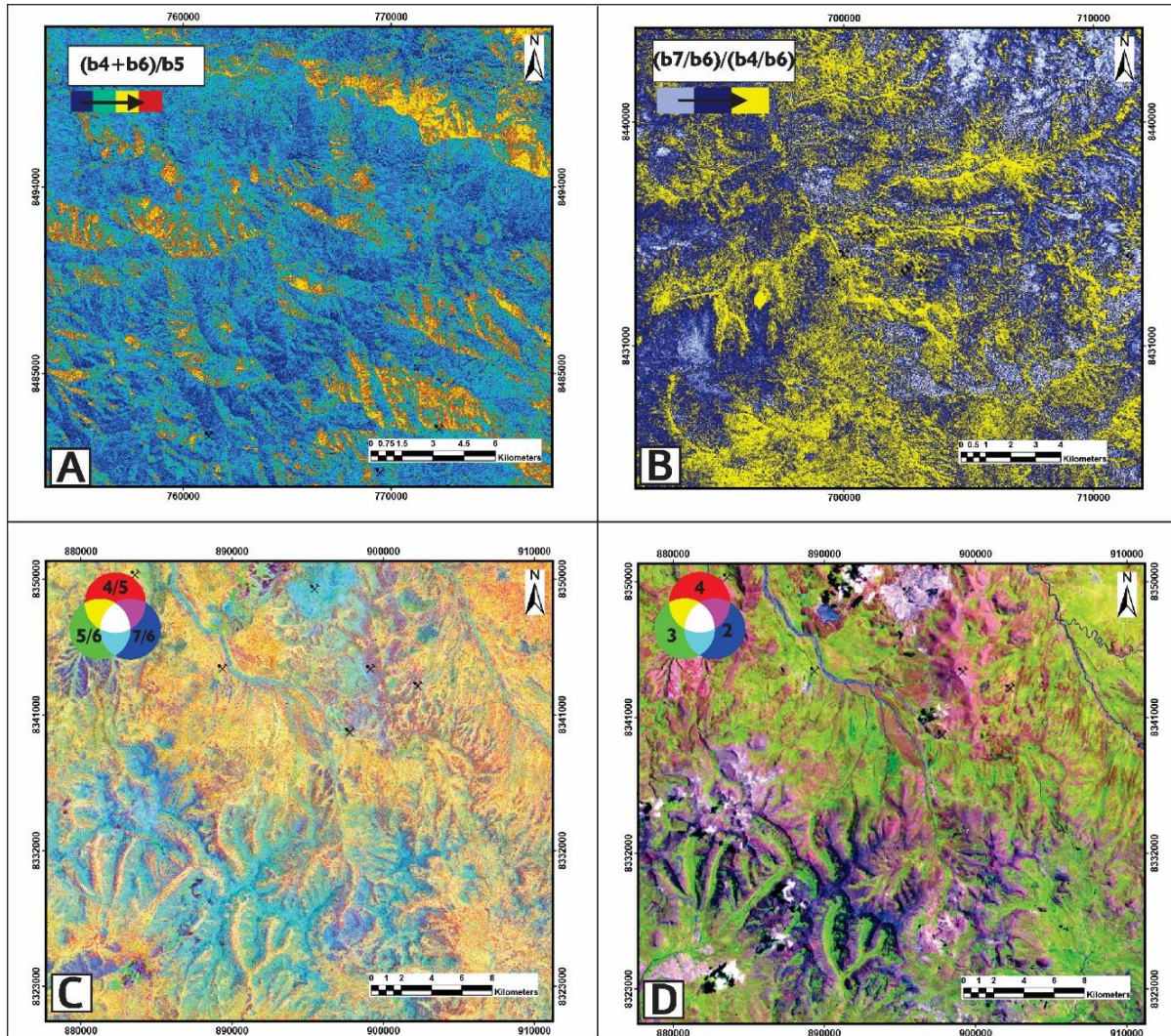
**Figura 61.** *Ubicación de Imágenes ASTER utilizadas. Estas se escogieron posterior a los análisis de imágenes Landsat-8 y de los análisis geoestadísticos y geológicos, donde se predeterminaron zonas de interés.*



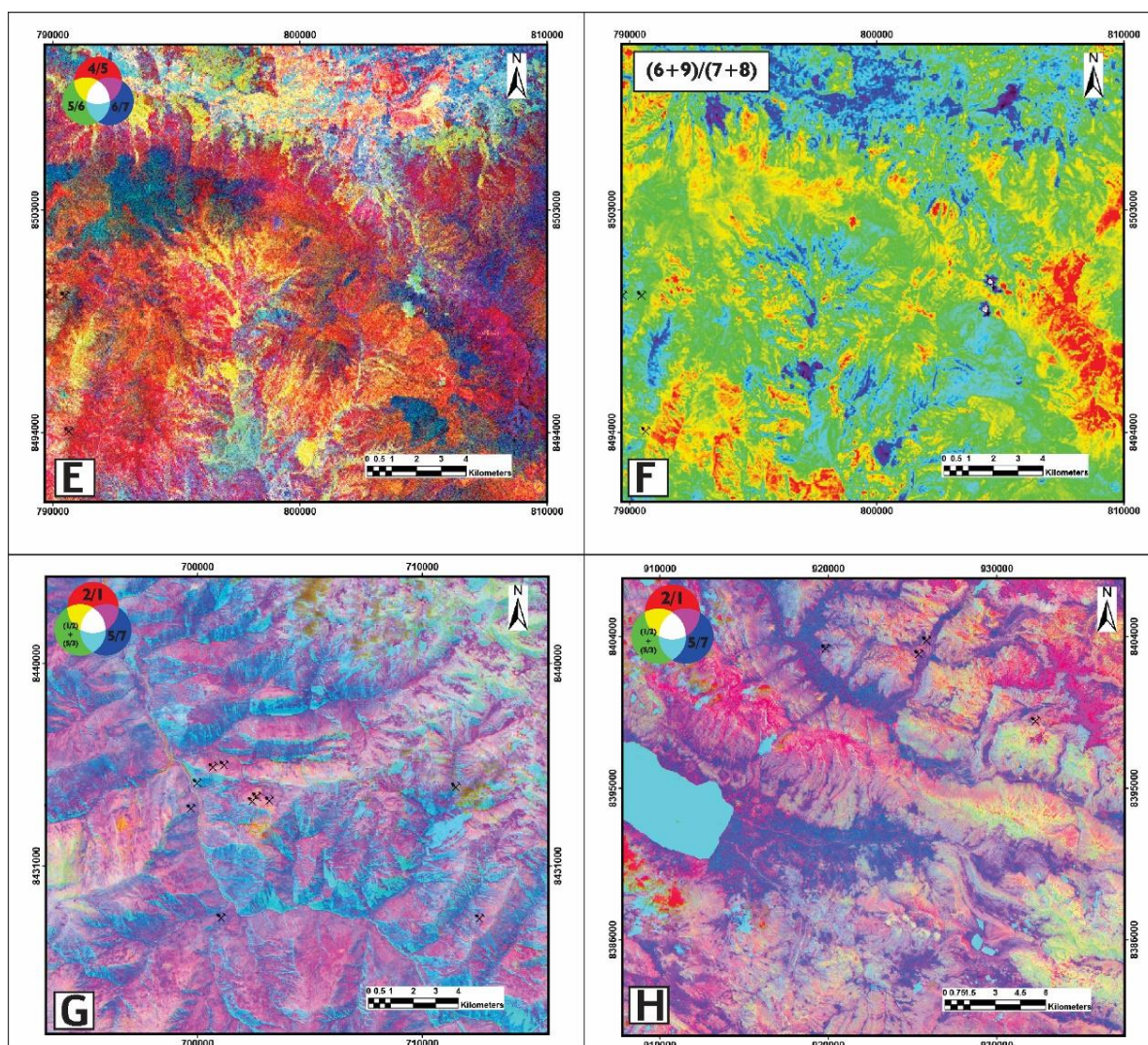
Las relaciones y combinaciones de bandas se realizaron utilizando las compiladas por Kalinowski & Oliver (2004) en el manual de procesamiento de índices minerales en ASTER. También se utilizaron los trabajos de (Mars & Rowan, 2006; Ninomiya, 2002, 2004; Sabins, 1999; Volesky *et al.*, 2003).

A continuación se mostraran algunas (**Figura 62-63**) de las combinaciones y relaciones de bandas realizadas para resaltar y mapear alteraciones hidrotermales, principalmente de tipo Argílica avanzada, propílica, argílica, calcosilicatada y fílica , estas se utilizaron ya que están estrechamente relacionadas con yacimientos de tipo Pórfidos de cobre (Mars & Rowan, 2006).

**Figura 62.** A. índice de alteración Argílica avanzada, las zonas amarillas-rojas indican la alta absorción de los minerales Alunita-pirofilita en la banda 5 (Mars & Rowan, 2006). B. índice de minerales alterados con contenido de OH (OHI). C. Composición en falso color para cálculo de alteración fílica. D. Combinación 432 de la misma zona de la imagen C.



**Figura 63.** E. combinación en falso color RGB para identificar alteración argílica, R: 4/5, G:5/6, B:6/7. Los yacimientos confirmados en la zona sirvieron de guía para identificar las tonalidades asociadas. F. la relación  $(6+9)/(7+8)$  se utilizó para encontrar zonas con contenido de epidota, clorita, anfíbolita, las zonas de mayor reflectancia se observan en rojo y están cerca de yacimientos de tipo skarn confirmados. G y H. Relaciones de banda en RGB para identificar zonas de óxidos férricos ( $2/1$ ). Óxidos ferrosos ( $1/2 + 5/2$ ) y arcillas 5/7.

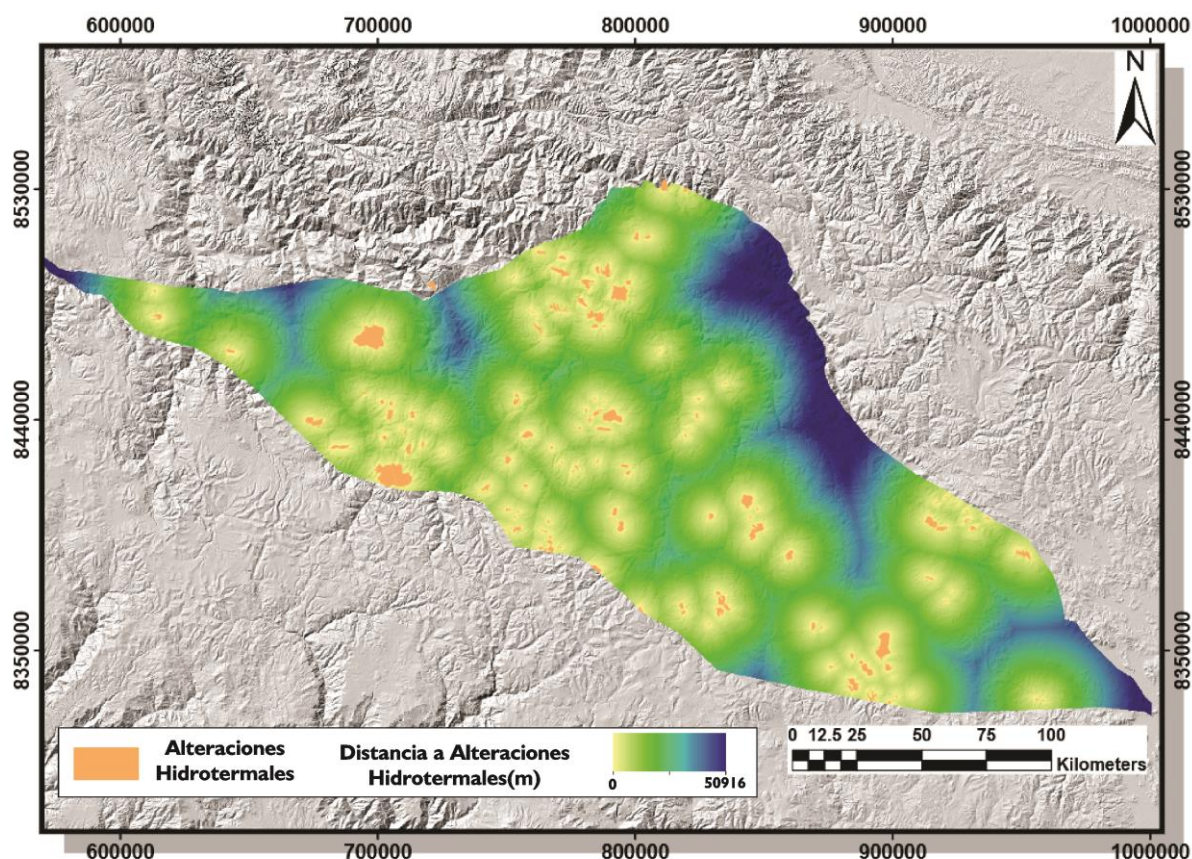


El análisis con imágenes Aster fue complejo debido a que en varias zonas en las que se encuentran yacimientos confirmados estos no mostraban respuestas de alteraciones hidrotermales en superficie como el caso de la mina Corocohuyco donde las mineralizaciones están por debajo de los 200 metros, cubiertas por depósitos coluviales (Maher, 2010), Las

principales zonas de alteración se encontraron en los intrusivos relacionados al Batolito de Abancay entre estos se resaltan en la zona suroeste las tonalitas de Plutones Oyoni, Negro Cocha, las calizas de las formaciones Arcurquina y Socosani (Cerpa *et al.*, 2004).

