UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA

La Universidad Católica de Loja

ÁREA TÉCNICA

INGENIERO EN GEOLOGÍA Y MINAS

TRABAJO DE TITULACIÓN

Validación mediante Análisis Multivariable de los tipos de litologías presentes en el sector suroeste (polígono 4) de la zona Tambo-Azuay

Autor: Rojas Guano, Sergio Estuardo

Director: Guartán Medina, José Arturo

LOJA – ECUADOR

2021

2021

Aprobación del director del trabajo de titulación

Loja, 10 de mayo de 2021

Magister.

José Arturo Guartán Medina

Coordinador de Titulación de Geología

De mi consideración:

El presente trabajo de titulación: Validación mediante Análisis Multivariable de los tipos de litologías presentes en el sector suroeste (polígono 4) de la zona Tambo-Azuay, realizado por Sergio Estuardo Rojas Guano, ha sido orientado y revisado durante su ejecución, por cuanto se aprueba la presentación del mismo**.** Así mismo, doy fe que dicho Trabajo de Titulación ha sido revisado por la herramienta antiplagio institucional.

Particular que comunico para los fines pertinentes

Atentamente;

Mgtr. José Arturo Guartán Medina

C.I.: 1103034615

Declaración de autoría y cesión de derechos

"Yo, Sergio Estuardo Rojas Guano, declaro y acepto en forma expresa lo siguiente:

- Ser autor del Trabajo de Titulación denominado: Validación mediante Análisis Multivariable de los tipos de litologías presentes en el sector suroeste (polígono 4) de la zona Tambo-Azuay, específicamente de los contenidos comprendidos en: Capítulo 1. Introducción y objetivos, Capitulo 2. Marco teórico, Capitulo 3. Metodología utilizada de la investigación, Capítulo 4. Antecedentes de la zona de estudio, Capítulo 5. Desarrollo del estudio de caso, Capítulo 6. Análisis de resultados, Conclusiones y Recomendaciones, siendo José Arturo Guartán Medina, director del presente trabajo; y, en tal virtud, eximo expresamente a la Universidad Técnica Particular de Loja y a sus representantes legales de posibles reclamos o acciones judiciales o administrativas, en relación con la propiedad intelectual. Además, ratifico que las ideas, conceptos, procedimientos y resultados vertidos en el presente trabajo investigativo son de mi exclusiva responsabilidad.
- Que mi obra, producto de mis actividades académicas y de investigación, forma parte del patrimonio de la Universidad Técnica Particular de Loja, de conformidad con el artículo 20, literal j), de la Ley Orgánica de Educación Superior; y, artículo 91 del Estatuto Orgánico de la UTPL, que establece: "Forman parte del patrimonio de la Universidad la propiedad intelectual de investigaciones, trabajos científicos o técnicos y tesis de grado que se realicen a través, o con el apoyo financiero, académico o institucional (operativo) de la Universidad".
- Autorizo a la Universidad Técnica Particular de Loja para que pueda hacer uso de mi obra con fines netamente académicos, ya sea de forma impresa, digital y/o electrónica o por cualquier medio conocido o por conocerse, sirviendo el presente instrumento como la fe de mi completo consentimiento; y, para que sea ingresada al Sistema Nacional de Información de la Educación Superior del Ecuador para su difusión pública, en cumplimiento del artículo 144 de la Ley Orgánica de Educación Superior.

Firma: .....

Autor: Sergio Estuardo Rojas Guano

C.I.: 1600450413

Dedicatoria

El presente trabajo investigativo lo dedico a mis padres, por su apoyo incondicional y esfuerzo en todos estos años, gracias a ellos he logrado culminar mi carrera.

Agradecimientos

En primer lugar, agradezco a Dios por haberme dado la vida y permitirme culminar mis estudios.

Agradezco a mis padres, a mis hermanos y a mi esposa por el apoyo incondicional.

Asimismo, deseo expresar mi agradecimiento al profesor José Arturo Guartán por haberme guiado en mi tesis y por su tiempo brindado constantemente para culminar este trabajo.

Índice de Contenido

Índice de figuras

Índice de tablas

Resumen

El objetivo del presente trabajo es Validar los diferentes tipos de litologías presentes en el sector Suroeste de la zona Tambo-Azuay a través de análisis multivariable; considerando la gran importancia que tiene la caracterización litológica para las diferentes etapas de la exploración-evaluación de recursos naturales; por lo cual será de gran utilidad para las empresas dedicadas a la exploración de recursos naturales.

Se parte de una base de datos geoquímicos de 1016 puntos, con la ayuda de Excel se eliminó todos los datos duplicados y valores anómalos obteniendo una base final de 956 puntos, a los cuales se aplicó el análisis EDA, análisis multivariable, dentro de ellos se destaca el Árbol de Decisiones, con el cual se identificó que el método Chaid exhaustivo es el método que mayor eficacia presenta con una certeza del 93.41%.

Con los valores a priori y posteriori obtenidos en el análisis discriminante se elaboró un nuevo mapa geológico, existiendo una modificación de elementos químicos en los diferentes tipos litologías. Finalmente se concluye que la litología Jubones prevalece en sector Suroeste de la zona Tambo-Azuay.

Palabras claves: EDA, XLSTAT, Azuay.

Abstract

The objective of this work is to validate the different types of lithologies present in the Southwest sector of the Tambo-Azuay area through multivariate analysis; considering the great importance of lithological characterization for the different stages of exploration-evaluation of natural resources; therefore, it will be very useful for companies dedicated to the exploration of natural resources.

It starts from a geochemical database of 1016 points, with the help of Excel, all duplicate data and anomalous values were eliminated, obtaining a final base of 956 points, to which the EDA analysis, multivariate analysis, was applied. The Decision Tree stands out, with which it was identified that the exhaustive Chaid method is the method with the highest efficiency with a certainty of 93.41%. With the a priori and posteriori values obtained in the discriminant analysis, a new geological map was elaborated, with a modification of chemical elements in the different lithology types. Finally, it is concluded that the Jubones lithology prevails in the Southwest sector of the Tamb0-Azuay area.

Keywords: EDA, XLSTAT, Azuay.

Introducción

El estudio “Análisis Multivariable” busca generar una metodología estándar de discriminación de variables categóricas en función de medidas continúas utilizando técnicas de minería de datos como el análisis de componentes principales, análisis discriminante, árbol de decisiones, etc., con el fin de encontrar posibles patrones ocultos en las grandes bases de información como las que se manejan en una mina.

El desarrollo minero consta de varias etapas cada una de ellas excluyentes en el avance hacia un desarrollo exitoso; como primera fase se debe realizar la Exploración, tarea que tiene como objetivo obtener información sobre el mineral al menor gasto financiero. En nuestro estudio se aplicó el análisis multivariable para determinar qué tipo de litología predomina dentro de la zona de estudio Tambo-Azuay.

Para evaluar la gran cantidad de datos (1016) existentes en el estudio, en primera instancia se realizó el análisis de datos (EDA) y posteriormente se aplicó el software estadístico XL-STAT 2020 llegado a determinar que la litología JUBONES es la que predomina en la zona de estudio polígono 4 en el sector TAMBO AZUAY.

El presente trabajo está dividido en las siguiente secciones; la primera parte contiene una introducción, antecedentes del trabajo y los objetivos planteados para el desarrollo de la tesis; el capítulo 2 menciona los principales conceptos y definiciones a ser utilizados en el desarrollo del estudio de caso; el capítulo 3, describe la metodología utilizada para el tratamiento de los datos; el capítulo 4 se menciona las características físico geográficas de la zona de estudio y la descripción de la geología regional; el capítulo 5 se detalla los resultados del desarrollo del estudio de caso, a continuación se presenta el análisis de los resultados y finalmente tenemos las conclusiones del trabajo.

Antecedentes

La caracterización litológica en zonas de interés es importante para las diferentes etapas de la exploración-evaluación de recursos naturales, se debe mencionar que la geoquímica es como llave fundamental para la resolución de problemas petrológicos. Trabajos sobre clasificación regionalizada (Oleas, 1999) en base a muestreos geoquímicos han sido caracterizados basándose en análisis de componentes principales y análisis discriminantes. (Barbosa, P., Oliveira, T., Silva, J., 2010.)

Es importante conocer la litología del área a estudiar, puesto que el desconocimiento del tipo de roca puede traer complicaciones en cuanto a la errada asignación de algún uso antrópico que se le asigne a una zona de interés. En base a investigación se puede obtener conocimientos sobre el material parental para obtener una buena cartografía geológica que representara los conocimientos litológicos y estructuras (U.S. Geological Survey (USGS), 2006) convirtiéndose en una herramienta importante para la exploración en la búsqueda de recursos minerales.

La aplicación de métodos estadísticos multivariables revela patrones y relaciones dentro de los datos atribuidos a procesos geológicos/geoquímicos. (Grunsky E. C., 2010), describió un enfoque sistemático para evaluar los datos geoquímicos que implica el examen de los datos geoquímicos como elementos individuales y asociaciones multivariadas. Técnicas como el análisis de componentes principales, el análisis discriminante y otros procedimientos de clasificación proporcionan un marco sistemático mediante el cual se identifican los procesos geoquímicos/geológicos.

En la actualidad los métodos estadísticos se han convertido en herramientas técnicas para el proceso de datos que han sido recolectados en campo.

Para identificar las correlaciones que existe entre variables categóricas y/o concentraciones de minerales como multi-elementos se recurre a análisis multivariable como Análisis de componentes principales (PCA), que permite proyectar los datos multivariados en subespacios de pequeña dimensión (típicamente, planos), mediante el círculo de correlaciones determinar la correlación cercana o lejana entre los elementos analizados.

Para la determinación de clasificación supervisada se utilizará la relación entre las clases observadas y las clases modeladas (categóricas y cualitativas) en una “matriz de confusión”. Para esto se utilizó los métodos de análisis Discriminante, Bosques Aleatorios, Árboles de Decisión.

Este trabajo denominado Análisis Multivariable desde el punto de vista estadístico determinístico, se lo aplico a una base de datos de valores geoquímicos, de la zona del Tambo ubicada en la provincia del Azuay, en el cual se aplica herramientas de análisis multivariable y poder predecir las litologías mediante una clasificación supervisada como arboles de decisión, análisis discriminante y bosque aleatorio.

Objetivos

Objetivo general

- Validar a través de análisis multivariable la litología presente en el sector Suroeste de la zona Tambo-Azuay.

Objetivos específicos

- Realizar la validación de datos utilizando clasificación regionalizada con la aplicación de Análisis Estadístico Multivariable.
- Definir el modelo geológico predictivo como clasificación regionalizada entre la geoquímica y litología.

Capítulo uno

Marco teórico

1.1 Litología

La corteza terrestre continental está compuesta en su mayoría por 11 elementos químicos, por intermedio del análisis químico de rocas son determinados primeramente la participación de los denominados elementos mayoritarios (Poldervaart, 1955); (Ronov & Yarosheysky, 1976) que son expresados en peso de los óxidos de elementos.

Las rocas y sus diversos tipos de litologías pueden ser estudiados a partir de la mineralogía o la prospección geoquímica que analiza la composición del sedimento superficial o del suelo (*s*), la cual varía espacialmente como resultado de los cambios del clima (*cl*), intervención de organismos (*o*), la forma de relieve (*r*), el material de origen o parental (*p*) y el tiempo de formación de los suelos (*t*). Jenny (1941) sostiene que la formación de los suelos lo expresa como: $s = f(cl, o, r, p, t)$. La validación cuantitativa de la relación entre los elementos químicos y la litología parental es posible por el desarrollo de grandes bases de datos digitales, por avances de capacidades analíticas y aplicaciones de métodos de análisis estadísticos multivariables. La documentación o mapeo de suelos y rocas con una buena capacidad predictiva se beneficia en el descubrimiento de conocimientos en pedología. (McBratney, A., et al, 2003)

En base a investigación se puede obtener conocimientos sobre el material parental, litologías, estructuras que llevarán a realizar una buena cartografía geológica (U.S. Geological Survey (USGS), 2006), convirtiéndose en una herramienta importante para determinar el tipo de suelo y su uso antrópico. Tener un conjunto grande de datos puede ser un desafío para reconocer el valor y el potencial que tienen los datos para ofrecer información sobre los procesos geológicos (Grunsky E. C., 2010). Investigaciones de la geoquímica de muestras de suelo en Sumatra, Indonesia (Grunsky & Smee, 1999) y de sedimentos lacustres de la Península de Melville, Canadá, indican que la geoquímica tiene un potencial de proporcionar información útil sobre la geología de las zonas muestreadas.