Las alteraciones encontradas **Figura 64** fueron principalmente argílica avanzada, fílica y propilítica, que corresponden a las más comunes en los yacimientos de tipo pórfido y skarn asociados al Batolito de Abancay (Carlotto *et al.*, 2009; Perelló *et al.*, 2003).

**Figura 64.** Mapa de alteraciones hidrotermales. Algunas fueron tomadas de la bases de datos del Geocatmin, (2019).

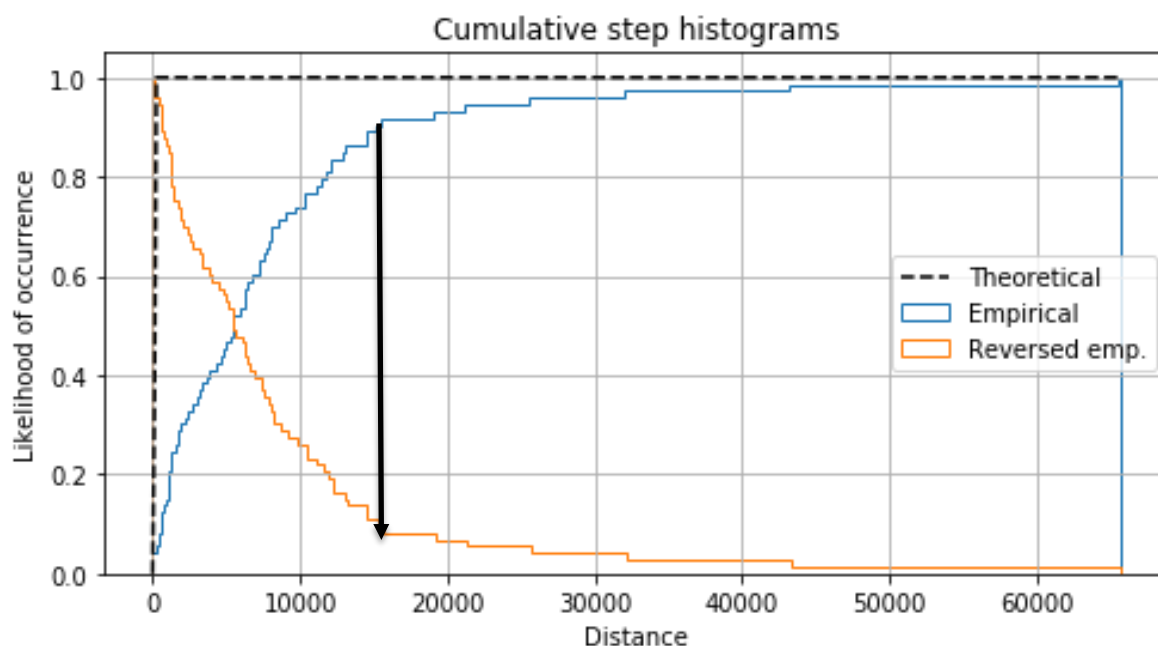


### 6.3 Machine Learning

#### 6.3.1 Resultados de la adquisición y preprocesamiento de datos de entrada

Los puntos de entrenamiento se escogieron teniendo en cuenta la metodología de Carranza 2009, quien plantea que el tamaño de celda del pixel sea menor a la distancia mínima entre yacimientos, con esto se garantiza que no haya más de un yacimiento por pixel. Las distancias entre yacimientos se calcularon mediante la generación de un código de patrones de puntos en Python (**Figura 65**) donde se calculó que la distancia mínima entre yacimientos pertenecientes a la zona de estudio es de 275 metros. También se realizó un filtrado a los depósitos tipo placer y redbeds ya que estos no guardan una relación directa con los sistemas mineralizantes de tipo pórfidos y Skarn.

**Figura 65.** Resultado del análisis de patrón de puntos mostrando la distancia de vecinos cercanos y la probabilidad de encontrar un nuevo yacimiento desde otro ya identificado. Distancia mínima[m]: 275.072, distancia máxima[m]: 65730.036.

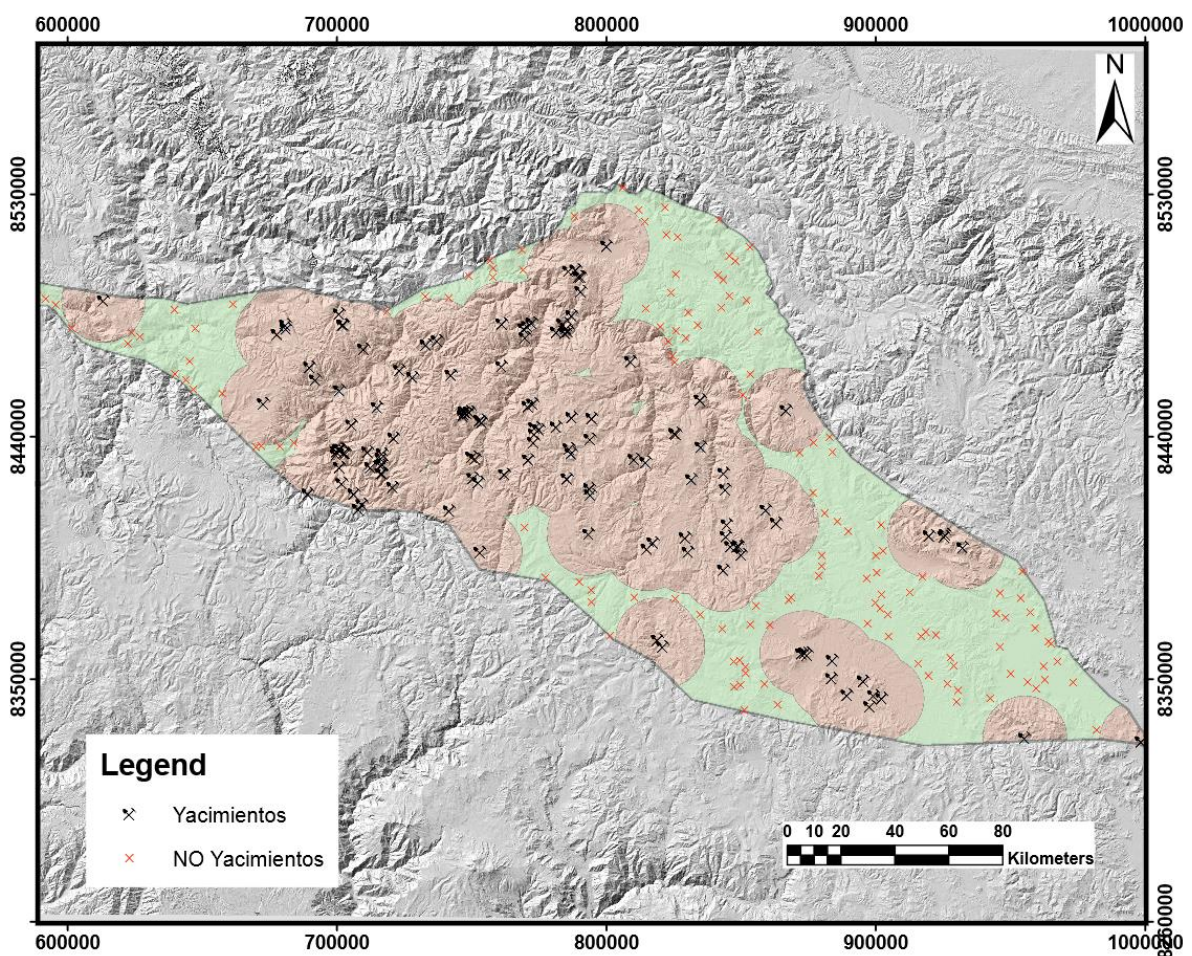


En la gráfica se observa que la distancia máxima entre un par de depósitos cualquiera, es de 65730 metros, es decir que hay un cien por ciento de probabilidad de ocurrencia de otro yacimiento dentro de este rango. En otras palabras, encontrar depósitos separados más allá de 65730 metros es improbable. Sin embargo, son pocos los puntos que se hallan en este rango.

En su lugar se seleccionó 15591 metros como distancia de *buffer* dentro del cual hay una probabilidad del 89% de encontrar un yacimiento vecino junto a cualquier depósito dado, esto se observa en la **Figura 65** cuando la curva comienza a aplanarse y la cantidad de yacimientos disminuye.

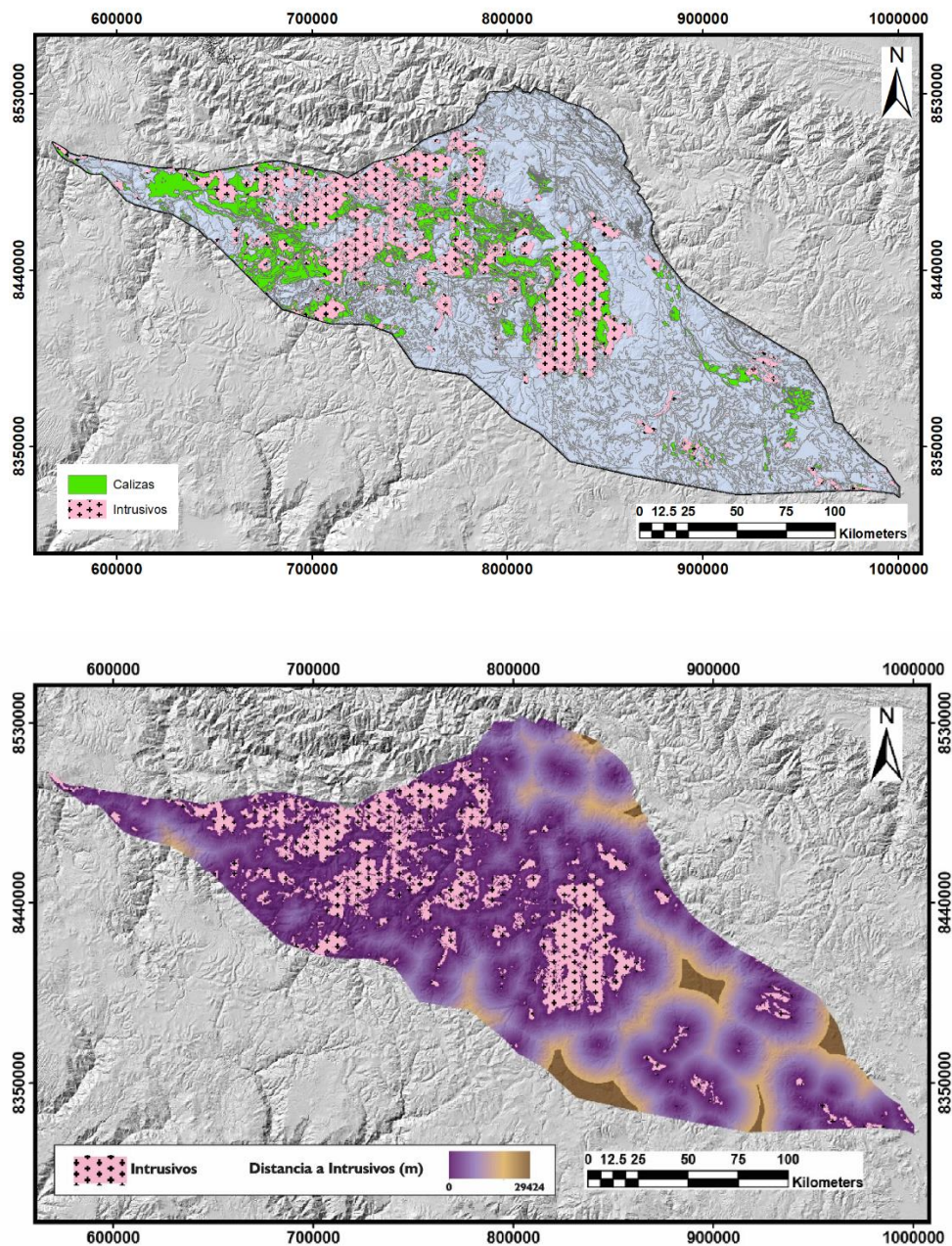
Con estos análisis estadísticos se calcularon las áreas de dominio que delimitaron las zonas de yacimiento y no yacimiento (**Figura 66**), proceso de vital importancia para el entrenamiento supervisado del modelo usado.