Los grandes conjuntos de datos geoquímicos de elementos múltiples pueden ser interpretados con mayor efectividad cuando se aplican procedimientos multivariables de reducción de dimensión. La aplicación de métodos estadísticos a menudo revela patrones y relaciones dentro de los datos atribuidos a procesos geológicos/geoquímicos. Grunsky (2010) describe un enfoque sistemático para evaluar los valores geoquímicos que implica el examen de estos datos como elementos individuales y asociaciones multivariables. (Grunsky E & Caritat P, 2020)

1.2 Análisis exploratorio de datos (EDA).

El Análisis Exploratorio de Datos (Tukey, 1977), conocido por sus siglas en inglés EDA (Exploratory Data Analysis), es considerado como un conjunto de procedimientos que proporcionar una visión más detallada y precisa de las variables en estudio, esto se apoya en procedimientos analíticos y descriptivos de carácter gráfico o semigráfico que muestran más y mejor todas las particularidades y caracteres de las variables sacando a la luz las estructuras ocultas de los datos. (Rodríguez, 2009)

El análisis estadístico de los datos; permite detectar valores atípicos o llamados outliers (Stevens, 1984) para dirigir la prueba específica de su hipótesis; determinar si todas las variables han sido medidas en cada sitio (muestreo isotópico) o si algunas variables están sub-muestreadas (muestreo heterotópico) (U.S. Geological Survey (USGS), 2006); (Barnett, 2015); además, de analizar la malla de muestreo para determinar si es representativa de la región de estudio o, al contrario, es preferencial; elegir las variables a estudiar y, si fuese necesario, realizar un cambio de variables (por ejemplo, cuando se está en presencia de restricciones composicionales o estequiométricas). (Alperin, 2013); (Martínez R., et al, 2009)

Los estudios geológicos en los cuales se utilizan métodos estadísticos se centran en las características que cambian su estado o expresión entre los diferentes elementos de la población. Se define formalmente el término variable (Tabla 1) como aquella característica, propiedad o atributo, con respecto a la cual los elementos de una población difieren de alguna forma. (Alperin, 2013)

Tabla 1.

Síntesis y ejemplos de tipos de variables

Variables cualitativas Los individuos se diferencian por los atributos que poseen. No permiten realizar operaciones algebraicas.	**Binarias:** describe solo dos estados (si/no, contaminado/sin contaminar) **Nominales:** describen los estados del atributo (estratificación, entrecruzada, plana, cruzada, plana, cruzada, etc.). **Ordinarias:** los estados tienen orden ascendente o descendente (muy seleccionado, bien seleccionado, probablemente seleccionado, mal seleccionado).
Variables cuantitativas. Los individuos se diferencian por grado o cantidad relativa expresada en un valor numérico. Permiten realizar operaciones algebraicas.	**Discretas:** sólo pueden tomar valores enteros (recuentos: 1, 2, …, 25). **Continuas:** pueden tomar cualquier valor real dentro de un intervalo (mediciones: 1.25; 1.32; …, 5,2845).

Nota. Fuente: Marta Alperín

1.3 Análisis multivariable

El análisis multivariable es la parte de la estadística y del análisis de datos que estudia, analiza, representa e interpreta los datos que resulten de observar un número de p >1 de variables estadísticas, sobre una muestra de n individuos. Las variables observadas son homogéneas y correlacionadas, sin que alguna predomine sobre las demás. La información es de carácter multidimensional, por lo tanto, la geometría, el cálculo matricial y las distribuciones multivariables juegan un papel fundamental. (Cuadras, 2007)

1.3.1 Análisis de componentes principales

El análisis de componentes principales una técnica no supervisada que transforma el conjunto de variables originales en un conjunto más pequeño de variables, las cuales son combinaciones lineales de las primeras, que contienen la mayor parte de la variabilidad presente en el conjunto inicial. (Diaz, 2007)

El análisis por componentes principales según Díaz 2007 en su libro indica que los objetivos del PCA, entre otros, son los siguientes:

- Generar nuevas variables que expresen la información contenida en un conjunto de datos.
- Reducir la dimensión del espacio donde están inscritos los datos.
- Eliminar las variables (si es posible) que aporten poco al estudio del problema.
- Facilitar la interpretación de la información contenida en los datos.

El análisis por componentes principales tiene como propósito central la determinación de unos pocos factores (componentes principales) que retengan la mayor variabilidad contenida en los datos. Las nuevas variables poseen algunas características estadísticas "deseables", tales como independencia (bajo el supuesto de normalidad) y no correlación. (Diaz, 2007)

1.3.2 Clasificación supervisada

1.3.2.1 Análisis discriminante. Una de las técnicas del análisis multivariable que permite asignar o clasificar nuevos individuos dentro de grupos previamente reconocidos o definidos. La finalidad del Análisis Discriminante es analizar si existen diferencias significativas entre grupos de objetos respecto a un conjunto de variables medidas sobre los mismos, y facilitar procedimientos de clasificación sistemática de nuevas observaciones de origen desconocido en uno de los grupos analizados. (De la Fuente Fernandez, S., 2011)

El Análisis Discriminante se puede considerar como un análisis de regresión donde la variable dependiente es categórica y tiene como categorías la etiqueta de cada uno de los grupos, mientras que las variables independientes son continuas y determinan a qué grupos pertenecen los objetos. (De la Fuente Fernandez, S., 2011)

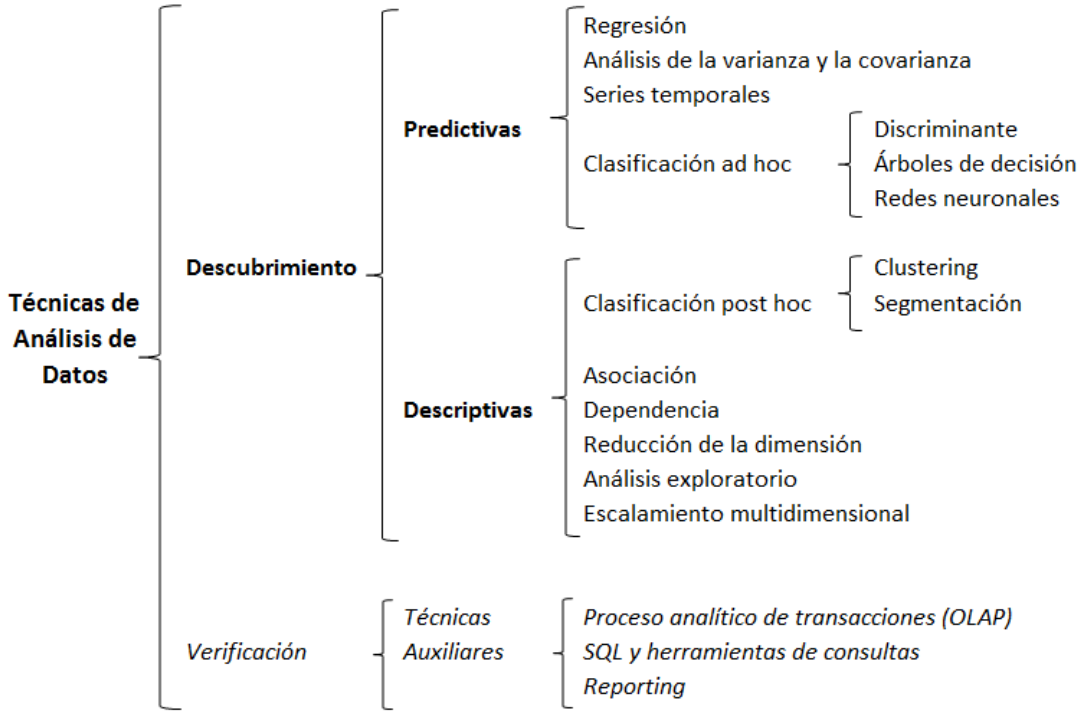
1.3.2.2 Arboles de decisión. Los árboles de decisión son una técnica estadística para la segmentación, la estratificación, la predicción, la reducción de datos y el filtrado de variables, la identificación de interacciones, la fusión de categorías y la discretización de variables continuas. (Berlanga, Rubio, & Vilà, 2013). La clasificación inicial de las técnicas de minería (Figura 1) de datos distingue entre técnicas predictivas, en las que las variables pueden clasificarse en dependientes e independientes; técnicas descriptivas, en las que

todas las variables tienen el mismo estatus y técnicas auxiliares, en las que se realiza un análisis multidimensional de datos. (Berlanga, Rubio, & Vilà, 2013)

Un árbol de decisión es una forma gráfica y analítica de representar todos los eventos (sucesos) que pueden surgir a partir de una decisión asumida en cierto momento. Nos ayudan a tomar la decisión más “acertada”, desde un punto de vista probabilístico, ante un abanico de posibles decisiones. Estos árboles permiten examinar los resultados y determinar visualmente cómo fluye el modelo. Los resultados visuales ayudan a buscar subgrupos específicos y relaciones que tal vez no encontraríamos con estadísticos más tradicionales. (Berlanga, Rubio, & Vilà, 2013)

Figura 1.

Clasificación de las técnicas de Data Mining



Fuente: Berlanga y Rubio, (2013)

1.3.2.2.1 ***Usos generales del análisis de árboles de decisión.*** Los árboles de decisión crean un modelo de clasificación basado en diagramas de flujo. Clasifican casos en grupos o pronostican valores de una variable dependiente (criterio) basada en valores de variables independientes (predictoras). (Berlanga, Rubio, & Vilà, 2013)

Las ventajas de un árbol de decisión según Berlanga et al (2013) son:

- Facilita la interpretación de la decisión adoptada.
- Facilita la comprensión del conocimiento utilizado en la toma de decisiones.
- Explica el comportamiento respecto a una determinada decisión.
- Reduce el número de variables independientes. (Berlanga, Rubio, & Vilà, 2013)

Como desventajas según Acosta (2014) tenemos:

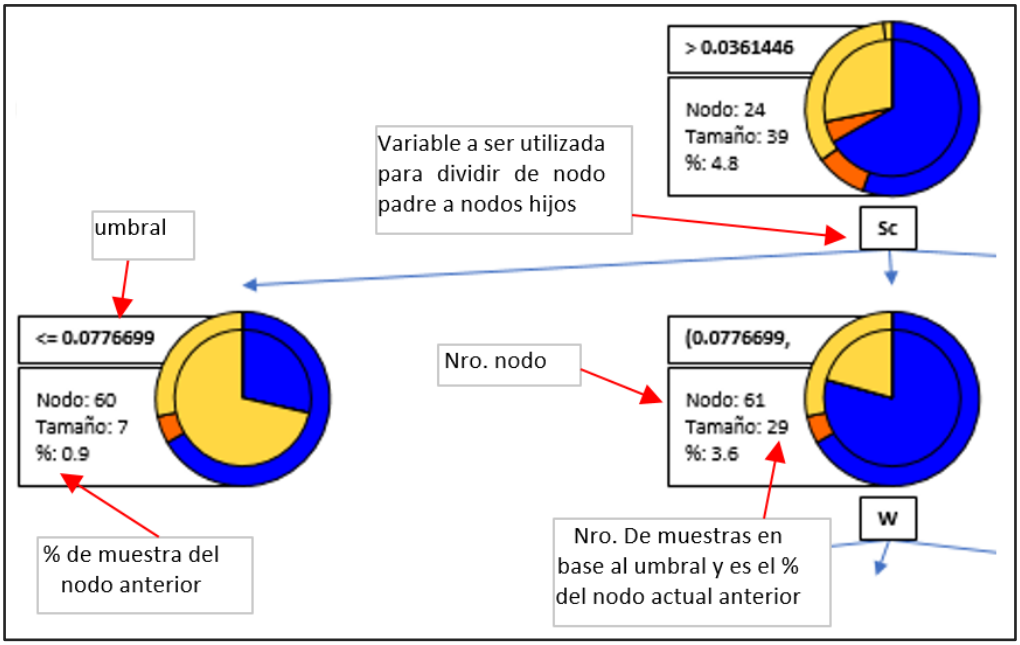
- Las reglas de asignación son bastantes sensibles a pequeñas perturbaciones en los datos.
- Existe dificultad para elegir el árbol óptimo.
- Ausencia de una función global de las variables y como consecuencia pérdida de la representación.
- Requieren un gran número de datos para asegurarse que la cantidad de las observaciones de los nodos hoja sea significativa. (Acosta, 2014)

En base a Berlanga et al, (2013) la terminología asociada a la técnica de los árboles es (Figura 2):

- **Nodo de decisión (nodo padre):** Nodo que indica que una decisión necesita tomarse en ese punto del proceso. Está representado por un cuadrado.
- **Nodo de probabilidad (nodos hijos):** Nodo que indica que en ese punto del proceso ocurre un evento aleatorio. Probabilidades de que ocurran los eventos posibles como resultado de las decisiones. Está representado por un círculo.
- **Nodo terminal:** Nodo en el que todos los casos tienen el mismo valor para la variable dependiente. Es un nodo homogéneo que no requiere ninguna división adicional, ya que es “puro”.
- **Rama:** Nos muestra los distintos caminos que se pueden emprender cuando tomamos una decisión o bien ocurre algún evento aleatorio. Resultados de las posibles interacciones entre las alternativas de decisión y los eventos. (Berlanga, Rubio, & Vilà, 2013)

Figura 2.

Partes que componen la estructura de un árbol de decisión.



1.3.2.2.2 ***Tipos de árboles de decisión.*** Diferentes algoritmos de árbol de decisión (Berlanga, Rubio, & Vilà, 2013) se conocen, los cuales se indican a continuación:

CHAID (Chi-square automatic interaction detector). Consiste en un rápido algoritmo de árbol estadístico y multidireccional que explora datos de forma rápida y eficaz, y crea segmentos y perfiles con respecto al resultado deseado. Permite la detección automática de interacciones mediante Chi-cuadrado. En cada paso, CHAID elige la variable independiente (predictora) que presenta la interacción más fuerte con la variable dependiente. Las categorías de cada predictor se funden si no son significativamente distintas respecto a la variable dependiente. (Berlanga, Rubio, & Vilà, 2013)

CHAID exhaustivo: Supone una modificación de CHAID que examina todas las divisiones posibles para cada predictor y trata todas las variables por igual, independientemente del tipo y el número de categorías.

Árboles de clasificación y regresión (CRT-Classification and regression trees): Consiste en un algoritmo de árbol binario completo que hace particiones de los datos y genera subconjuntos precisos y homogéneos. CRT divide los datos en segmentos para que sean lo más homogéneos posible respecto a la variable dependiente. (Berlanga, Rubio, & Vilà, 2013)

QUEST (Quick, unbiased, efficient, statistical tree): Consiste en un algoritmo estadístico que selecciona variables sin sesgo y crea árboles binarios precisos de forma rápida y eficaz. Con cuatro algoritmos tenemos la posibilidad de probar métodos diferentes de crecimiento de los árboles y encontrar el que mejor se adapte a nuestros datos. Es un método rápido y que evita el sesgo que presentan otros métodos al favorecer los predictores con muchas categorías. Sólo puede especificarse QUEST si la variable dependiente es nominal. (Berlanga, Rubio, & Vilà, 2013).

1.3.2.2.3 ***Validación del modelo.*** La validación permite evaluar la bondad de la estructura de árbol cuando se generaliza para una mayor población. Se puede estimar el riesgo del árbol mediante tres métodos distintos (Acosta, 2014):

- **La reestimación de toda la muestra:** El método más sencillo para calcular el riesgo es el de la reestimación, pero normalmente subestima el riesgo verdadero.
- **Validación por división muestral:** es un buen método cuando el conjunto de datos es suficientemente grande. El riesgo se calcula a partir de la muestra de comprobación.
- **Validación cruzada:** implica dividir la muestra en una serie de muestras más pequeñas y se calcula como el promedio de todos los árboles generados.

1.3.3 Bosque Aleatorio

Como desventajas de los árboles de decisión es su alta inestabilidad, las técnicas de agregación o combinación de modelos, en especial el procedimiento bagging y los modelos de Bosques Aleatorios (“Random Forests”) reducen considerablemente esa inestabilidad. Las técnicas de combinación de modelos consisten en la agregación de cierto número de modelos para obtener una clasificación o predicción a partir de los distintos clasificadores o predictores previamente generados. La construcción de cada uno de los modelos elementales puede depender de aspectos como los siguientes:

- Definición del conjunto de entrenamiento (muestras bootstrap, reponderación).
- Selección de variables.

- Elección del modelo (por ejemplo, todos árboles de decisión, o una mezcla de modelos de distinta naturaleza). (Pino, 2017)

Los modelos Random Forests se basan en las técnicas bagging, que se presentan en primer lugar. El término bagging viene de “Bootstrap aggregating”, la técnica fue propuesta por (Breiman, 1996). Un concepto fundamental en este contexto es el de muestra bootstrap. Una muestra bootstrap es una muestra aleatoria de tamaño n extraída con reemplazamiento a partir de la muestra disponible. (Pino, 2017)

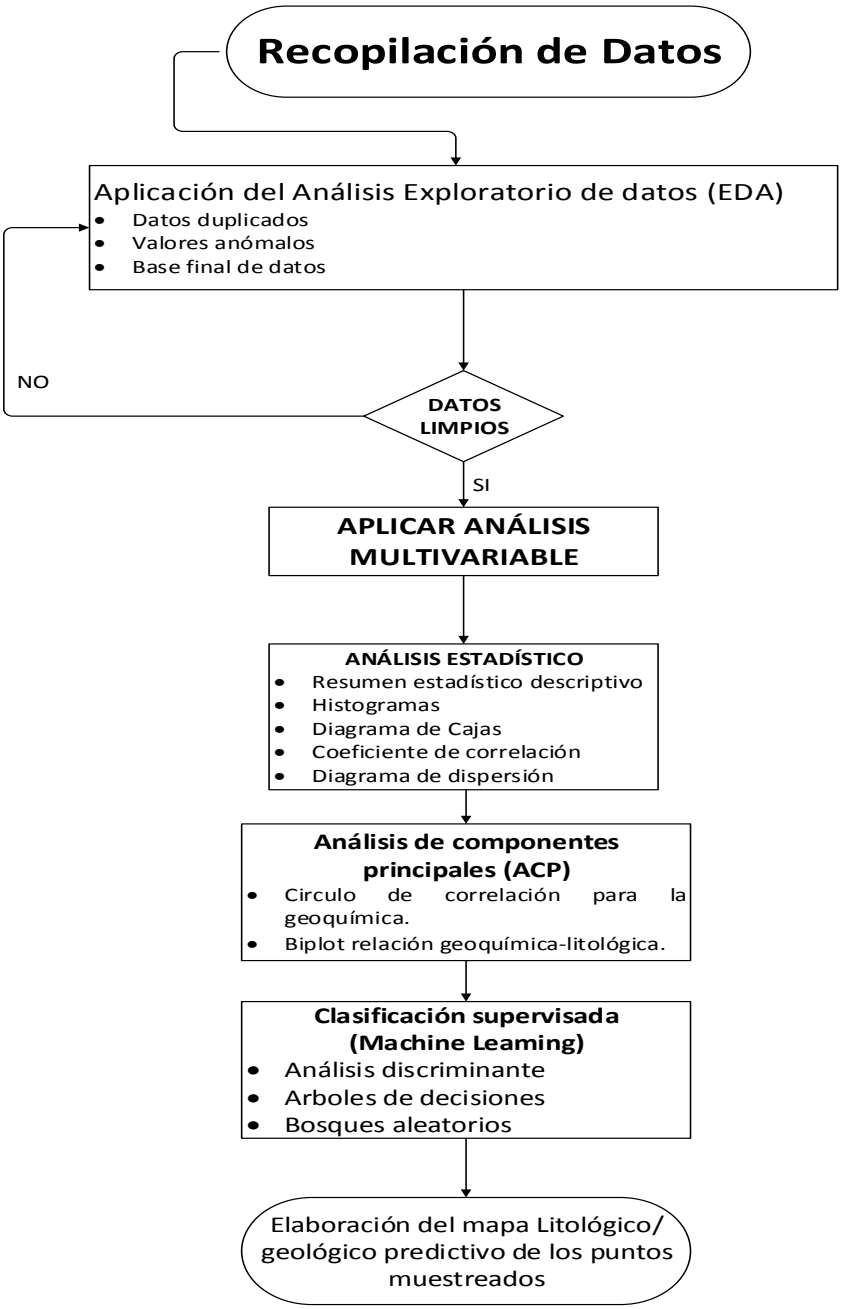
Capítulo dos

Metodología

Para conseguir los objetivos de este trabajo se planteó una metodología (Figura 3) que involucra un análisis multivariado de las variables estudiadas. Por temas de confidencialidad se omite el nombre de la Empresa que proporciono la base de datos; de igual manera el tipo de muestreo y análisis químicos de los diferentes elementos químicos.

Figura 3.

Metodología utilizada en la tesis para conseguir los objetivos planteados



Como herramientas de análisis se utilizó: programa Excel, estadístico XLSTAT, ArcGIS. Se parte de una base original de 1016 puntos que contiene valores de las concentraciones de los elementos químicos, y la georreferenciación para cada punto referenciales del sector el Tambo-Cuenca.

2.1 Análisis exploratorio de datos

A la base inicial de datos utilizando el programa Excel se realizó un Análisis Exploratorio de datos (EDA) para encontrar datos duplicados en coordenadas Y (norte) y X (este) los cuales fueron eliminados, también se buscó valores anómalos en las variables de los elementos químicos, no encontrándose este tipo de valores como, por ejemplo -99999.

Algunos datos de las variables geoquímicas tienen valores negativos (bajo límite de detección por el instrumentos o técnica usado en laboratorio) se procedieron a dividir para dos (Reiman, Filmozer, & Dutter, 2008), el resultado es reportado en forma absoluta para eliminar el signo negativo.

La base de datos final está compuesta de 956 puntos de muestreo georreferenciados, cada muestrea tiene valores de 36 elementos químicos que será la geoquímica de la zona de estudio.

2.1.1 *Determinación de las litologías*

Las litologías para cada punto de muestreo fueron obtenidas del mapa geológico de Cuenca (NV-F Cuenca Hoja 53) y del mapa geológico de Girón (NV-F Girón Hoja 54); utilizando el programa ArcGIS y de acuerdo con las coordenadas de cada punto de muestra se obtuvo la litología respectiva.

2.2 Análisis estadístico.

Los resultados de las variables cuantitativas, en este trabajo corresponde a los 36 elementos químicos, estadísticamente son presentados en una tabla resumen de estadística descriptiva en el que se detalla la media, mediana, moda, varianza, desviación estándar, valor mínimo, valor máximo, entre otros.

2.2.1 *Histogramas*

Se realizó la representación gráfica en Histogramas de algunos elementos químicos los más representativos. Con estas gráficas nos permiten visualizar el centramiento y la dispersión de datos, la distribución entorno a un rango de variación; es decir, que tan apartados están los datos con respecto a su media aritmética y como es la distribución.

2.2.2 *Diagramas de cajas*

Los diagramas de caja (Williamson, 1989), también llamados box Plots en inglés (Tukey, 1977) cual corresponde al 25 % y 75 % de los datos para el primer y tercer cuartil respectivamente; y el 50% correspondiente al segundo cuartil o llamo la mediana. Estos diagramas fueron utilizados para presentar sintéticamente los aspectos más importantes de una distribución de frecuencias: posición, dispersión, asimetría, longitud de las colas, puntos anómalos. (Alperin, 2013)

2.2.3 *Coeficiente de correlación:*

El coeficiente de correlación de Pearson fue introducido por (Galton F., 1877), permitió conocer cuantitativamente la fuerza y dirección de la relación entre los elementos químicos como variables cuantitativas. La correlación lineal es medida entre 1 y -1, siendo el 1 una correlación lineal perfecta positiva y el -1 una correlación lineal perfecta pero negativa y cero indica que tiene ausencia de correlación lineal. Estos valores ayudaran a comprender la relación entre los diferentes elementos y que pueden ser representados mediante un diagrama de dispersión. (Martínez R., et al, 2009); (Peralta, 2018)

2.2.4 *Diagramas de dispersión*

Diagrama de dispersión conocido también como scatter Plots, (Touchette, P. E., MacDonald, R. F., & Langer, S. N., 1985), permitió realizar una representación gráfica para analizar la relación que existe entre dos variables, representando como una nube de puntos. (Stevens, 1984)

2.3 Análisis multivariable

Para esta parte se realizó el análisis de los datos por medio de componentes principales; y para la clasificación supervisada mediante el análisis discriminante y árboles de decisión.