**Figura 66.** Localización de depósitos y selección de puntos de no depósito. Se realizó un *buffer* de 15.591 metros que representa el límite. Los puntos de no yacimiento se generaron mediante la herramienta de ArcGis Create Random Points, fuera del *buffer* realizado.



Una vez definidos los puntos de yacimiento y no yacimiento para el entrenamiento supervisado, los datos fueron tratados en el software de ESRI, ArcGis, donde se visualizaron y generaron mapas de distancia a fallas y lineamientos, pliegues, intrusivos, contactos de intrusivos y calizas, alteraciones hidrotermales, alteraciones de óxidos de hierro como se observan en las **Figura 67-72** con la herramienta *Euclidean Distance* la cual describe la relación de cada celda con un origen o un conjunto de orígenes basándose en la distancia de la línea recta (ESRI, 2019). La información para realizar los mapas se tomó de planchas a escala 1:100k del INGEMMET. También se utilizó un mapa de densidad de fallas junto con los mapas de anomalías geoquímicas creados y presentados en los resultados de geoestadística y Machine Learning,

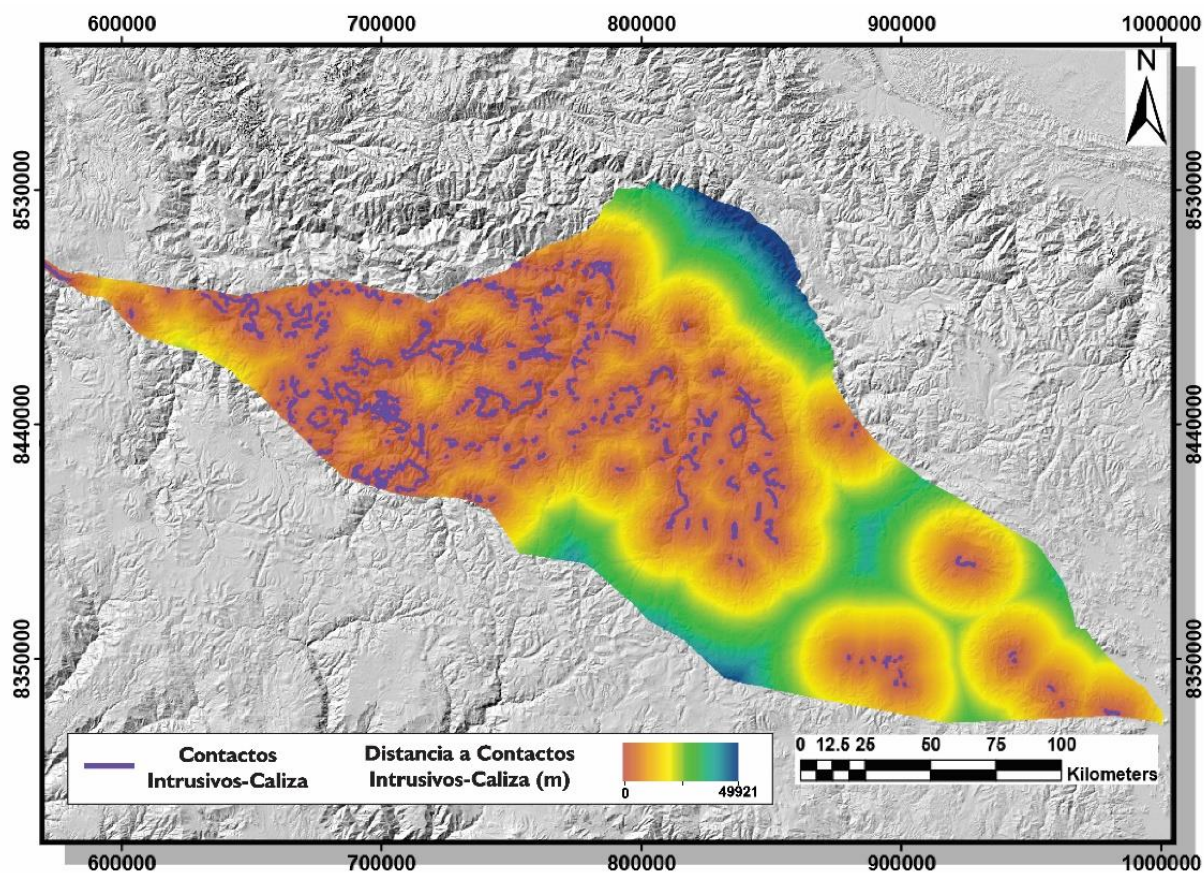
**Figura 67.** Arriba. Mapa geológico utilizado para el cálculo de distancias a intrusivos y contactos Intrusivos-Calizas donde se resaltan los intrusivos (relacionados al batolito de Abancay) y las calizas (principalmente de Ferrobamba), en azul las demás litologías. Abajo. Mapa de proximidad a intrusivos aflorando en la zona de estudio. Los datos fueron tomados del Mapa Geológico del Perú a escala 1:100000 del INGEMMET.



Los intrusivos de la zona pertenecen al Batolito de Andahuaylas- Yaurí (BHY) ,constan de 2 etapas, la primera son rocas máficas (principalmente gabros de la unidad Cotabambas-Plutón Curahuasi) y dioritas (de las unidades Lambrama, Llajua, Parco, Progreso, Ocobamba y Coyllurqui) emplazadas entre ~48 y ~43 Ma (Cerpa *et al.*, 2004). Estas unidades se encuentran intruidas a su vez por las rocas de la segunda etapa que son pulsos de granodioritas (Plutones Porgreso, Parco, Huasani) y cuarzo-monzodioritas (Plutones Ocobamba y Taquebamba) emplazadas entre ~42 y ~30 Ma (Perelló *et al.*, 2003).

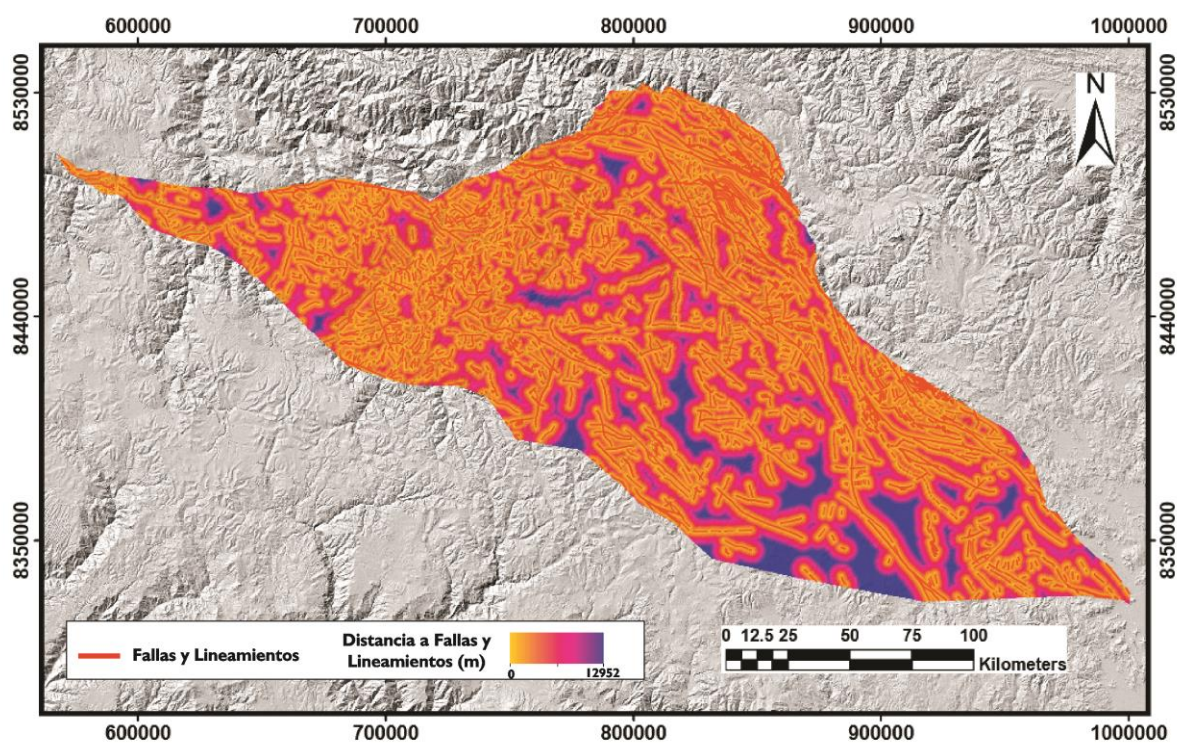
En la franja aflora el borde noroeste del BHY donde recibe el nombre de Batolito de Abancay San Miguel o Batolito de Apurímac (Cerpa *et al.*, 2004; Rivera *et al.*, 2011). El BHY es de composición calco alcalina y edad Eoceno-Oligoceno inferior (Carlotto, 1998)

**Figura 68.** Mapa de proximidad a contactos Intrusivos-Calizas. Se realizó en base al Mapa Geológico a escala 1:100.000 del Perú.



Los cuerpos intrusivos que afloran en la Franja Metalogenética XV se encuentran emplazando dentro de rocas sedimentarias clásticas de la Formación Mara (Aptiano), el Grupo Yura (Jurásico-Cretácico) y las calizas de Ferrobamba (Albiano-Turoniano; Carlotto *et al.*, 2009). La gran cantidad de contactos entre intrusivos con calizas como roca encajante genera un ambiente propicio para las ocurrencias de yacimientos de tipo Pórfidos y Skarn.

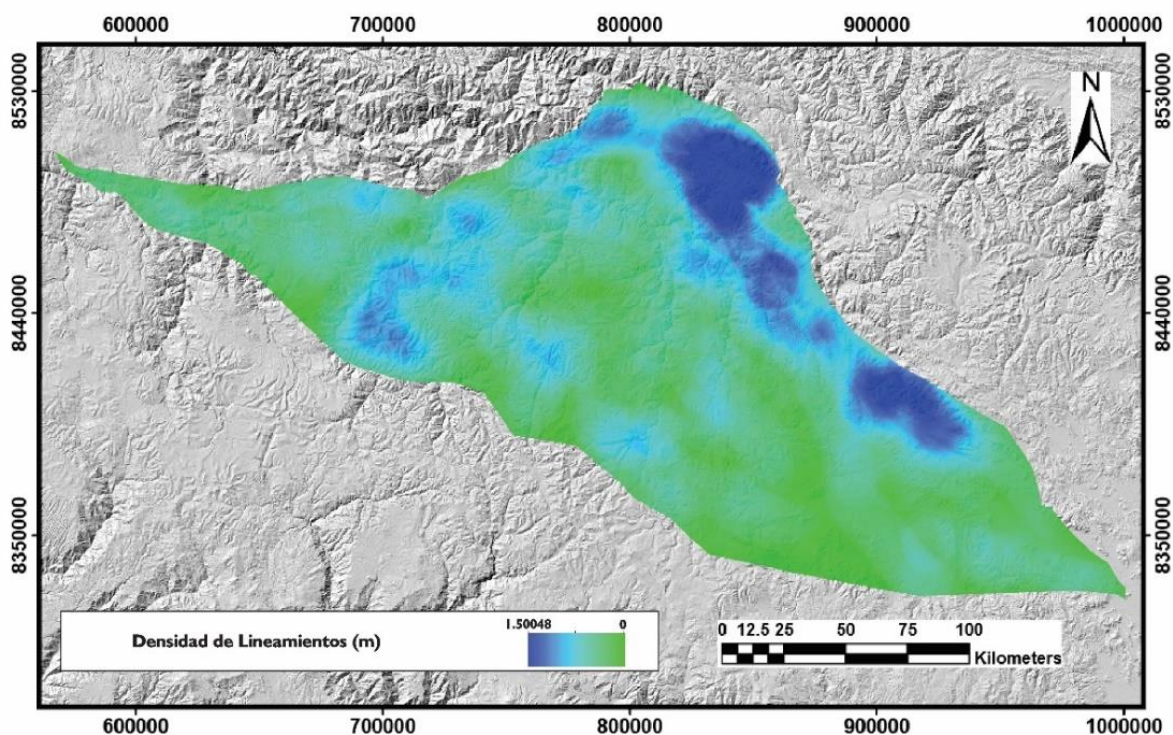
**Figura 69.** Los datos se adquirieron de Mapas Geológicos de la zona a escala 1:1M, 1:100K y 1:50K. Esa información se completó realizando una extracción de lineamientos en Arcgis con las ayuda de imágenes satelitales y DEMs.



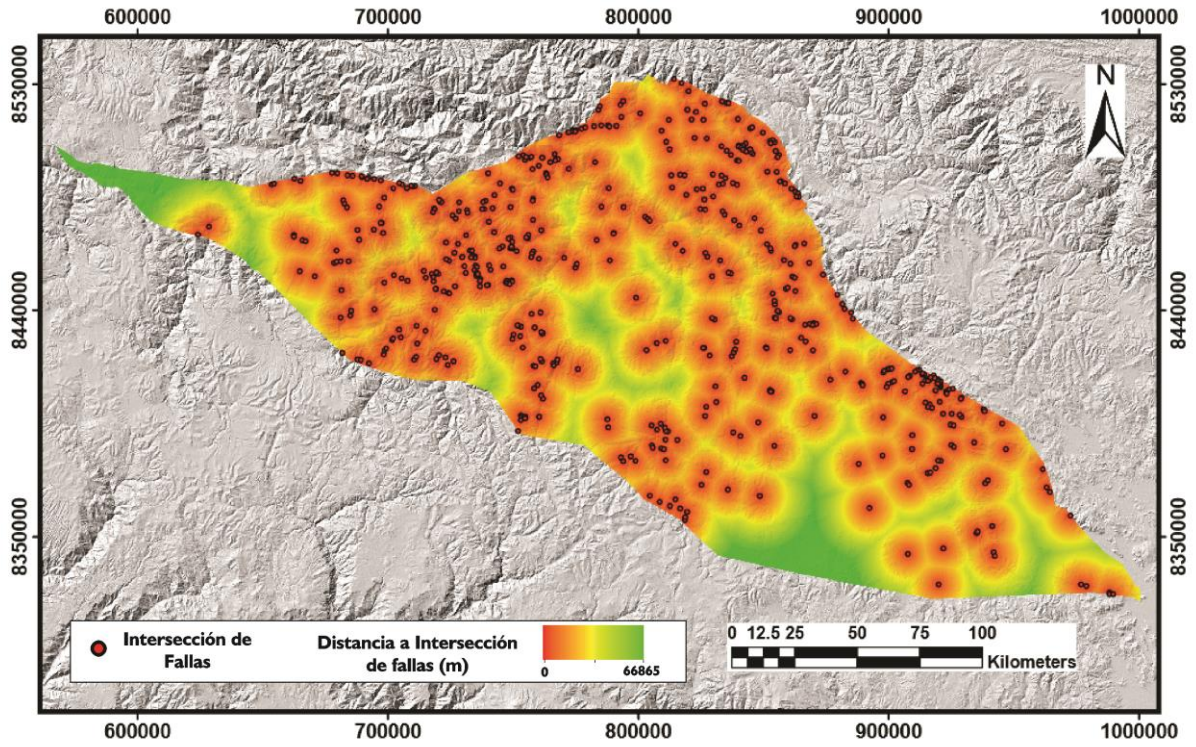
En la franja el comportamiento de las fallas está controlado por los distintos dominios estructurales divididos en: Cordillera Oriental (CO), Cordillera Occidental (CO), y los Altiplanos Oriental (ALO) y Occidental (AIOC) delimitados por los sistemas de fallas Urcos-Sicuani-Ayaviri, Cusco-Lagunillas-Mañazo, Abancay-Andahuaylas-Totos-Chincheros-Lincapa (Carlotto *et al.*, 2005). La zona noreste y sureste donde se encuentran los dominios

CO, ALO y ALOC presenta fallas principalmente con cinemática normal con componente sinistral, dirección preferencial NW-SE y buzamientos de entre 65°-70°, las principales son las fallas Tambomacha, Sangarará, Capi, San Lorenzo y Angostura-Condorama de los sistemas de fallas Zurite-Cusco-Urcos-Sicuani u Casacunca-Acomoyo-Langui-Layo (Cerpa *et al.*, 2004; INGEMMET, 1995). Los lineamientos extraídos en Arcgis siguen esta tendencia NW-SE. La parte suroeste donde se encuentra el dominio de la CO algunas fallas normales tienen una tendencia SO como los del sistema de fallas Zona Espinar, los lineamientos extraídos no seguían ninguna tendencia en específico. Para el análisis no se tuvieron en cuenta las fallas inversas pertenecientes a los sistemas de fallas del Altiplano, Accos y Zurite-Cusco-Urcos-Sicuani, ya que estas no son tan eficaces como conductoras de fluidos ricos en minerales como los régimen transtensionales o extensionales de las fallas normales (Hulin, 1929).

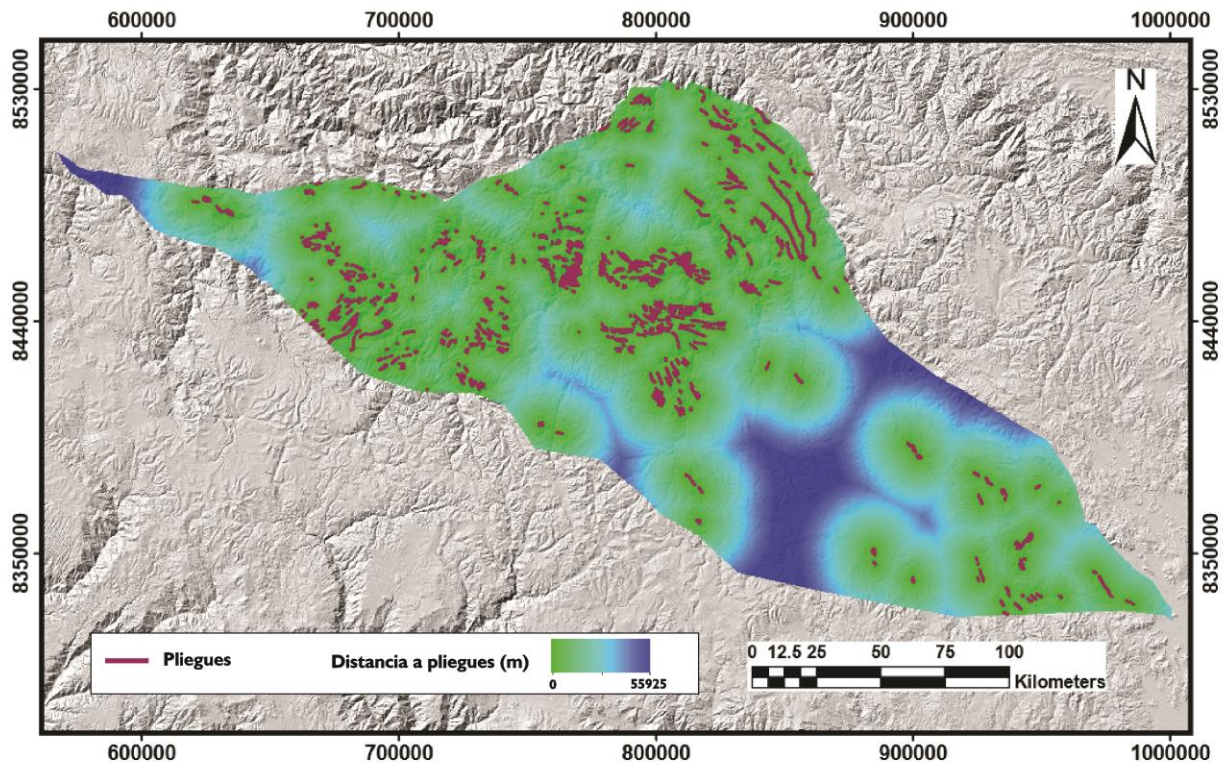
**Figura 70.** Mapa de densidad de lineamientos. Se generó con las fallas y lineamientos de la Figura 69.



**Figura 71.** Mapa de proximidad a intersección de fallas en roca encajante (calizas). Se utilizaron datos del mapa geológico del Perú a escala 1:50.000.



**Figura 72.** Mapa de proximidad a pliegues, la información se tomó del mapa geológico del Perú a escala 1:100k del INGEMMET. Yacimientos de Skarn de cobre como el Coroccohuayco en el distrito de Tintaya y el de Accha, las mineralizaciones se encuentran en la zona de charnela o flancos pliegues (Boni et al., 2009; Maher, 2010).



### 6.3.2 Resultados de la aplicación del modelo

A continuación se mostrará parte de los códigos de entrenamiento de la red (**Figura 73**) y de las predicciones (**Figura 74**).

**Figura 73.** Código realizado en Jupyter Notebook de entrenamiento de la red neuronal artificial.

```

"""# 2. Carga de datos del set de entrenamiento y asignación de "x" y "y"""
data = pd.read_csv("/content/sample_data/yacinoayaci.csv")
print(data.head())

#Verificar que se seleccionen las columnas adecuadas en el siguiente paso
df_x = data.iloc[:,5:24]
df_y = data.iloc[:,4]

print(df_x.head())
print(df_y.head())

"""# 3. Preprocesamiento

## 3.1 Normalización
"""

scaler = MinMaxScaler()
Xscaled= scaler.fit_transform(df_x)
Yscaled= df_y.to_numpy()

#Guarda los parámetros de escalado
dp(scaler, open("/content/sample_data/scaler.pkl", "wb"))

"""## 3.2 Dividir dataset entrenamiento y validación"""
X_train, X_test, y_train, y_test = train_test_split(Xscaled, Yscaled, test_size=0.2, random_state = 4)

"""# 4. Modelo ML

### 4.1 Entrenamiento
"""

#Instancia del modelo ML
clf= MLPClassifier(hidden_layer_sizes=(6,100), max_iter=1000, activation='logistic',
                  solver='adam', alpha=0.0001,learning_rate='constant',
                  learning_rate_init=0.001,random_state=0, verbose=True)

#Entrenamiento o ajuste del modelo
clf.fit(X_train,y_train)

"""## 4.2 Verificación de resultados"""

#Imprime en pantalla resultados del modelo
print("Predicción:\n",clf.predict(X_test),"\n")
print("Data_entrada\n",y_test,"\n")
print("Probabilidades\n",clf.predict_proba(X_test),"\n")
trainscores = clf.score(X_train,y_train)
testscores = clf.score(X_test,y_test)

# report performance
print('Accuracy on train set: %.3f (%.3f)' % (mean(trainscores), std(trainscores)))
print('Accuracy on test set: %.3f (%.3f)' % (mean(testscores), std(testscores)))

precision_score(y_test, clf.predict(X_test))

"""## 4.3 Exporte del modelo"""
"""## 5.3 Predicción y exporte"""

predict=clf.predict(Xall_scaled)
print(predict)
proba=clf.predict_proba(Xall_scaled).transpose()[0]
probi=clf.predict_proba(Xall_scaled).transpose()[1]
print(probi)

#Creación de dataframe con predicciones
result=pd.DataFrame({
    "Predicción":predict,
    "Probabilidad_0":proba,
    "Probabilidad_1":probi
})

print(result.head())

data_export=data_all_1.copy()
data_export=data_all_1.reset_index()
print(data_export)

#Creación de nuevo dataframe con la unión de dataframe de entrada con dataframe de predicciones
data_export= data_export.join(result)
print(data_export)

#Exportada de nuevo dataframe
data_export.to_csv("/content/sample_data/yaciclassified.csv")