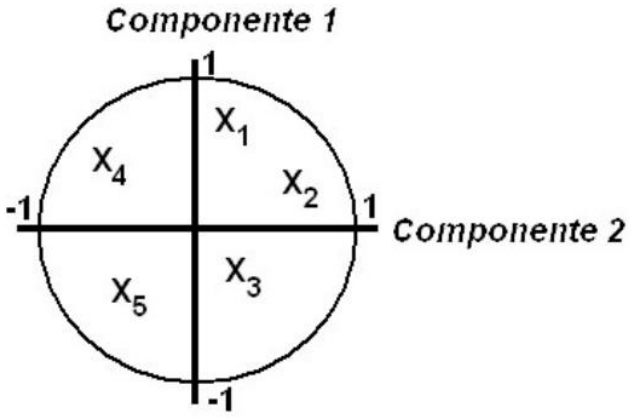
2.3.1 *Análisis de componentes principales (ACP)*

El ACP se utilizó para representar mediante un círculo de correlación (Figura 4) la mayor variabilidad de los datos, los vectores propios pueden llegar de 0 a -1 o +1, que indicara la fuerza de correlación entre las variables, por lo que todas las variables estarán contenidas dentro de un círculo de radio unidad. (De la Fuente, 2011)

La tabla de correlación de Pearson es representada mediante el círculo de correlación, para visualizar rápidamente la similitud entre todas las variables.

Figura 4.

Representación de variables en un círculo de correlación con dos componentes de mayor variabilidad



2.3.2 *Clasificación supervisada*

2.3.2.1 Análisis discriminante. El Análisis Discriminante permitirá analizar si existen diferencias significativas entre grupos de litologías con respecto a los valores de los elementos químicos, medidas sobre los mismos puntos, esto tendrá una clasificación sistemática lineal de nuevas observaciones, con la precisión en base a la matriz de confusión podremos mejorar con el uso de los árboles de decisión.

2.3.2.2 Arboles de decisión. Para determinar el mejor algoritmo de clasificación se realizó varios cálculos con cinco métodos (cart, chaid exh, quest, rdf, chaid), la precisión más alta de la clasificación es usada para el resto de los análisis. Con esto permite mediante una tabla de confusión la variable más significativa y su valor que proporciona los mejores conjuntos homogéneos de población.

Con el método con alta puntuación, se procede a utilizar en la base de datos, realizando una validación cruzada entre un subconjunto de entrenamiento y un subconjunto de prueba; el primero para estimar los parámetros del modelo y un segundo subconjunto para comprobar el comportamiento del modelo estimado.

2.4 Mapa predictivo

La validación de la litología a priori junto con los datos geoquímicos utilizado la clasificación supervisada se obtiene un mapa con las nuevas clasificaciones litológicas; estos nuevos resultados se proceden a graficar en ARGIS para obtener el mapa litológico predictivo que corroborara las litologías iniciales.

Capítulo tres

Antecedentes geográficos de la zona de estudio

3.1 Características físico-geográficas

3.1.1 *Ubicación y acceso*

El polígono de estudio se localiza en la parroquia de Chaucha, al sureste de la ciudad de Cuenca provincia del Azuay; aproximadamente a unos 60 km. El área aproximada de estudio es de 9.8km^2, enmarcado en las coordenadas (Psad56): P1. 680300-9669400; P2. 686050-9669400; P3. 686050-9667600; P4. 680300-9667600 (Figura 5). Forma parte de la zona subandina de la cordillera occidental de los Andes, (ILION,SISTEMS, 2019), para su acceso se tiene varias vías de tercer orden y como puntos de referencia se tiene la hacienda Soldados y el sector de Pimo.

3.1.2 *Clima*

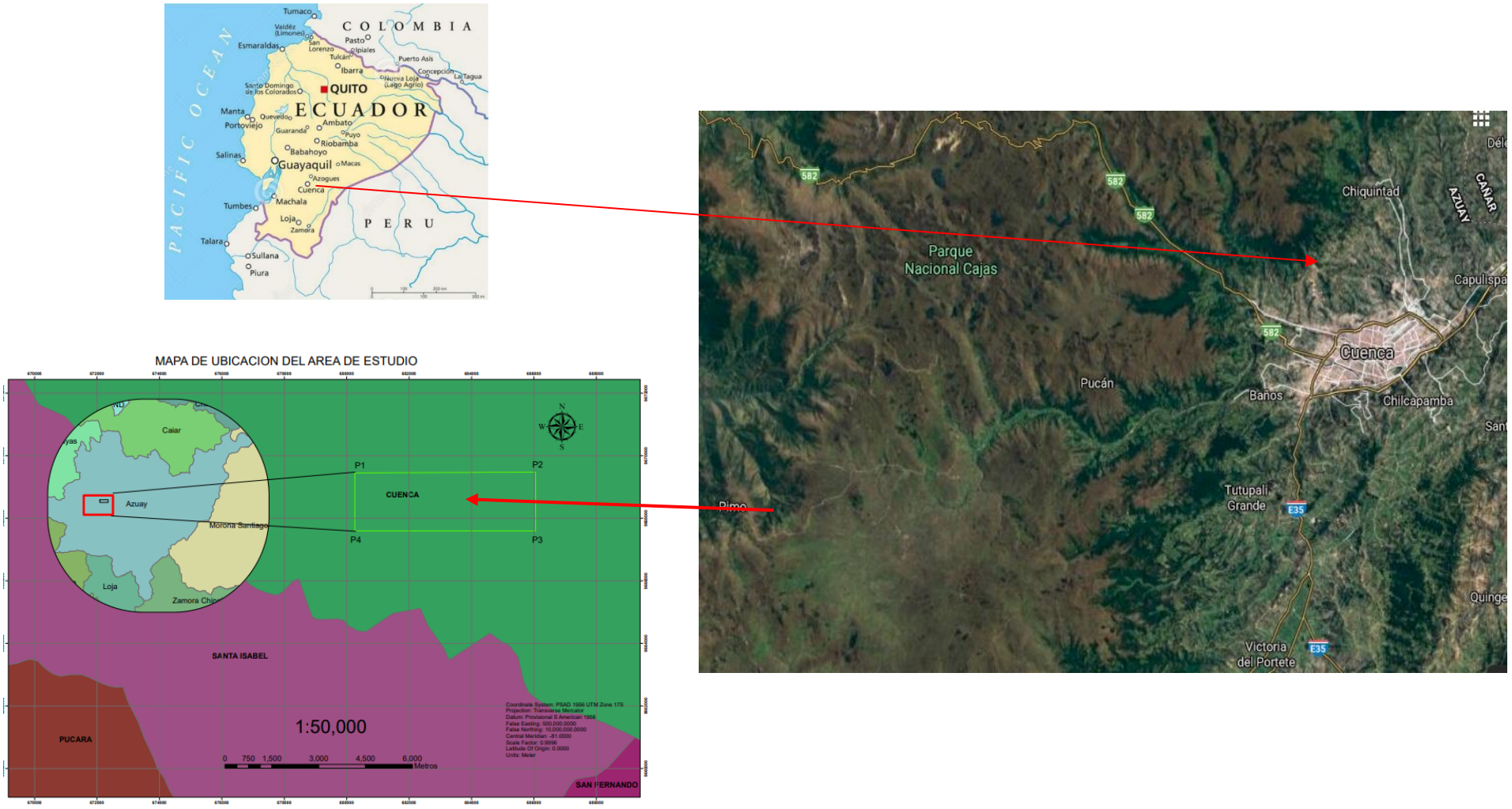
Existe frio de los páramos andinos, donde las temperaturas oscilan entre 8 °C - 10 °C; y, clima templado entre los 15 °C – 17 °C en los valles interandinos. En términos generales, el clima de toda la zona está determinado por la presencia de la cordillera de los Andes, por su variante topografía influye directamente en la temperatura y pluviosidad, resultando distintos microclimas (CLIRSEN,SIGAGRO-MAGAP, 2011).

3.1.3 *Hidrogeología*

Constituida por una serie de lagunas vertientes, drenajes menores, al nor-este de Chaucha, se encuentra las lagunas: Doublas, Napale, Estrellas Cocha, Atrancaderos, Atugpamba, Negra, Pallacocha, y las quebradas de Jerez y Canoas. Por el sur- oeste se encuentran los ríos Pita y Pingullo, localizados el primero en la comunidad de Naranjos y el segundo en la comunidad de Yubar Potrero. Estas lagunas, quebradas y ríos forman río Angas, Mientras que, en el sureste de la parroquia, en las comunidades de Pimo y Can, se origina el río Galgal (Espinoza, 2015).

Figura 5.

Ubicación geográfica del polígono de estudio en relación a la ciudad de Cuenca (Ecuador)



3.2 Geología regional de la zona de estudio

3.2.1 Descripción de las formaciones geológicas

3.2.1.1 Rocas metamórficas (M). Se extienden desde Molleturo hasta el S de Chaucha y constituyen el complejo basamento metamórfico. Comprenden metasedimentos con filitas cizalladas y esquistos grafíticos, biotíticos y moscovíticos intercalados con psamitas y conglomerados locales, se presentan como inliers controlados por fallas con dirección NE y techos pendientes en el Batolito de Chaucha. El grado metamórfico es bajo, reconociéndose en lugares granate, sillimanita y andalucita (Dunkley, P & Gaibor, A, 1997). De acuerdo con (Litherland, M; et al, 1994) constituirían el terreno metamórfico Chaucha, compuesto por sedimentos metamórficos tardíos afectados por una orogénia Triásica. (INSTITUTO GEOGRÁFICO MILITAR (IGM) DEL ECUADOR, 2002)

3.2.1.2 Formación chanlud (OScd) (Oligoceno temprano). Según (Dunkley, P & Gaibor, A, 1997) comprende lavas andesíticas subhorizontales con brechas e intercalaciones menores de sedimentos volcánicos y tobas. Las brechas son autoclasticas, piroclásticas o epiclasticas, las cuales en algunos lugares son volumétricamente más importantes que las lavas. La Formación Chanlud puede alcanzar un espesor máximo de 1000 m compuesto de flujos daciticos o riolíticos con varios diques basálticos, descansa sobre rocas plegadas de la Formación Soldados en el sur y toba riolítica de la Formación Cerro Caucay en el norte. (INSTITUTO GEOGRÁFICO MILITAR (IGM) DEL ECUADOR, 2002)

3.2.1.3 Formación plancharumi. (Ry) (Oligoceno Tardío), corresponde a una secuencia pobremente litificada de depósitos volcanoclásticos riolíticas y sedimentos fluvio lacustres que afloran al sur de Soldados. Descansa discordantemente sobre la Formación Soldados y a su vez sobreyacida con una discordancia angular por las formaciones Jubones y Quimsacocha. En el cerro Plancharumi, ocurre una secuencia estratificada de hasta 400 m de espesor, de tobas de flujo de cenizas blancas ricas en pómez, brechas riolíticas de flujos de masa, areniscas tobaceas, limolitas laminadas y tobas finas blancas. La toba superior es vidriosa, densa y extremadamente soldada con una fuerte fábrica eutaxitica (Dunkley, P & Gaibor, A, 1997). Los afloramientos al NE de Pimo presentan lavas riolíticas intensamente

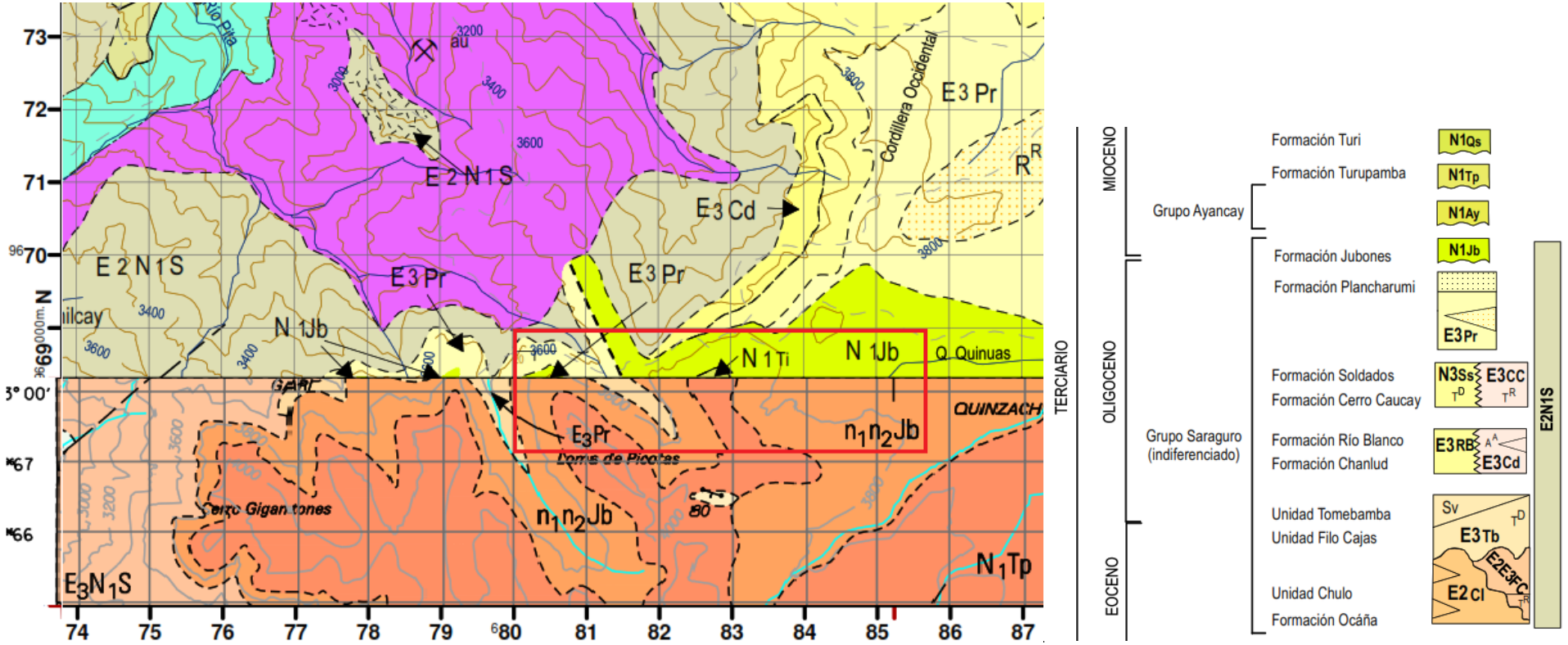
bandeadas por flujo, están intercaladas con sedimentos y tobas de ceniza fina que contienen abundante lapilli acrecionados que se interpretan como depósitos primarios de caída en aire. (INSTITUTO GEOGRÁFICO MILITAR (IGM) DEL ECUADOR, 2002)

3.2.1.4 Formación jubones (Msj) (Mioceo): (Pratt, W., et al, 1997); (Hungerbuhler D, et al, 1997). Descansa discordantemente sobre las formaciones Plancharumi y Soldados y esta sobre yacida con una fuerte discordancia por la Formación Quimsacocha. Las mejores exposiciones ocurren al S y SE de Pimo con alrededor de 200m de espesor de la toba riolítica de flujo de ceniza, fuertemente soldada, muy rica en feldespato, cuarzo y biotita. Su edad corresponda al Mioceno. (INSTITUTO GEOGRÁFICO MILITAR (IGM) DEL ECUADOR, 2002)

Figura 6.

Mapa geológico regional de la zona de estudio, cartas geológicas de Cuenca y Girón a escala 1:1000000

(INSTITUTO GEOGRÁFICO MILITAR (IGM) DEL ECUADOR, 2002)



3.2.1.5 Formación turupamba (MTu) (Miocena). Hacia el Sur la formación es extensa y comprende tobas y tobas retrabajadas ácidas con lapilli de pómez, cristales de cuarzo y fragmentos de carbón (Pratt, W., et al, 1997). Sobreyace a la Formación Turi y probablemente esta post datada por la Formación Quimsacocha. (INSTITUTO GEOGRÁFICO MILITAR (IGM) DEL ECUADOR, 2002)

3.2.1.6 Grupo Saraguro indiferenciado (E-Ms) (Eoceno Tardío). El grupo Saraguro fue redefinido por (Dunkley, P & Gaibor, A, 1997), como una secuencia de rocas volcánicas subáreas, calco-alcalinas, intermedias a ácidas, de edad Eoceno medio-tardío a Mioceno temprano. El grupo descansa inconformemente sobre o está fallado contra la Unidad Pallatanga y rocas metamórficas. Predominan composiciones andesíticas a dacíticas, pero son comunes rocas riolíticas. Estas rocas han sido termo metamorfizadas e alteradas por el batolito de Chaucha. (INSTITUTO GEOGRÁFICO MILITAR (IGM) DEL ECUADOR, 2002)

3.2.1.7 Rocas intrusivas. El Batolito de Chaucha es el de mayor importancia, consiste en granodiorita y tonalita (Misión Belga, 1986), con biotita – hornblenda. Este intruye rocas metamórficas, las Unidades Pallatanga y Yunguilla y el Grupo Saraguro. Además, hay la presencia de stocks jóvenes de cuarzo diorita o dacita porfídica las cuales intruyen la tonalita y se cree que están relacionados a la mineralización porfídica. Se conocen dos edades K/Ar de 9.77+/-0.29 Ma y 12+/-0.6 Ma (Snelling, 1970). Domos intrusivos subvolcánicos de riolita, y sills son comunes dentro del Grupo Saraguro. Sills grandes de melanodiorita y diques de andesita instruyen la Formación Chandul y dioritas de grano fino instruyen la Formación Río Blanco. (INSTITUTO GEOGRÁFICO MILITAR (IGM) DEL ECUADOR, 2002)

Capítulo cuatro

Resultados de la zona de estudio

4.1 Análisis exploratorio de datos (EDA)

Los 1016 puntos de la base de datos inicial fue sujeto a un Análisis Exploratorio de Datos, quedando un total de 956 puntos de muestras, cada punto de muestreo contiene valores de 36 elementos químicos (Tabla 2) reportados como elementos trazas en ppm con muy bajas concentraciones, típicamente < 0.1% (1000ppm) y elementos mayoritarios, reportados en porcentaje con concentraciones mayores de 1.0 %.

Para la identificación y codificación de las litologías en cada punto de muestreo se basó mediante el mapa geológico de Cuenca y Girón a escala 1:100000 (Figura 6). Las litologías para considerar en el análisis multivariable son mencionadas en la Tabla 3.

Tabla 2.

Elementos químicos en porcentaje y en ppm

Definición	Variable	Símbolo	Variable	Símbolo	Variable	Símbolo
Elementos mayores (%)	Hierro	Fe	Magnesio	Mg	Sodio	Na
	Calcio	Ca	Titanio	Ti	Potasio	K
	Fosforo	P	Aluminio	Al	Azufre	S
Elementos trazan (ppm)	Molibdeno	Mo	Uranio	U	Bario	Ba
	Cobre	Cu	Torio	Th	Boro	B
	Plomo	Pb	Estroncio	Sr	Wolframio	W
	Zinc	Zn	Cadmio	Cd	Escandio	Sc
	Plata	Ag	Antimonio	Sb	Talio	Tl
	Níquel	Ni	Bismuto	Bi	Mercurio	Hg
	Cobalto	Co	Vanadio	V	Selenio	Se
	Manganeso	Mg	Lantano	La	Teluro	Te
	Arsénico	As	Cromo	Cr	Galio	Ga

Tabla 3.

Litologías identificadas en la zona de estudio en base al mapa regional de Cuenca y Girón.

Tipo de roca	Litología	Codificación	Nro. Datos
Andesitas, lavas y brechas	Cd	2	23
Andesitas y dacitas	Sar	3	53
Riolitas	Ry	5	66
Tobas riolíticas ricas en cristales	Jub	7	776
Tobas Turupamba	Tp	8	38
		Total	***956***

4.2 Análisis estadístico de datos

El comportamiento de las variables cualitativas se verifico mediante el análisis estadístico descriptiva (Tabla 4 y 5), para conocer sus medidas de posición y dispersión, que después serán verificados en los histogramas y box Plots**.**

Tabla 4.

Resumen estadístico de los elementos mayoritarios y minoritarios reportados en porcentaje

Resumen estadístico	**Fe**	**Ca**	**P**	**Mg**	**Ti**	**Al**	**Na**	**K**	**S**
Media	2.395	0.115	0.048	0.158	0.029	2.857	0.007	0.066	0.064
Mediana	2.170	0.050	0.044	0.120	0.020	2.790	0.005	0.060	0.060
Moda	2.400	0.030	0.063	0.040	0.002	2.230	0.001	0.020	0.010
Desviación Estándar	1.569	0.166	0.032	0.139	0.028	0.894	0.006	0.052	0.045
Varianza	2.462	0.028	0.001	0.019	0.001	0.799	0.000	0.003	0.002
Curtosis	28.834	13.395	5.054	8.089	8.170	0.654	15.044	6.588	4.840
Coef. Asimetría	4.092	3.314	1.563	2.103	2.174	0.440	2.977	1.954	1.458
Rango	19.170	1.235	0.251	1.255	0.260	5.920	0.060	0.405	0.359
Mínimo	0.270	0.005	0.003	0.005	0.001	0.480	0.001	0.005	0.001
Máximo	19.440	1.240	0.254	1.260	0.261	6.400	0.061	0.410	0.360

Tabla 5.

Resumen estadístico de elementos trazas reportados en ppm.

Resumen Estadístico	**Mo**	**Cu**	**Pb**	**Zn**	**Ag**	**Ni**	**Co**	**Mn**	**As**	**U**	**Th**	**Sr**	**Cd**	**Ga**
Media	1.042	23.613	18.050	98.534	0.149	3.491	4.548	277.987	24.143	1.198	2.181	22.270	0.109	8.003
Mediana	0.660	19.340	13.310	72.400	0.095	3.000	2.600	136.500	10.200	1.000	1.400	12.950	0.080	7.800
Moda	0.470	15.910	12.540	45.400	0.066	2.400	0.800	12.000	2.100	0.600	0.200	8.500	0.040	8.200
Desviación Estándar	2.013	17.560	15.864	83.133	0.199	2.354	19.159	656.066	51.069	0.805	2.251	26.922	0.123	2.385
Varianza	4.053	308.366	251.672	6911.150	0.040	5.541	367.075	430423.240	2608.037	0.647	5.065	724.798	0.015	5.688
Curtosis	87.558	22.582	55.620	9.240	68.840	25.933	768.967	125.651	81.725	11.611	3.081	26.590	42.323	0.768
Coef. Asimetría	8.416	3.460	5.909	2.590	6.872	3.834	26.513	9.966	7.708	2.347	1.701	4.194	5.171	0.455
Rango	26.515	208.390	231.310	652.500	2.614	29.200	564.800	9996.000	680.350	8.350	14.150	292.900	1.645	15.800
Mínimo	0.005	0.140	1.250	3.800	0.009	0.300	0.100	4.000	0.050	0.050	0.050	1.700	0.005	1.400
Máximo	26.520	208.530	232.560	656.300	2.623	29.500	564.900	10000.000	680.400	8.400	14.200	294.600	1.650	17.200

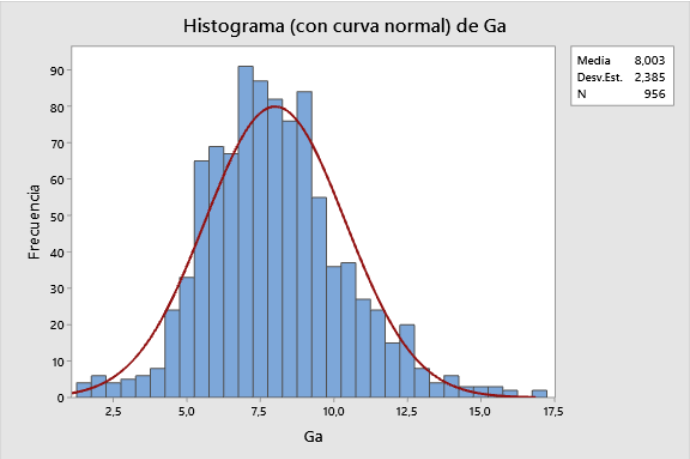
Resumen Estadístico	**Sb**	**Bi**	**V**	**La**	**Cr**	**Ba**	**B**	**W**	**Sc**	**Tl**	**Hg**	**Se**	**Te**
Media	2.621	0.933	47.128	10.496	10.776	118.163	0.766	0.063	2.485	0.258	0.261	1.211	1.014
Mediana	0.480	0.290	44.000	9.200	8.300	90.800	0.500	0.050	2.200	0.190	0.083	0.700	0.190
Moda	0.160	0.100	34.000	6.000	6.300	82.500	0.500	0.050	1.800	0.170	0.038	0.400	0.010
Desviación Estándar	5.101	1.752	20.624	6.132	8.142	96.343	0.478	0.033	1.437	0.554	0.493	1.649	2.081
Varianza	26.019	3.068	425.368	37.595	66.296	9281.894	0.228	0.001	2.064	0.307	0.243	2.718	4.330
Curtosis	15.875	45.083	2.381	13.596	10.511	15.031	18.268	35.288	5.166	220.087	23.134	25.872	49.422
Coef. Asimetría	3.525	5.396	1.080	2.805	2.853	2.754	3.526	3.091	1.789	13.476	4.255	4.032	5.262
Rango	43.820	22.660	170.000	56.500	66.600	1042.400	4.500	0.490	12.000	11.380	4.835	19.750	30.430
Mínimo	0.040	0.010	5.000	1.100	1.400	7.900	0.500	0.010	0.300	0.010	0.001	0.050	0.010
Máximo	43.860	22.670	175.000	57.600	68.000	1050.300	5.000	0.500	12.300	11.390	4.836	19.800	30.440

4.2.1 Distribución de elementos químicos

4.2.1.1 Histogramas. Se realiza la representación de frecuencia de algunos elementos químicos para conocer su distribución mediante histogramas, teniendo una distribución asimétrica positiva para la mayoría de los elementos químicos.