```

**Figura 74.** Código realizado en Jupyter Notebook de predicción del algoritmo de RNA.

```

2. Carga de datos del set de entrenamiento y asignación de "X" y "y"

In [24]: data = pd.read_csv("/content/sample_data/yacinoayaci.csv")
print(data.head())

#Verificar que se seleccionen las columnas adecuadas en el siguiente paso
df_x = data.iloc[:,5:24]
df_y = data.iloc[:,4]

   FID  OBJCTID  PDMT_X ...  Idw_Plomo  Idw_Niquel  Idw_Arseni
0  0      180  808085.476656 ...  11.656799  14.242780  10.320281
1  1      144  821745.085196 ...  27.479700  19.426800  11.352300
2  2      145  812035.086242 ...  18.134000  16.617901  9.847940
3  3      137  750235.129797 ...  22.750200  19.634100  10.130900
4  4      216  841838.316754 ...  22.169201  18.495500  10.104400

[5 rows x 24 columns]

In [ ]: print(df_x.head())
print(df_y.head())

   E_fall1Lin  E_Interval  Euc_Pilegu ...  Idw_Plomo  Idw_Niquel  Idw_Arseni
0  254.918084  8262.259766  3076.129317 ...  11.656799  14.242780  10.320281
1  687.003000  2631.640003  1990.000000 ...  27.479700  19.426800  11.352300
2  2470.120000  8156.589244  5963.438276 ...  18.134000  16.617901  9.847940
3  447.123300  3626.260287  2461.570600 ...  22.750200  19.634100  10.130900
4  105.113900  4970.100150  803.094971 ...  22.169201  18.495500  10.104400

3. Preprocesamiento

3.1 Normalización

In [14]: scaler = MinMaxScaler()
Xscaled= scaler.fit_transform(df_x)
Yscaled= df_y.to_numpy()

#Guarda los parámetros de escalado
dp(scaler, open("/content/sample_data/scaler.pkl", "wb"))

3.2 Dividir dataset entrenamiento y validación

In [15]: X_train, X_test, y_train, y_test = train_test_split(Xscaled, Yscaled, test_size=0.2, random_state = 4)

4. Modelo ML

4.1 Entrenamiento

In [16]: #Instancia del modelo ML
clf= MLPClassifier(hidden_layer_sizes=(6,100), max_iter=1000, activation='logistic',
                  solver='adam', alpha=0.0001,learning_rate='constant',
                  learning_rate_init=0.001,random_state=0, verbose=True)

4.2 Verificación de resultados

In [18]: #Imprime en pantalla resultados del modelo
print("Predicción:\n",clf.predict(X_test),"\n")
print("Data_entrada\n",y_test,"\n")
print("Probabilidades\n",clf.predict_proba(X_test),"\n")
trainscores = clf.score(X_train,y_train)
testscores = clf.score(X_test,y_test)

# report performance
print('Accuracy on train set: %.3f (%.3f)' % (mean(trainscores), std(trainscores)))
print('Accuracy on test set: %.3f (%.3f)' % (mean(testscores), std(testscores)))

Predicción:
[8 1 1 1 1 0 0 1 0 0 0 0 1 0 1 0 0 0 1 1 0 1 1 1 1 0 1 0 1 0 1 0
 1 1 0 1 1 0 1 0 1 0 0 0 1 1 1 0]

Data_entrada
[0 1 0 1 0 0 0 0 0 0 1 0 1 0 0 0 1 1 0 1 1 0 1 1 0 0 0 1 0 1 0
 1 1 0 1 1 0 1 0 1 0 0 0 1 1 1 0]

5.3 Predicción y exporte

In [33]: predict=clf.predict(Xall_scaled)
print(predict)
proba=clf.predict_proba(Xall_scaled).transpose()[0]
probi=clf.predict_proba(Xall_scaled).transpose()[1]
print(probi)

#Creación de dataframe con predicciones
result=pd.DataFrame({
    "Predicción":predict,
    "Probabilidad_0":proba,
    "Probabilidad_1":probi
})

print(result.head())

data_export=data_all_1.copy()
data_export=data_all_1.reset_index()
print(data_export)

#Creación de nuevo dataframe con la unión de dataframe de entrada con dataframe de predicciones
data_export= data_export.join(result)
print(data_export)

#Exportada de nuevo dataframe
data_export.to_csv("/content/sample_data/yaciclassified.csv")

```

La generación de resultados comienza con la aplicación del modelo creado a los datos extraídos a partir de las 19 capas de información en el área total de estudio. En total fueron 14'569.042 puntos con la información contenida en píxeles de 50x50 metros.

No existe una relación ideal para seleccionar la cantidad de datos de entrenamiento-prueba, esta depende de las características de cada proyecto (Raschka, 2019). En este caso el entrenamiento se realizó con un set de datos del 80% del total y la prueba de la red neuronal

artificial con el 20% restante.

**Figura 75.** El código recibió el nombre de 'clf', 'MLPClassifier': es el nombre del algoritmo supervisado utilizado, 'activation': es la función de activación de las capas ocultas, 'solver': determina el peso de la optimización, 'alpha': es un parámetro de regular.

```
clf= MLPClassifier(hidden_layer_sizes=(3,), max_iter=1800,
                  activation='logistic', solver='adam', alpha=0.0001,
                  learning_rate='constant', learning_rate_init=0.001,random_state=0)
```

El entrenamiento de la red fue complejo, se requirieron múltiples cambios en los parámetros mostrados en la **Figura 75** para alcanzar un modelo que contara con predicciones confiables, el algoritmo alcanzó una precisión del 89.9% para el conjunto de datos de entrenamiento y del 94.5% para el set de prueba. La precisión de modelo se calculó mediante la **Ecuación 10**:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

Esta fórmula se utiliza para clasificaciones binarias como en el presente caso de estudio: yacimiento=1 o No yacimiento=0. Donde:

TP= positivos verdaderos (*True Positives*) para los casos donde el algoritmo predice un depósito y este resulta ser uno confirmado.

TN= negativos verdaderos (*True Negatives*) para los casos donde el algoritmo predice un punto de No Yacimiento correctamente.

FP= falso positivo (*False Positives*) para los casos donde el algoritmo predice un punto de No Yacimiento como Yacimiento.

FN= falso negativo (*False Negatives*) para los casos donde el algoritmo predice un punto de Yacimiento como No Yacimiento.

Esta fórmula de exactitud también puede ser expresada como:

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \quad (9)$$

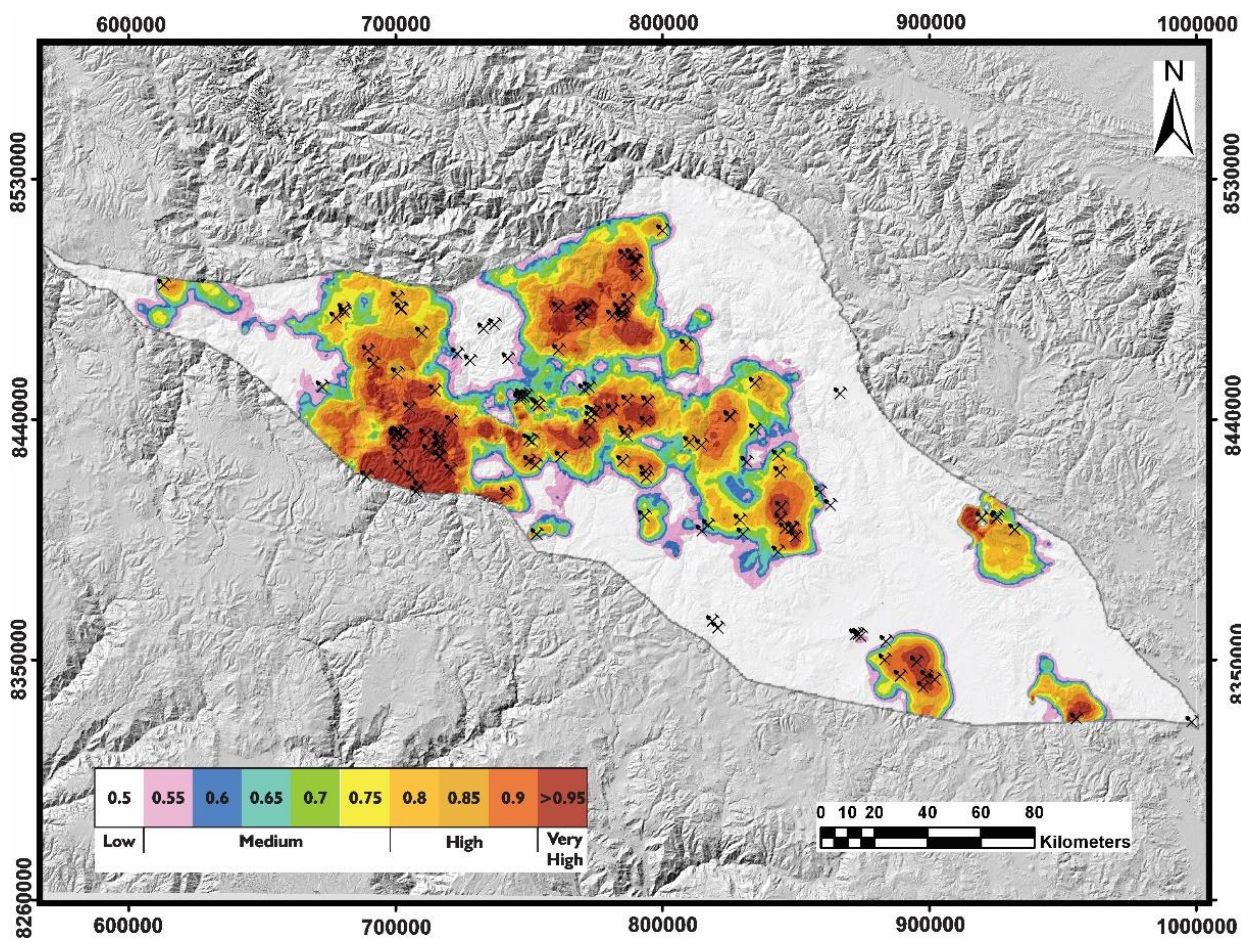
Donde: *Number of correct predictions*= número de predicciones correctas y *Total number of predictions*= número total de predicciones.

Es importante tener en cuenta que una precisión demasiado alta es perjudicial, ya que se corre el riesgo de generar un modelo con ‘overfitting’, también conocido como sobreajuste, lo que hace que el modelo sea solamente útil para los datos con los que fue entrenado, memorizando una gran cantidad de ejemplos pero perdiendo la flexibilidad necesaria para ‘aprender’ a detectar las características que le permitirán predecir exitosamente en casos que difieran del modelo original (Hawkins, 2004).

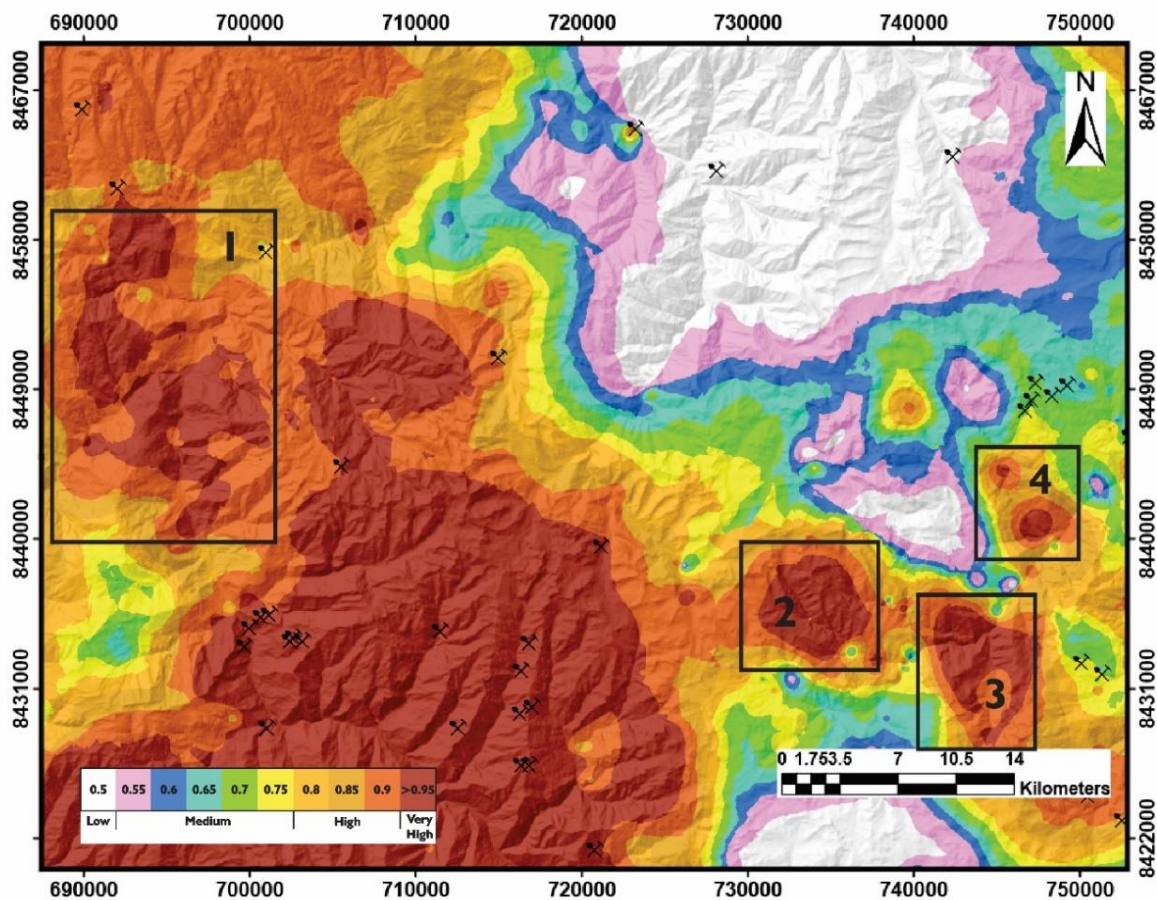
Para el entrenamiento de la red se realizaron un máximo de 1800 iteraciones, estos valores se determinan de manera empírica, ejecutando el modelo y cambiando los parámetros hasta alcanzar la convergencia, que indica el número de iteraciones a las cuales el modelo alcanza su nivel máximo de exactitud (Raschka, 2019).

El área con potencial prospectivo generado correspondió al 47.9% del área estudiada, donde se encuentran el 94% de los yacimientos confirmados. Sin embargo este potencial definido a partir de valores mayores a 0.5 en una decisión binaria (0-1), no genera un nivel de certidumbre contundente al momento de considerar o no, un lugar como prospectivo. Para esto se realizó una clasificación de probabilidad de encontrar un yacimiento (baja, media, alta y muy alta) como se observa en la **Figura 76**.

Figura 76. Mapa de predictividad de ocurrencia mineral.



La zona prospectada con potencial alto-muy alto ocupa el 18% del área de estudio y se encuentran el 83.4% de los yacimientos confirmados. A continuación se muestran las zonas de interés en las que no hay ningún yacimiento confirmado.

**Figura 77.** Acercamiento a zonas con probabilidad de ocurrencia mineral mayor al 95%.

En el sector suroeste se encontraron 4 zonas con una probabilidad mayor al 95% de ocurrencias minerales (**figura 77**). La zona 1 de aproximadamente 104 km<sup>2</sup> se encuentra circunscrita principalmente a las calizas de la Formación Arcuquina y al Plutón Linque.

Presenta alteraciones hidrotermales y de óxidos de hierro alrededor, la atraviesan múltiples pliegues y fallas en dirección suroeste que podrían funcionar como trampas físicas y canales de movilización de fluidos respectivamente, además se evidencian fuertes anomalías de Mo, Cu, Fe, Zn, As y zonas de intersección de fallas en las calizas de la Formación Arcuquina. Múltiples contactos entre intrusivos y calizas están en la zona lo que la hace realmente interesante y prospectable para yacimientos de tipo Skarn o pórfido de Cu.

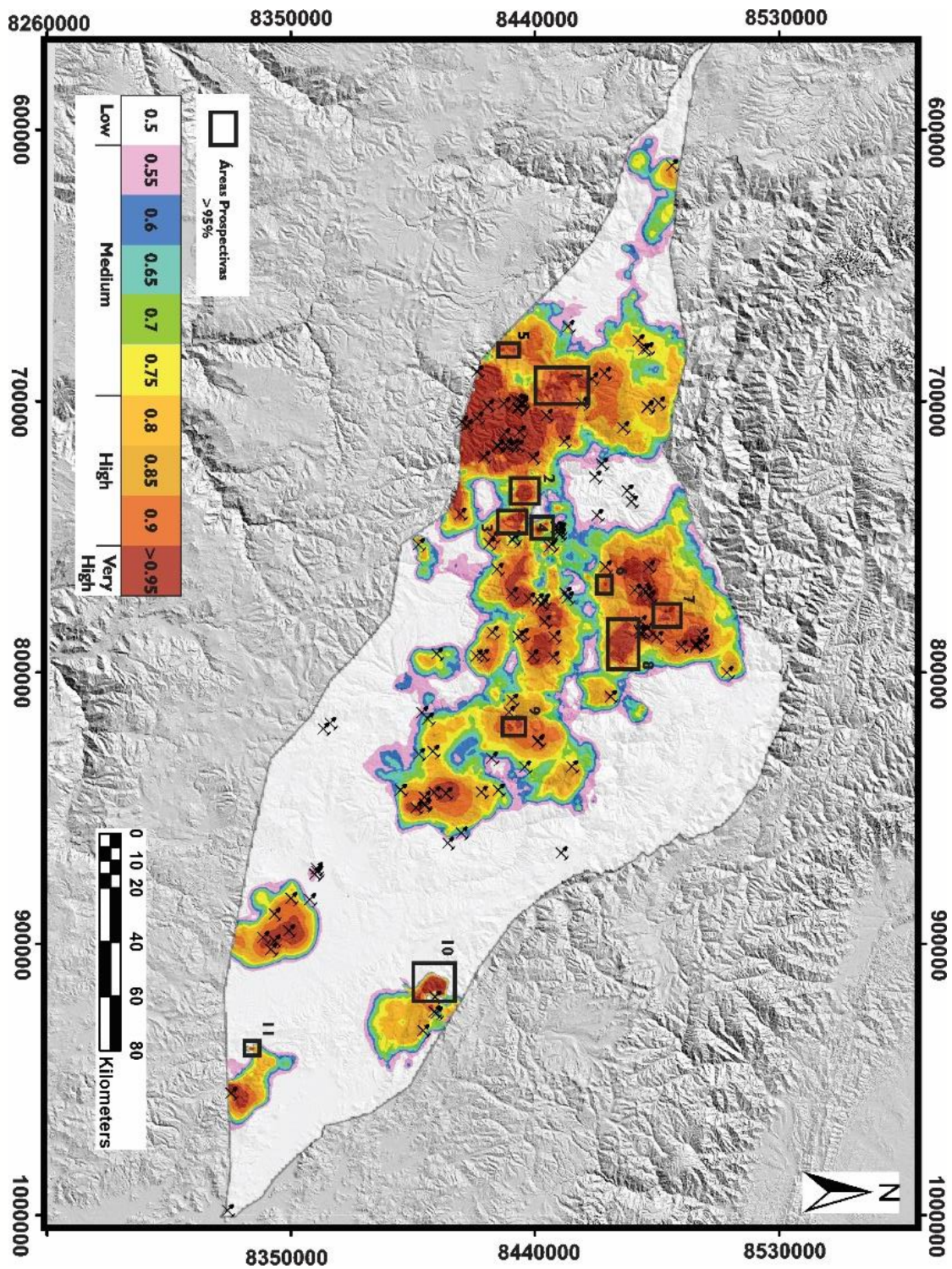
La zona 2 de aproximadamente 30 Km<sup>2</sup> encuentra en el dominio de las areniscas cuarzosas de la formaciones Cachíos, Labra y Puente. Aunque no presenta alteraciones hidrotermales o de óxidos de hierro visibles, hacia los límites se presentan contactos entre intrusivos-calizas, múltiples lineamientos y pliegues en dirección noroeste. Dentro de la zona 2 también hay una gran anomalía de Mo con concentraciones superiores a 33 ppm y anomalías de menor tamaño de Cu, Ag, Cd, Zn y Ni.

La zona 3 con un área aproximada de 28 km<sup>2</sup> se encuentra en el dominio de las areniscas de las Formación Labra, Puente y las calizas de la Formación Socosani. Presenta múltiples lineamientos y pliegues (sinclinales) en dirección SE-NW, contactos entre las granodioritas del Plutón Progreso con calizas de las formaciones Socosani. Las anomalías encontradas son de Mo, Cu, Cr, Cd, Zn, Ni y As.

La zona 4 está compuesta por 2 áreas de 3 Km<sup>2</sup> y 700 m<sup>2</sup> que se encuentran en los dominios de las areniscas cuarzosas de la Formaciones Puente y Hualhuani. Presentan fallas normales en dirección suroeste. Se encuentran anomalías con grandes extensiones de Sb, Cu, Ag y Pb.

Este proceso se realizó para todas las zonas con probabilidad superior a 95% **Figura 78** y se resumieron en la **Figura 79**

**Figura 78.** Ubicación y descripción de zonas con probabilidad mayor al 95% de ocurrencias minerales donde no hay yacimientos confirmados hasta la fecha.



**Figura 79.** Tabla resumen de las propiedades de las zonas prospectadas con probabilidad superior al 95% de encontrar ocurrencias minerales.

Zona	1	2	3	4	5	6	7	8	9	10	11
<b>Geología</b>	Calizas Fm. Arcuquina, Plutón Linque	Areniscas Fm. Cachios, Labray Puente	Areniscas Fm. Labray Puente, Calizas Fm. Sposari, Plutón Progreso	Areniscas Fm. Puente, Huahuaní	Areniscas Fm. Huahuaní, Fm. Murco, Labra, Dacta, Uti Sañayca	Areniscas Fm. Huahuaní, Fm. Murco, Labra, Dacta, Uti Sañayca	Calizas Fm. Arcuquina, Granodioritas Plutón Coyllurqui	Calizas Fm. Arcuquina, Areniscas Fm. Huahuaní, Tonalitas de Plutón Coyllurqui	Tonalitas Plutón Psuropata, Areniscas Fm. Huahuaní y Fm. Labra, Calizas Fm. Arcuquina, Tonalitas Plutón Colquemarca	Areniscas y conglomerados del grupo Mitú, areniscas Fm. Auzangate, limolitas grupo Tarma	Fijos andesíticos y areniscas del Grupo Tacaza, lignitritas del Grupo Páca
<b>Área (km²)</b>	104	30	28	3 y 0.7	2.5 y 2.2	2.5 y 2.3	1.6	105	7.10 y 5.22	57	2
<b>Fallas-Lineamientos</b>	Normales	Normales	Normales (SE-NW)	Normales (SW)	Lineamientos (SW)	Lineamientos (SW)	Fallas Normales (SW)	Fallas Normales (SN)	Fallas Normales con componente strike-slip (E-W)	Fallas Normales del sistema Urcos-Sacani-Ayaviri	Lineamientos (SE-NW)
<b>Plegues</b>	Sinclinales	Anticlinales (SW)	Sinclinales (SW)	x	Anticlinales-Sinclinales (SW)	Anticlinales-Sinclinales (SW)	Anticlinales-Sinclinales (NW)	x	Sinclinales (E-W)	x	Anticlinales (SE-NW)
<b>Cort. Intru-caliza</b>	✓	✓	✓	x	✓	✓	✓	✓	✓	x	x
<b>Alt. hidrotermal</b>	Arg. avanzada	x	x	x	Filica, Arg. Avanzada	Filica, Arg. Avanzada	x	Argilica, Calc-Silicatada.	Argilica avanzada, Sericítica	Argilica avanzada	x
<b>Alt. óxido Fe</b>	✓	x	x	x	✓	✓	x	✓	✓	✓	x
<b>Intersecc fallas</b>	✓	x	x	x	✓	x	x	✓	x	✓	x
<b>Intrusivos</b>	✓	✓	✓	✓	✓	✓	✓	✓	✓	x	x
<b>Den. lineamientos</b>	Alta	Baja	Baja	Medía	Medía	Medía	Medía	Medía-Alta	Baja	Baja	Medía
<b>Anomalías Geoquímicas</b>	Mo, Cu, Fe, Zn, As	Mo, Cu, Ag, Cd, Zn, Ni	Mo, Cu, Cr, Cd, Zn, Ni, As	Sb, Cu, Ag, Pb	Mo	Cu	Pb, Zn, Cr, Ag	Sb, Ag, Zn, As	Cu, Cr, Fe	Sb, Cu, Ag, Cr, Zn, Pb, As, Fe	Al, Sb, Cu, Ag, Zn, Pb, As

Se encontraron similitudes evidentes entre las zonas altamente prospectivas generadas a partir del modelo y las zonas con alta densidad de ocurrencias minerales. Las principales características son (1) la estrecha relación geológica de los cuerpos intrusivos como los basaltos de la Formación Anta de 29 m.a. y los pórfidos tonalíticos en la región norte, que datan de 34.2 m.a.; hacia el centro las granodioritas y monzonitas de 34.2 y 36 millones de años respectivamente, pertenecientes al Batolito Andahuaylas-Yauri finalizando con la región sureste e intrusivos de diorita también del BAY de 36.2 m.a (Perelló *et al.*, 2003; Rivera, 2009). (2) las de calizas como rocas encajantes, (3) la recurrente presencia de anomalías de los mismos grupos de elementos (Cu, Pb, Zn, Fe, As, Sb), (4) la cercanía a pliegues sinclinales y anticlinales que funcionan como trampas físicas (Maher, 2010) y (5) fallas normales en dirección SW. Características como cercanía a intersección de fallas, densidad de lineamientos, alteraciones hidrotermales y de óxidos de hierro aunque son importantes no son tan decisivas a la hora de determinar a una zona como prospectable, ya que no son común denominador en todas las áreas.