Figura 7.

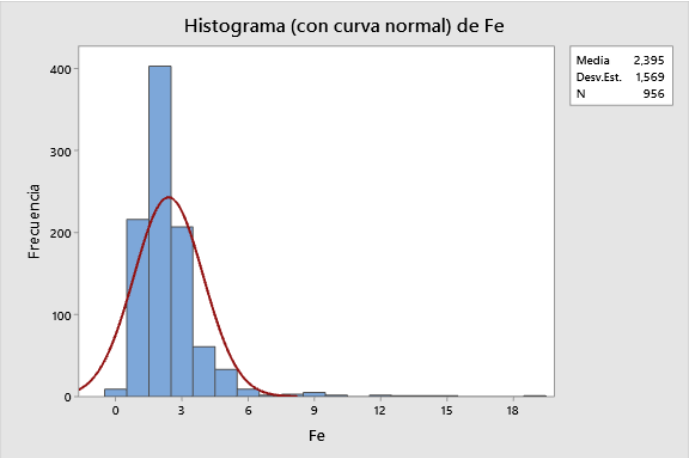
Histograma de Galio



Nota: El Galio tiene una leve distribución simetría, la media, mediana y moda están alrededor del valor de 8.00

Figura 8.

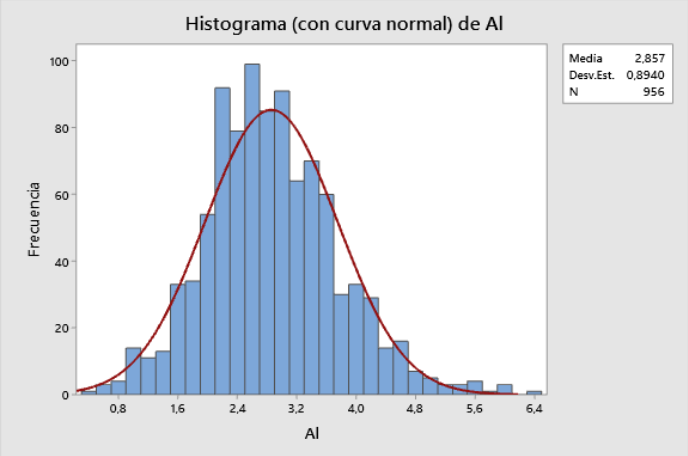
Histograma para el Hierro



Nota: Elemento con alta concentración de valores en torno a la media de 2.395, tiene una asimetría positiva, con valores extremos altos superiores a 12, lo que refleja en el valor de la desviación estándar o varianza

Figura 9.

Histograma para el aluminio

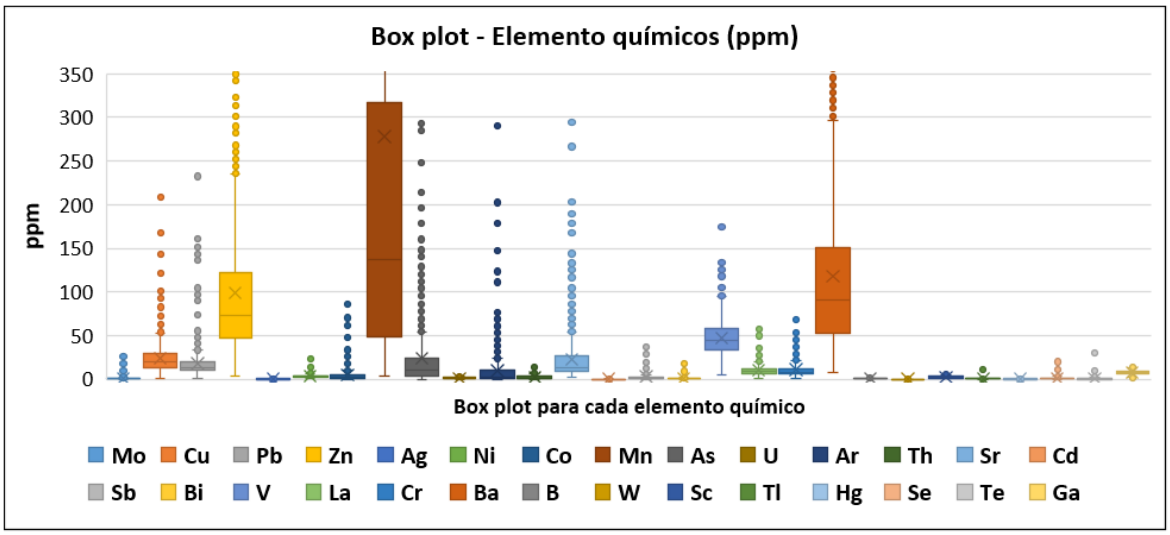


Nota: Aluminio con leve distribución asimétrica positiva, con una media de 2.857. Los valores no están muy dispersos.

4.2.1.2 Diagrama de cajas. Para comparar las distribuciones entre los elementos químicos en trazas (ppm) se utilizó el diagrama de caja que representa gráficamente la serie de datos numéricos a través de sus cuartiles. En la Figura 10 se visualiza las medidas de posición y dispersión; y valores atípicos de los elementos traza.

Figura 10.

Gráfica de cajas de elementos químicos en ppm

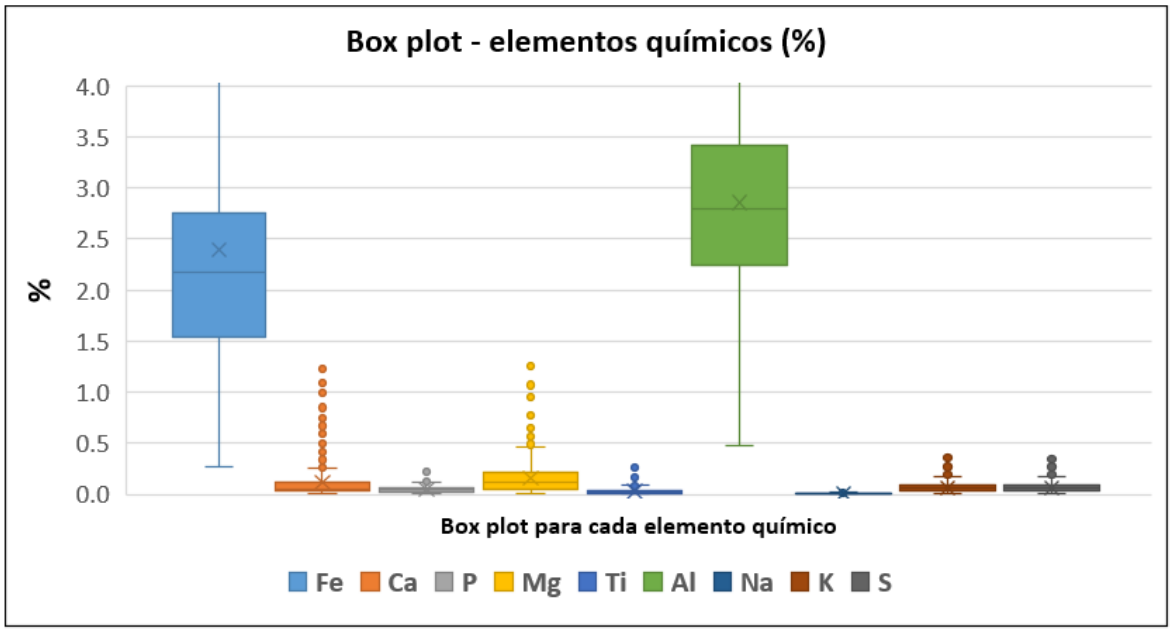


Se pueden apreciar tres elementos químicos que tienen valores altos de concentración como el zinc, manganeso y bario; los demás elementos químicos se encuentran en pequeñas proporciones. Si correlacionamos los elementos indicados podemos decir a priori que es una zona relacionada a mineralización de pórfidos.

En la Figura 11, el hierro y aluminio son los elementos con más altos contenidos de sus valores respectivos. La media de los elementos están entre el cuartil Q2 y Q3, sobre la media, lo que indica que tienen una distribución asimétrica positiva. Los elementos minoritarios (bajo el 0.5%) tienen valores atípicos u outliers.

Figura 11.

Gráfica de cajas de elementos químicos en porcentaje



4.3 Análisis de correlación

4.3.1 Correlación directa e inversamente proporcional

El coeficiente de correlación de Pearson**,** indica una correlación lineal entre 0 a +1, entre 0 a -1; en la Tabla 6 se muestra la matriz de correlación entre todas las variables en estudio.

Los elementos molibdeno (Mo), cobre (Cu) y plomo (Pb) entre ellos tienen correlación > +0.8, implicando ser elementos calcófilos con índices de busqueda de mineralizaciones en exploración de recursos.

Tabla 6.

Matriz de correlaciones (Pearson (n)):