## 7. Conclusiones

El estudio demostró que los análisis de geoestadística multivariable (Correlación lineal de Pearson, dendogramas y matrices de correlación lineal) son una herramienta poderosa para encontrar asociaciones de elementos geoquímicos. En el proyecto desarrollado los elementos del Grupo 1 mostraron la asociación entre los elementos formadores de depósitos (Au, Mo, Ag, Zn, Cu) y los elementos indicadores o *pathfinders* (As, Sb, Cd, Sn, Pb). Los cálculos estadísticos establecieron satisfactoriamente los valores de fondo y umbrales que fueron de vital importancia a la hora de calcular correctamente el potencial mineral. La interpolación geoestadística (IDW-Inverse distance Weighted) permitió identificar espacialmente las anomalías y sugerir áreas con potencial para una exploración más detallada. Las asociaciones

entre elementos del grupo 1- 2 y los yacimientos ya identificados en la zona de estudio (Pórfidos-Skarn Cu-Mo,Mo-Zn) indican un origen mantélico (granito tipo I) con cierta contaminación cortical (granito tipo s).

El tratamiento de imágenes satelitales Mediante combinaciones y relaciones de bandas fueron útiles para encontrar alteraciones hidrotermales y de óxidos de hierro confirmadas con la presencia de operaciones mineras a pequeña y gran escala. Si bien son una herramienta útil, no siempre son del todo efectivas y están limitadas a las expresiones más superficiales de los cuerpos de roca alterados. Es por esto, que una integración con la geología y los datos de geoquímica de sedimentos activos se hace necesaria y es efectuada gracias al algoritmo de redes neuronales.

Las mineralizaciones se encuentran principalmente relacionadas a los intrusivos en contacto con las calizas de las formaciones Socosani y Arcurquina, estos son de composición basáltica, diorítica, granodiorítica y monzonítica y pertenecen al Batolito de Abancay, varían en un rango de edades de entre 29-63 Ma y limitan la temporalidad de las mineralizaciones respecto a intrusivos más antiguos como la granodiorita de Cochabamba (222Ma). Esto se explica por la depositación tardía de las unidades carbonatadas que imposibilitó la interacción necesaria con los cuerpos intrusivos.

Estructuralmente las zonas mineralizadas están influenciadas por fallas con tendencia SW, estas son importantes porque proveen conductos por los cuales penetran fluidos de los que precipitarán minerales de interés. Los pliegues son estructuras ideales para el entrapamiento físico de las mineralizaciones, en varios de los depósitos estudiados las zonas de interés económico se encuentran hacia los flancos y zona de charnela.

El modelo fue capaz de integrar exitosamente información proveniente de numerosas capas de datos geoquímicos, geológicos y de imágenes satelitales, y entrega un mapa que

predice correctamente la ubicación de los yacimientos y reduce efectivamente el área con potencial prospectivo esto tiene una repercusión directa en la reducción de riesgo exploratorio, por ende también en los costos. Este mapa propone 11 zonas con probabilidad muy alta que hasta ahora no han sido explotadas y que pueden ser utilizadas a nivel local para prospecciones futuras. Las zonas predichas con alto y muy alto potencial ocuparon el 16% del área de estudio y capturaron el 84% de los depósitos confirmados.

Los resultados del modelo predictivo de redes neuronales mostraron que la cercanía a intrusivos con rocas encajantes calizas y las anomalías geoquímicas principalmente de Fe, Cu, Au y Pb son las capas de mayor peso en modelo predictivo, las variables de menor interés fueron densidad de lineamientos e intersección de fallas.

### Referencias Bibliográficas

- Abedi, M., Norouzi, G., & Bahroudi, A. (2012). Computers & Geosciences Support vector machine for multi-classification of mineral prospectivity areas. *Computers and Geosciences*, 46, 272–283. <https://doi.org/10.1016/j.cageo.2011.12.014>
- Acosta, J., Rivera, R., Valencia, M., Chirif, H., Huanacuni, D., Rodríguez, I., Villarreal, E., & Paico, D. (2009). Mapa metalogenético del Perú 2009. *Ingemmet*, 2008, 1–2.
- Acosta Ale, J. G., Villarreal Jaramillo, E., Heo, C. H., Kim, Y. D., Yang, S. J., Yu, B. H., & Im, O. (2019). Estudio metalogenético de la mineralización de Cu-Au en las franjas de pórfidos epitermales de Apurímac: Sureste del Perú. Región Apurímac.
- Aiken, L., West, S., Pitts, S., Baraldi, A., & Wurpts, I. (2009). Multiple linear regression. In *Bmj* (Vol. 338). Liang, S., Li, X., & Wang, J. (Eds.). (2012). *Advanced remote sensing : Terrestrial information extraction and applications*. ProQuest Ebook Central <https://ebookcentral.proquest.com>
- Anderson, & Darling. (1930). The Annals of Mathematical Statistics. *The Annals of Mathematical Statistics*, 1(1), 1–2. <https://doi.org/10.1214/aoms/1177733256>
- Bellman, R., & Kalaba, R. E. (1965). *Dynamic programming and modern control theory* (Vol. 81). New York: Academic Press.
- Boni, M., Gluseppina Balassone, Vernon Arseneau, & Paul Schmidt. (2009). The nonsulfide zinc deposit at accha (Southern Peru): Geological and mineralogical characterization. *Economic Geology*, 104(2), 267–289. <https://doi.org/10.2113/gsecongeo.104.2.267>
- Brown, W. M., Gedeon, T. D., Groves, D. I., & Barnes, R. G. (2000). *Artificial neural networks : a new method for mineral prospectivity mapping*. 757–770.
- Carlotto-Caillaux, V. (1998). *Évolution andine et raccourcissement au niveau de Cusco* (13-

16 S), Pérou: enregistrement sédimentaire, chronologie, contrôles paléogéographiques, évolution cinématique (Doctoral dissertation, Grenoble 1).

Carlotto, Victor, Carlier, G., Jaillard, E., Sempéré, T., & Mascle, G. (1999). Sedimentary and structural evolution of the Eocene-Oligocene Capas Rojas basin: evidence for a late Eocene lithospheric delamination event in the southern Peruvian Altiplano (1999). *Isag*, 998, 1–6. <http://en.scientificcommons.org/51326450>

Carlotto, Victor, Carlier, G., Jaillard, E., Cardenas, J. O. S. É., Cerpa, L., Flores, T., & Latorre, O. (2005). Las cuencas terciarias sinorogénicas en el Altiplano y en la Cordillera Occidental. *Boletín de la Sociedad geologica del Perú*, 6(spécial), 103-à126.

Carlotto, Víctor, Quispe, J., Acosta, H., Rodríguez, R., Romero, D., Cerpa, L., Mamani, M., Díaz-Martínez, E., Navarro, P., Jaimes, F., Velarde, T., Lu, S., & Cueva, E. (2009). Dominios Geotectónicos Y Metalogénesis Del Perú Geotectonic Domains As Tool for Metalogenetic Mapping in Peru. *Bol. Soc. Geol. Perú*, 103, 1–89.

Carranza, E. J. M. (2009). Objective selection of suitable unit cell size in data-driven modeling of mineral prospectivity. *Computers and Geosciences*, 35(10), 2032–2046. <https://doi.org/10.1016/j.cageo.2009.02.008>

Ccapa Cuti, D. (2019). Procesos de Operación Ore Control en el Yacimiento Porfido-Skarn de la Mina Antapaccay Compañía Glencore-Cusco-Perú.

Cerpa, L., Carlotto, V., Perelló, J., & Cárdenas, J. (2004). *Características estructurales en la Región de Andahuaylas-Yauri (Sur de Perú) y la relación con la Cordillera de Domeyko (Norte de Chile)*.

Chuvieco, E. (1991). Fundamentos de teledetección espacial. *Estudios Geográficos*, 52(203), 371.

- Demirel, N., & Düzgün, H. S. (2011). Principles of remote sensing. In *Principles of remote sensing*. (Issue 2011, pp. 13–34).
- Devore, J. (2016). *Probabilidad y estadística para ingeniería y ciencias*. (9a. ed.) Cengage. <http://bibliotecavirtual.uis.edu.co:2168/?il=2753>
- Forina, M., Armanino, C., & Raggio, V. (2002). Clustering with dendrograms on interpretation variables. *Analytica Chimica Acta*, 454(1), 13-19.
- Friedl, M. A., & Brodley, C. E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote sensing of environment*, 61(3), 399-409.
- Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern Recognition Letters*, 27(4), 294–300. <https://doi.org/10.1016/j.patrec.2005.08.011>
- Hagen, A. (2003). Fuzzy set approach to assessing similarity of categorical maps. *International Journal of Geographical Information Science*, 17(3), 235–249. <https://doi.org/10.1080/13658810210157822>
- Hale, M. (1981). Pathfinder applications of arsenic, antimony and bismuth in geochemical exploration. In *Developments in Economic Geology* (Vol. 15, Issue C). Elsevier Scientific Publishing Company. <https://doi.org/10.1016/B978-0-444-42012-1.50024-6>
- Harris, D., Zurcher, L., Stanley, M., Marlow, J., & Pan, G. (2003). A Comparative Analysis of Favorability Mappings by Weights of Evidence , Probabilistic Neural Networks , Discriminant Analysis , and Logistic Regression. 12(4).
- Harris, D., Zurcher, L., Stanley, M., Marlow, J., & Pan, G. (2003). A Comparative Analysis of Favorability Mappings by Weights of Evidence , Probabilistic Neural Networks , Discriminant Analysis , and Logistic Regression. 12(4).

- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.
- Hawkins, D. M. (2004). The problem of overfitting. *Journal of chemical information and computer sciences*, 44(1), 1-12.
- Howarth, R. J. (1983). Statistics and Data Analysis in Geochemical Prospecting. In *Geochimica et Cosmochimica Acta* (Vol. 47, Issue 9). [https://doi.org/10.1016/0016-7037\(83\)90194-1](https://doi.org/10.1016/0016-7037(83)90194-1)
- Hulin, C. D. (1929). Structural control of ore deposition. *Economic Geology*, 24(6), 657–663. <https://doi.org/10.2113/gsecongeo.24.6.657>
- Hunt, G. R., & Ashley, R. P. (1979). Spectra of altered rocks in the visible and near infrared. *Economic Geology*, 74(7), 1613-1629.
- Ibañez, B. (2005). Sistemas de información geográfica y gestión del territorio. *Curso de Formación de GvSIG*, 120.
- INGEMMET, I. G. M. y M.-. (1995). IMGEMMET BOLETIN No. A 55: Geología del Perú. *Ingemmet*, A(55), 177. <https://hdl.handle.net/20.500.12544/176>
- Isaaks, E. H., & Srivastava, M. R. (1989). *Applied geostatistics* (No. 551.72 ISA).
- Jayalakshmi, T., & Santhakumaran, A. (2011). Statistical Normalization and Back Propagation for Classification. *International Journal of Computer Theory and Engineering*, 3(1), 89–93. <https://doi.org/10.7763/ijcte.2011.v3.288>
- Jones, B. K. (1992). Application of metal zoning to gold exploration in porphyry copper systems. *Journal of Geochemical Exploration*, 43(2), 127–155. [https://doi.org/10.1016/0375-6742\(92\)90003-Q](https://doi.org/10.1016/0375-6742(92)90003-Q)
- Kalinowski, A., & Oliver, S. (2004). *ASTER Mineral Index Processing Manual* (Issue

October).

Keim, D., Hao, M., Dayal, U., Janetzko, H., & Bak, P. (2010). *Generalized scatter plots*. 9(November 2009), 301–311. <https://doi.org/10.1057/ivs.2009.34>

Khorram, S., Koch, F. H., van der Wiele, C., & Nelson, S. (2012). Remote Sensing. In *Progress In Electromagnetics Research C*. <https://doi.org/10.2528/PIERC13030401>

Lei Ba, J., Kiros, J. R., & Hinton, G. E. (2016). Layer normalization. *ArXiv e-prints*, arXiv-1607. <http://arxiv.org/abs/1607.06450>

Liang, S., Li, X., & Wang, J. (Eds.). (2012). *Advanced remote sensing : Terrestrial information extraction and applications*. ProQuest Ebook Central. <https://ebookcentral.proquest.com>

Maher, K. C. (2010). Skarn alteration and mineralization at coroccohuayco, Tintaya District, Peru. *Economic Geology*, 105(2), 263–283. <https://doi.org/10.2113/gsecongeo.105.2.263>

Mariethoz, G., & Caers, J. (2014). *Multiple-point geostatistics: stochastic modeling with training images*. John Wiley & Sons.

Marocco, R. (1978). *Géologie des Andes péruviennes: un segment EW de la chaîne des Andes péruviennes: la déflexion d'Abancay, étude géologique de la Cordillère orientale et des hauts plateaux entre Cuzco et San Miguel, Sud du Pérou (12° 30'S à 14° 00 S)*.

Mars, J. C., & Rowan, L. C. (2006). Regional mapping of phyllic- and argillic-altered rocks in the zagros magmatic arc, Iran, using advanced spaceborne thermal emission and reflection radiometer (ASTER) data and logical operator algorithms. *Geosphere*, 2(3), 161–186. <https://doi.org/10.1130/GES00044.1>

Mingers, J. (1989). An empirical comparison of selection measures for decision-tree induction. *Machine learning*, 3(4), 319-342.

- Ninomiya, Y. (2002, March). Mapping quartz, carbonate minerals, and mafic-ultramafic rocks using remotely sensed multispectral thermal infrared ASTER data. In *Thermosense XXIV* (Vol. 4710, pp. 191-202). International Society for Optics and Photonics. <https://doi.org/10.1117/12.459566>
- Ninomiya, Y. (2004). Lithologic mapping with multispectral ASTER TIR and SWIR data. *Sensors, Systems, and Next-Generation Satellites VII*, 5234(July), 180. <https://doi.org/10.1117/12.511902>
- Ordóñez, J. L. (2012). Espectro electromagnético y espectro radioeléctrico. *Manual formativo de ACTA*, (62), 17-31.
- Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1), 217–222. <https://doi.org/10.1080/01431160412331269698>
- Palarea-Albaladejo, J., & Martín-Fernández, J. A. (2015). Treatment of Zeros, Left-Censored and Missing Values in Compositional Data Sets. *Chemometrics and Intelligent Laboratory Systems*, 143, 85–96. <https://doi.org/10.1016/j.chemolab.2015.02.019>
- Pecho Gutiérrez, V. (1981). Geología de los cuadrángulos de Chalhuanca, Antabamba y Santo Tomás. Hojas: 29-p, 29-q, y 29-r-[Boletín A 35].
- Perelló, J., Carlotto, V., Zarate, A., Ramos, P., Posso, H., Neyra, C., Caballero, A., Fuster, N., & Muhr, R. (2003). Porphyry-Style Alteration and Mineralization of the Middle Eocene to Early Oligocene Andahuaylas-Yauri Belt, Cuzco Region, Peru. *Economic Geology*, 98, 1575–1605. <https://doi.org/10.2113/gsecongeo.98.8.1575>.
- Perelló, J., Carlotto, V., Zárate, A., Ramos, P., Posso, H., Neyra, C., Caballero, A., Fuster, N., & Muhr, R. (2003). Porphyry-style alteration and mineralization of the Middle Eocene to

- Early Oligocene Andahuaylas-Yauri Belt, Cuzco Region, Peru. *Economic Geology*, 98(8), 1575–1605. <https://doi.org/10.2113/gsecongeo.98.8.1575>
- Peters, J., De Baets, B., Verhoest, N. E. C., Samson, R., Degroeve, S., Becker, P. De, & Huybrechts, W. (2007). Random forests as a tool for ecohydrological distribution modelling. *Ecological Modelling*, 207(2–4), 304–318. <https://doi.org/10.1016/j.ecolmodel.2007.05.011>
- Piringer, H., Kosara, R., & Hauser, H. (2004, July). Interactive focus+ context visualization with linked 2d/3d scatterplots. In *Proceedings. Second International Conference on Coordinated and Multiple Views in Exploratory Visualization, 2004*. (pp. 49-60). IEEE.
- Pons, X., & Solé-Sugrañes, L. (1994). A simple radiometric correction model to improve automatic mapping of vegetation from multispectral satellite data. *Remote Sensing of Environment*, 48(2), 191–204. [https://doi.org/10.1016/0034-4257\(94\)90141-4](https://doi.org/10.1016/0034-4257(94)90141-4)
- Porwal, A., Carranza, E. J. M., & Hale, M. (2003). Artificial neural networks for mineral-potential mapping: A case study from Aravalli Province, Western India. *Natural Resources Research*, 12(3), 155–171. <https://doi.org/10.1023/A:1025171803637>
- Pour, A. B., & Hashim, M. (2015). Hydrothermal alteration mapping from Landsat-8 data, Sar Cheshmeh copper mining district, south-eastern Islamic Republic of Iran. *Journal of Taibah University for Science*, 9(2), 155–166. <https://doi.org/10.1016/j.jtusci.2014.11.008>
- Raschka, S., & Mirjalili, V. (2017). Python Machine Learning: Machine Learning and Deep Learning with Python. *Scikit-Learn, and TensorFlow. Second edition ed.*
- Rencz, A. N., & Ryerson, R. A. (Eds.). (1999). *Manual of remote sensing, remote sensing for the earth sciences* (Vol. 3). John Wiley & Sons.

- Rigol-Sanchez, J. P., Chica-Olmo, M., & Abarca-Hernandez, F. (2003). Artificial neural networks as a tool for mineral potential mapping with GIS. *International Journal of Remote Sensing*, 24(5), 1151–1156. <https://doi.org/10.1080/0143116021000031791>
- Rivera Cornejo, R. (2009). Alcances metalogenéticos de los Pórfidos de Au-Cu-Mo y sus relaciones en la exploración minera (Región Cajamarca)(TT-160).
- Rivera Cornejo, R., Acosta Ale, J. G., Heo, C. H., Ryoo, C. R., Bustamante Romaní, A., Villarreal Jaramillo, E., & Santisteban Angeldonis, A. (2011). Metalogenia del Batolito Andahuaylas–Yauri y sus implicancias en la exploración de pórfidos Cu–Au.
- Rivera Cornejo, R., Bustamante Romaní, A., Acosta Ale, J. G., & Santisteban Angeldonis, A. (2010). Evaluation of ore deposits potential in the Andahuaylas–Yauri Batholith.
- Rivera Cornejo, R., & Condori Gutiérrez, N. K. (2010). El uso de la geoquímica regional para localizar provincias metalogenéticas" una visión para la exploración minera".
- Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*, 71, 804–818. <https://doi.org/10.1016/j.oregeorev.2015.01.001>
- Rodriguez-galiano, V F, Chica-olmo, M., Abarca-hernandez, F., Atkinson, P. M., & Jeganathan, C. (2012). Remote Sensing of Environment Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sensing of Environment*, 121, 93–107. <https://doi.org/10.1016/j.rse.2011.12.003>
- Rodriguez-galiano, Victor F, & Chica-rivas, M. (2012). *International Journal of Digital Earth Evaluation of different machine learning methods for land cover mapping of a Mediterranean area using multi-seasonal Landsat images and Digital Terrain Models.*

December 2012, 37–41. <https://doi.org/10.1080/17538947.2012.748848>

Rodríguez Pérez, D. Sánchez Carnero, N. y Domínguez Gómez, J. A. (2015). *Cuestiones de teledetección*. UNED - Universidad Nacional de Educación a Distancia.

<https://bibliotecavirtual.uis.edu.co:4259/es/lc/uis/titulos/48836>

Rojas, R. (1996). The Backpropagation Algorithm. In *Neural Networks* (pp. 149–182).

[https://doi.org/10.1007/978-3-642-61068-4\\_7](https://doi.org/10.1007/978-3-642-61068-4_7)

Rowan, L. C., Goetz, A. F., & Ashley, R. P. (1977). Discrimination of hydrothermally altered and unaltered rocks in visible and near infrared multispectral images. *Geophysics*, 42(3), 522-535.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536. <https://doi.org/10.1038/323533a0>

Sabins, F. F. (1999). Remote sensing for mineral exploration. *Ore geology reviews*, 14(3-4), 157-183.

Sarikaya, A., & Gleicher, M. (2017). Scatterplots: Tasks, data, and designs. *IEEE transactions on visualization and computer graphics*, 24(1), 402-412.

Sazlı, M. H. (2006). A brief review of feed-forward neural networks.

Schowengerdt, R. A. (2006). *Remote sensing: models and methods for image processing*. Elsevier.

Sholarin, E. A., & Awange, J. L. (2015). Remote sensing. *Environmental Science and Engineering (Subseries: Environmental Science)*, 9783319276496, 231–238. [https://doi.org/10.1007/978-3-319-27651-9\\_11](https://doi.org/10.1007/978-3-319-27651-9_11)

Sillitoe, R. H. (2010). Porphyry copper systems. *Economic Geology*, 105(1), 3–41.

<https://doi.org/10.2113/gsecongeo.105.1.3>

Spiegel, M. (2009). *Estadística*. (4a. ed.) McGraw-Hill Interamericana.

<http://bibliotecavirtual.uis.edu.co:2168/?il=608>

Sun, T., Chen, F., Zhong, L., Liu, W., & Wang, Y. (2019). GIS-based mineral prospectivity mapping using machine learning methods: A case study from Tongling ore district, eastern China. *Ore Geology Reviews*, 109(August 2018), 26–49.

<https://doi.org/10.1016/j.oregeorev.2019.04.003>

Triola, M. (2018). *Estadística*. (12a. ed.) Pearson Educación. Tomado de

<http://bibliotecavirtual.uis.edu.co:2168/?il=7386>

Vapnik, V. (1964). A note one class of perceptrons. *Automation and remote control*.

Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20, 273-297.

Vargas, L. E., Cuyubamba, V., Geldres, M., Aljovín, C., & Labó, R. (2017). *ATLAS GEOQUÍMICO DEL PERU*. <https://ci.nii.ac.jp/naid/40021243259/>

Volesky, J. C., Stern, R. J., & Johnson, P. R. (2003). Geological control of massive sulfide mineralization in the Neoproterozoic Wadi Bidah shear zone, southwestern Saudi Arabia, inferences from orbital remote sensing and field studies. *Precambrian Research*, 123(2–4), 235–247. [https://doi.org/10.1016/S0301-9268\(03\)00070-6](https://doi.org/10.1016/S0301-9268(03)00070-6)

Zuo, R., John, E., & Carranza, M. (2011). Computers & Geosciences Support vector machine : A tool for mapping mineral prospectivity. *Computers and Geosciences*, 37(12), 1967–

1975. <https://doi.org/10.1016/j.cageo.2010.09.014>