Variables	Mo	Cu	Pb	Zn	Ag	Ni	Co	Mn	As	U	Th	Sr	Cd	Sb	Bi	V	La	Cr	Ba	B	W	Sc	Tl	Hg	Se	Te	Ga	Fe	Ca	P	Mg	Ti	Al	Na	K	S
Mo	1	0,210	0,045	-0,125	0,005	-0,009	-0,019	-0,013	0,032	-0,197	-0,184	-0,109	0,030	0,021	0,036	-0,088	-0,167	-0,019	-0,103	0,111	0,130	-0,176	0,034	-0,009	0,125	0,053	-0,096	0,053	-0,080	0,081	-0,113	-0,164	-0,195	-0,078	-0,126	0,088
Cu	0,210	1	0,265	0,085	0,310	0,165	-0,001	-0,007	0,185	-0,144	-0,352	-0,192	0,119	0,298	0,304	0,078	-0,138	0,219	-0,275	0,052	-0,008	0,094	0,068	0,280	0,528	0,320	0,216	0,104	-0,139	0,216	-0,208	-0,125	-0,006	-0,141	-0,403	0,307
Pb	0,045	0,265	1	0,160	0,323	-0,103	-0,036	-0,075	0,309	-0,033	-0,101	-0,144	-0,067	0,624	0,640	0,050	-0,103	0,091	-0,244	0,026	-0,025	-0,049	-0,023	0,526	0,324	0,631	0,133	-0,018	-0,186	-0,083	-0,246	-0,056	-0,051	-0,228	-0,245	0,056
Zn	-0,125	0,085	0,160	1	0,043	0,013	0,019	-0,035	-0,009	0,032	-0,036	-0,017	-0,048	0,214	0,260	0,161	0,005	0,139	-0,085	-0,212	-0,236	0,051	-0,025	0,237	0,030	0,171	0,213	-0,072	-0,059	-0,107	-0,148	0,078	0,069	-0,083	-0,166	0,054
Ag	0,005	0,310	0,323	0,043	1	-0,032	-0,034	-0,030	0,221	-0,051	-0,272	-0,117	0,143	0,240	0,216	-0,158	0,082	0,062	-0,203	0,018	0,109	-0,130	0,012	0,251	0,104	0,173	-0,016	-0,103	0,000	0,199	-0,147	-0,158	-0,105	-0,126	-0,146	0,180
Ni	-0,009	0,165	-0,103	0,013	-0,032	1	0,139	0,357	-0,053	0,038	-0,117	0,094	0,410	-0,132	-0,099	0,284	0,044	0,480	0,278	0,068	0,087	0,270	0,249	-0,157	-0,055	-0,114	0,256	0,109	0,190	0,263	0,409	0,309	0,368	0,094	-0,028	0,165
Co	-0,019	-0,001	-0,036	0,019	-0,034	0,139	1	0,694	-0,015	0,033	0,038	0,037	0,087	-0,052	-0,051	0,050	0,026	0,028	0,219	-0,023	-0,002	0,094	0,549	-0,044	-0,017	-0,041	0,064	0,280	0,039	0,035	0,085	0,045	0,053	0,011	0,031	-0,038
Mn	-0,013	-0,007	-0,075	-0,035	-0,030	0,357	0,694	1	-0,030	0,096	0,025	0,087	0,332	-0,120	-0,106	0,021	0,152	-0,002	0,395	0,005	0,018	0,110	0,725	-0,110	-0,031	-0,094	-0,004	0,257	0,138	0,179	0,169	0,026	0,073	0,067	0,087	-0,017
As	0,032	0,185	0,309	-0,009	0,221	-0,053	-0,015	-0,030	1	-0,136	-0,135	-0,091	0,002	0,444	0,327	0,001	-0,060	0,095	-0,158	0,056	-0,019	0,001	0,079	0,284	0,272	0,450	0,002	0,286	-0,007	0,090	-0,164	-0,128	-0,120	-0,068	-0,152	0,037
U	-0,197	-0,144	-0,033	0,032	-0,051	0,038	0,033	0,096	-0,136	1	0,486	0,278	-0,009	-0,150	-0,127	0,166	0,484	-0,051	0,335	-0,103	-0,044	0,305	0,114	-0,138	-0,170	-0,159	0,088	-0,077	0,222	-0,136	0,125	0,330	0,130	0,291	0,293	-0,131
Th	-0,184	-0,352	-0,101	-0,036	-0,272	-0,117	0,038	0,025	-0,135	0,486	1	0,209	-0,282	-0,130	-0,170	0,264	0,224	-0,209	0,362	-0,151	-0,128	0,427	0,020	-0,149	-0,227	-0,133	0,003	0,069	0,049	-0,515	0,207	0,429	0,151	0,123	0,495	-0,502
Sr	-0,109	-0,192	-0,144	-0,017	-0,117	0,094	0,037	0,087	-0,091	0,278	0,209	1	0,103	-0,147	-0,136	0,001	0,210	0,017	0,591	0,008	-0,084	0,222	-0,012	-0,125	-0,168	-0,145	-0,217	-0,079	0,671	-0,051	0,263	0,185	-0,084	0,611	0,336	-0,156
Cd	0,030	0,119	-0,067	-0,048	0,143	0,410	0,087	0,332	0,002	-0,009	-0,282	0,103	1	-0,160	-0,103	-0,189	0,232	0,057	0,170	0,278	0,190	-0,190	0,205	-0,144	-0,007	-0,168	-0,184	-0,049	0,380	0,583	0,090	-0,094	-0,103	0,144	0,055	0,445
Sb	0,021	0,298	0,624	0,214	0,240	-0,132	-0,052	-0,120	0,444	-0,150	-0,130	-0,147	-0,160	1	0,777	0,179	-0,174	0,092	-0,307	0,049	-0,087	0,033	-0,022	0,732	0,397	0,854	0,217	0,096	-0,190	-0,136	-0,310	-0,039	-0,036	-0,236	-0,340	0,014
Bi	0,036	0,304	0,640	0,260	0,216	-0,099	-0,051	-0,106	0,327	-0,127	-0,170	-0,136	-0,103	0,777	1	0,120	-0,146	0,107	-0,273	-0,002	-0,107	-0,039	-0,051	0,628	0,332	0,756	0,219	0,024	-0,178	-0,056	-0,312	-0,050	-0,038	-0,207	-0,344	0,071
V	-0,088	0,078	0,050	0,161	-0,158	0,284	0,050	0,021	0,001	0,166	0,264	0,001	-0,189	0,179	0,120	1	-0,097	0,317	0,052	-0,029	-0,065	0,518	-0,031	0,057	0,012	0,176	0,629	0,353	-0,113	-0,246	0,188	0,564	0,471	-0,124	-0,078	-0,165
La	-0,167	-0,138	-0,103	0,005	0,082	0,044	0,026	0,152	-0,060	0,484	0,224	0,210	0,232	-0,174	-0,146	-0,097	1	-0,097	0,317	-0,061	0,065	0,123	0,188	-0,124	-0,256	-0,235	-0,211	-0,121	0,311	0,180	0,045	-0,046	-0,041	0,205	0,295	-0,022
Cr	-0,019	0,219	0,091	0,139	0,062	0,480	0,028	-0,002	0,095	-0,051	-0,209	0,017	0,057	0,092	0,107	0,317	-0,097	1	-0,066	-0,091	-0,024	0,208	-0,080	0,067	-0,060	0,047	0,284	0,066	0,038	0,079	0,182	0,203	0,182	0,016	-0,208	0,109
Ba	-0,103	-0,275	-0,244	-0,085	-0,203	0,278	0,219	0,395	-0,158	0,335	0,362	0,591	0,170	-0,307	-0,273	0,052	0,317	-0,066	1	-0,030	-0,049	0,188	0,340	-0,282	-0,224	-0,284	-0,159	0,022	0,446	-0,020	0,336	0,301	0,068	0,437	0,572	-0,199
B	0,111	0,052	0,026	-0,212	0,018	0,068	-0,023	0,005	0,056	-0,103	-0,151	0,008	0,278	0,049	-0,002	-0,029	-0,061	-0,091	-0,030	1	0,280	-0,143	0,006	-0,033	0,086	0,028	-0,052	-0,028	0,157	0,281	-0,026	0,023	-0,084	0,016	0,010	0,327
W	0,130	-0,008	-0,025	-0,236	0,109	0,087	-0,002	0,018	-0,019	-0,044	-0,128	-0,084	0,190	-0,087	-0,107	-0,065	0,065	-0,024	-0,049	0,280	1	-0,126	0,003	-0,089	-0,101	-0,164	-0,100	-0,059	0,036	0,202	0,097	-0,022	-0,064	0,010	0,056	0,099
Sc	-0,176	0,094	-0,049	0,051	-0,130	0,270	0,094	0,110	0,001	0,305	0,427	0,222	-0,190	0,033	-0,039	0,518	0,123	0,208	0,188	-0,143	-0,126	1	0,062	0,003	0,024	0,063	0,364	0,310	0,112	-0,318	0,254	0,328	0,313	0,095	-0,003	-0,338
Tl	0,034	0,068	-0,023	-0,025	0,012	0,249	0,549	0,725	0,079	0,114	0,020	-0,012	0,205	-0,022	-0,051	-0,031	0,188	-0,080	0,340	0,006	0,003	0,062	1	-0,033	0,080	-0,017	0,021	0,181	0,032	0,153	-0,027	-0,046	0,035	-0,002	0,057	-0,003
Hg	-0,009	0,280	0,526	0,237	0,251	-0,157	-0,044	-0,110	0,284	-0,138	-0,149	-0,125	-0,144	0,732	0,628	0,057	-0,124	0,067	-0,282	-0,033	-0,089	0,003	-0,033	1	0,377	0,708	0,084	0,027	-0,192	-0,127	-0,328	-0,121	-0,124	-0,180	-0,344	0,016
Se	0,125	0,528	0,324	0,030	0,104	-0,055	-0,017	-0,031	0,272	-0,170	-0,227	-0,168	-0,007	0,397	0,332	0,012	-0,256	-0,060	-0,224	0,086	-0,101	0,024	0,080	0,377	1	0,513	0,116	0,207	-0,142	0,036	-0,256	-0,215	-0,099	-0,161	-0,333	0,186
Te	0,053	0,320	0,631	0,171	0,173	-0,114	-0,041	-0,094	0,450	-0,159	-0,133	-0,145	-0,168	0,854	0,756	0,176	-0,235	0,047	-0,284	0,028	-0,164	0,063	-0,017	0,708	0,513	1	0,233	0,173	-0,201	-0,143	-0,284	-0,064	-0,023	-0,237	-0,354	0,001
Ga	-0,096	0,216	0,133	0,213	-0,016	0,256	0,064	-0,004	0,002	0,088	0,003	-0,217	-0,184	0,217	0,219	0,629	-0,211	0,284	-0,159	-0,052	-0,100	0,364	0,021	0,084	0,116	0,233	1	0,123	-0,323	-0,100	0,005	0,397	0,704	-0,355	-0,277	-0,022
Fe	0,053	0,104	-0,018	-0,072	-0,103	0,109	0,280	0,257	0,286	-0,077	0,069	-0,079	-0,049	0,096	0,024	0,353	-0,121	0,066	0,022	-0,028	-0,059	0,310	0,181	0,027	0,207	0,173	0,123	1	-0,061	0,067	0,045	0,060	0,107	-0,040	-0,106	-0,093
Ca	-0,080	-0,139	-0,186	-0,059	0,000	0,190	0,039	0,138	-0,007	0,222	0,049	0,671	0,380	-0,190	-0,178	-0,113	0,311	0,038	0,446	0,157	0,036	0,112	0,032	-0,192	-0,142	-0,201	-0,323	-0,061	1	0,243	0,281	0,033	-0,207	0,591	0,311	0,024
P	0,081	0,216	-0,083	-0,107	0,199	0,263	0,035	0,179	0,090	-0,136	-0,515	-0,051	0,583	-0,136	-0,056	-0,246	0,180	0,079	-0,020	0,281	0,202	-0,318	0,153	-0,127	0,036	-0,143	-0,100	0,067	0,243	1	-0,004	-0,214	0,005	0,107	-0,047	0,623
Mg	-0,113	-0,208	-0,246	-0,148	-0,147	0,409	0,085	0,169	-0,164	0,125	0,207	0,263	0,090	-0,310	-0,312	0,188	0,045	0,182	0,336	-0,026	0,097	0,254	-0,027	-0,328	-0,256	-0,284	0,005	0,045	0,281	-0,004	1	0,393	0,333	0,222	0,481	-0,209
Ti	-0,164	-0,125	-0,056	0,078	-0,158	0,309	0,045	0,026	-0,128	0,330	0,429	0,185	-0,094	-0,039	-0,050	0,564	-0,046	0,203	0,301	0,023	-0,022	0,328	-0,046	-0,121	-0,215	-0,064	0,397	0,060	0,033	-0,214	0,393	1	0,465	0,170	0,361	-0,083
Al	-0,195	-0,006	-0,051	0,069	-0,105	0,368	0,053	0,073	-0,120	0,130	0,151	-0,084	-0,103	-0,036	-0,038	0,471	-0,041	0,182	0,068	-0,084	-0,064	0,313	0,035	-0,124	-0,099	-0,023	0,704	0,107	-0,207	0,005	0,333	0,465	1	-0,246	0,048	-0,074
Na	-0,078	-0,141	-0,228	-0,083	-0,126	0,094	0,011	0,067	-0,068	0,291	0,123	0,611	0,144	-0,236	-0,207	-0,124	0,205	0,016	0,437	0,016	0,010	0,095	-0,002	-0,180	-0,161	-0,237	-0,355	-0,040	0,591	0,107	0,222	0,170	-0,246	1	0,352	0,102
K	-0,126	-0,403	-0,245	-0,166	-0,146	-0,028	0,031	0,087	-0,152	0,293	0,495	0,336	0,055	-0,340	-0,344	-0,078	0,295	-0,208	0,572	0,010	0,056	-0,003	0,057	-0,344	-0,333	-0,354	-0,277	-0,106	0,311	-0,047	0,481	0,361	0,048	0,352	1	-0,233
S	0,088	0,307	0,056	0,054	0,180	0,165	-0,038	-0,017	0,037	-0,131	-0,502	-0,156	0,445	0,014	0,071	-0,165	-0,022	0,109	-0,199	0,327	0,099	-0,338	-0,003	0,016	0,186	0,001	-0,022	-0,093	0,024	0,623	-0,209	-0,083	-0,074	0,102	-0,233	1

Los tres elementos antes mencinados forman óxidos simples, el molibdemo y plomo son siderófilos con afinidad por el hierro, se encuentran en el núcleo de la tierra; mientras que el cobre tiene tendencia a combinarse con azufre (calcófilos). El lantanio (La) con el cobre (Cu) tiene una correlación de -0.9, el La es un elemento de las tierras raras que por lo general no tiene correlación con sulfuros de cobre.

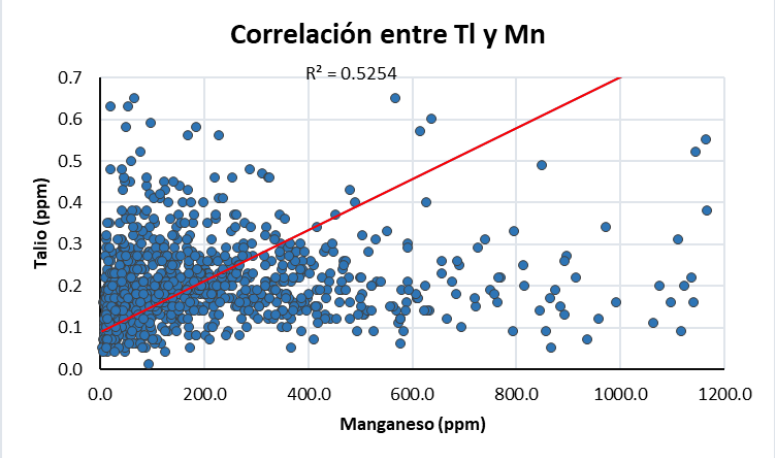
Entre los elementos litofilos torio (Th) y uranio (U) existe correlación positiva, pero estos con los siderofilos molibdeno (Mo), plata (Ag) y azufre (S) tienen correlación negativa.

4.3.1.1 Diagrama de dispersión. Los diferentes rangos de valores de correlación puede ser representado mediante diagramas tipo scatter plots (nubes de correlación, diagramas de dispersión). A continuación mostramos representación de algunas bivariables con su respectivo análisis.

La correlación entre Talio (Tl) y Manganeso (Mn) (Figura 12) es alta de 0.725. Son elementos trazas que están presentes en las rocas tipo volcánicas félsicas. En base a Nordberg (s.f) el talio está ampliamente distribuido en la corteza terrestre, aunque en concentraciones muy bajas; y, también se encuentra asociado con otros metales pesados en piritas y blendas y en los nódulos de manganeso en el lecho de los océanos. (Nordberg, s.f.)

Figura 12.

Correlación positiva entre Talio y Manganeso

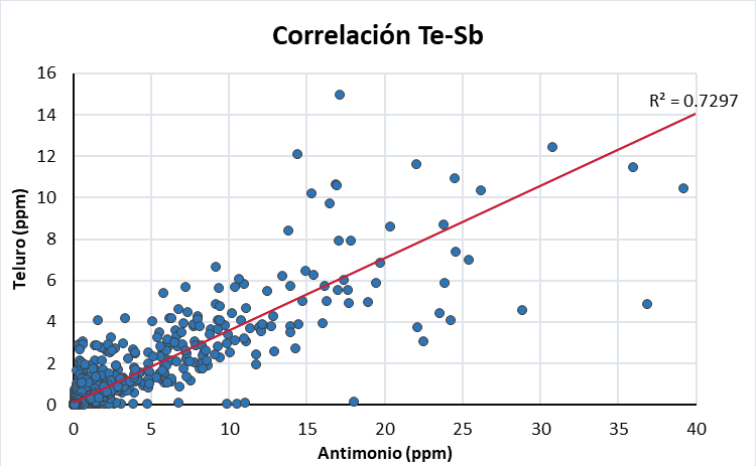


El Teluro (Te) y Antimonio (Sb) (Figure 13) tienen una fuerte correlación positiva de 0.854, estos son considerados metaloides. Además, en la tabla periódica los 2 elementos se

encuentran juntos, presentan características similares son calcófilos (con afinidad por el azufre, concentrados en sulfuros), semiconductores y presentan un estado de oxidación opuesto entre sí.

Figura 13.

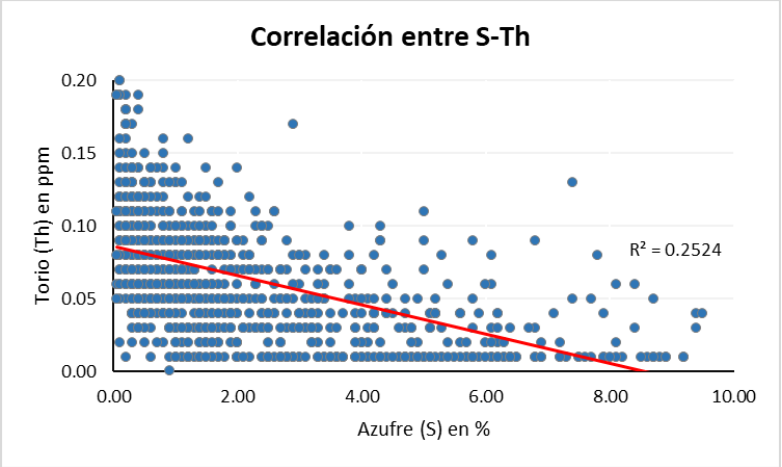
Correlación entre Selenio y Antimonio



Entre elementos de tipo calcófilos y litófilos por lo general no existe correlación, en la Figura 14 se muestra una correlación negativa entre el Azufre (calcófilo) y el Torio (litófilo), su correlación es negativa de -0.502 debido a que no son compatibles por la afinidad geoquímica ya que los litófilos se encuentran en la fase de silicatos líquidos y los calcófilos se encuentran en la fase de sulfuros líquidos dentro del manto de la tierra.

Figura 14.

Correlación negativa entre Azufre-Torio

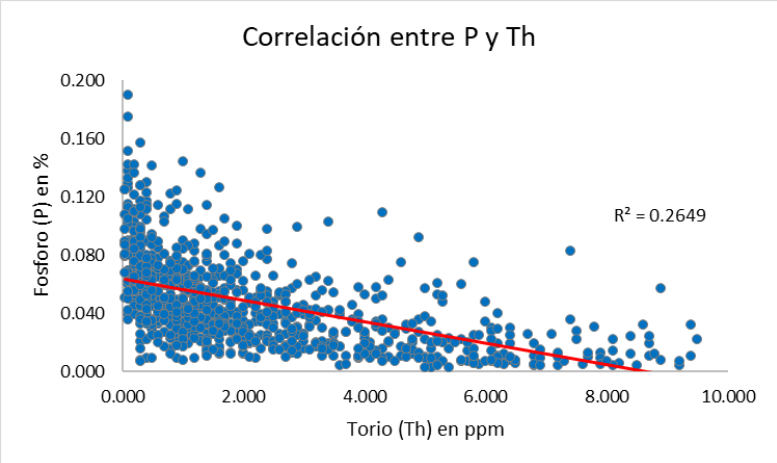


Además, los elementos calcófilos se encuentran formando compuestos estables con el azufre (sulfuros), tienen electronegatividades más altas que las de los elementos litófilos y presentan enlaces covalentes (Sarango, 2013), mientras que los litófilos tienen afinidad al oxigeno (O) y el silicio (Si), se encuentran formando silicatos, óxidos, entre otros, presentan enlaces iónicos. (Vásquez, 2017)

Con el mismo antecedente descrito anteriormente en la Figura 15, se muestra una correlación negativa entre fósforo (biófilo) y el Torio (litófilo), su correlación es -0.515. Son elementos altamente reactivos que al encontrase en la corteza terrestre reaccionan con el oxígeno y dentro de la tabla periódica se encuentran opuestos entre sí. Durante la cristalización magmática estos elementos químicos se asocian en la fase pegmatítica. Los minerales que se forman son silicatos ricos en sílice (cuarzo, ortosa, albita), en grupos hidroxilo (micas) y en elementos como el boro (turmalina), el fósforo (apatito), el flúor (fluorita), etc. (Servicio geológico Mexicano, 2017)

Figura 15.

Correlación entre Azufre y Torio



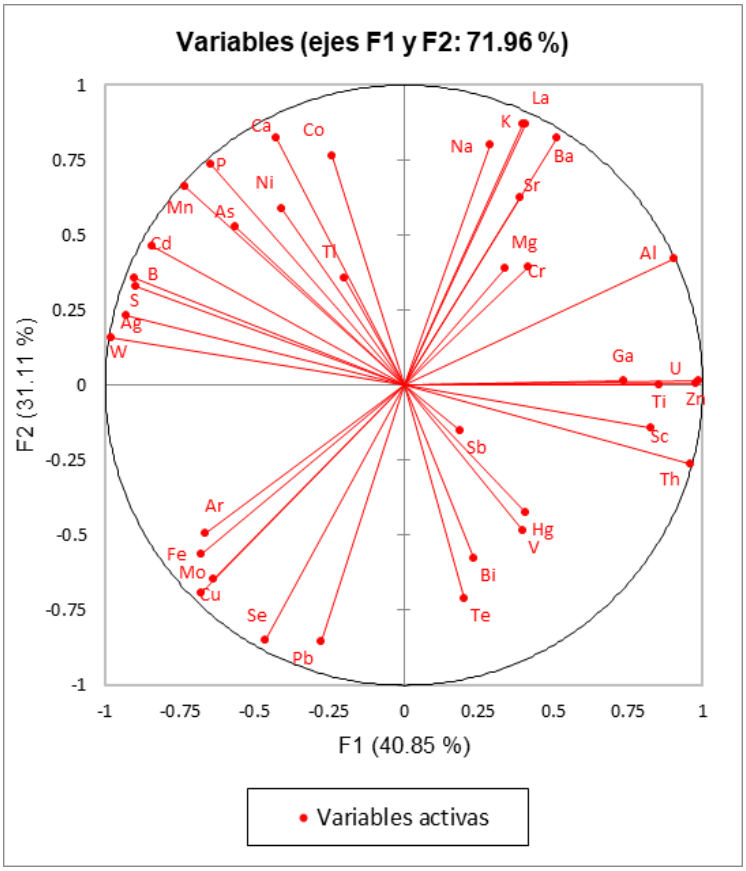
4.3.2 Análisis de Componentes Principales

El coeficiente de correlación indicado en la matriz de correlación de la sección 5.3.1, puede ser examinado por medio del Análisis de Componentes Principales (PCA, siglas en ingles), esto es util cuando se tiene una base multivariable (multi-elementos). El PCA mediante un circulo de correlación manejable a simple vista y con ello poder realizar una

buena interpretación de los datos. En la Figura 16, se muestra mediante la correlación entre los 36 elementos mediante vectores propios, que pueden llegar a +1 y -1. Los vectores propios entre mas cercanos esten, tendrán una buena correlación positiva, independientemente en el cuandrante que se encuentren. Vectores propios que separan entre 90° a 180° tendrán una correlación negativa.

Figura 16.

Círculo de correlación para la identificación de correlación entre los 36 elementos químicos.



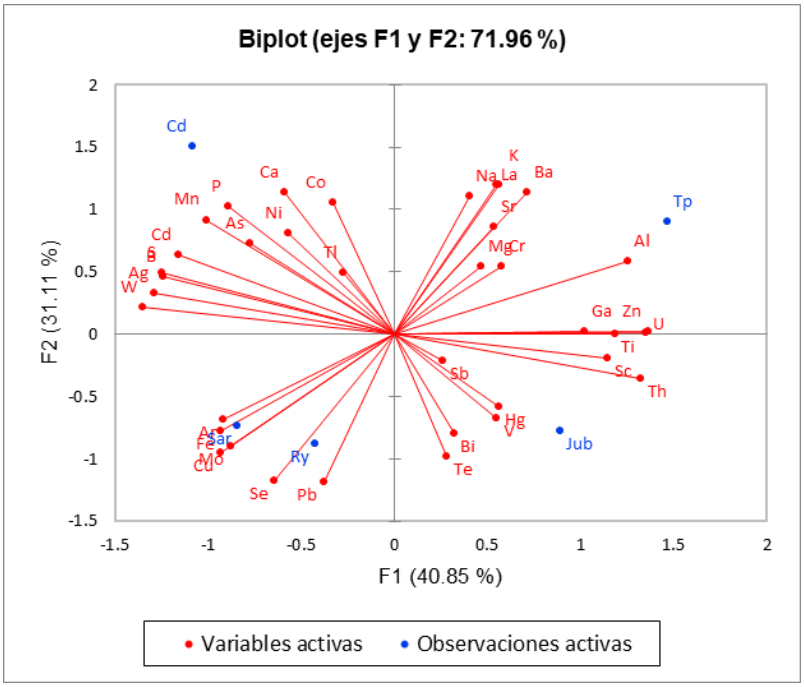
Anteriormente se mencionó que entre Mo-Cu-Pb tenían una correlación positiva, lo cual se evidencia en el Circulo de correlación, estos tres elementos estan cercanos. La asociación geoquímica entre Mo-Cu-Pb representa una mineralización tipo pórfido, formado en base a los elementos químicos que mineralizaron la roca y se asociaron para formar el pórfido. Los tres elementos antes mencionados son calcófilos y geoquímicamente son afines entre sí.

Entre el La (litófilo) y el Cu (calcófilo) se puede apreciar una correlación negativa, sus vectores propios estan en posición opuesta formando un angulo cercano a 180°. Geoquímicamente el La y el Cu no son compatibles entre sí, debido a que cada elemento se encuentra en fases minerales diferentes. El Lantanio es considerado como tierra rara que comunmente se encuentran rocas ígneas, sedimentarias y metamórficas, las cuales se han enriquecido de estos elementos de las tierras raras mediante procesos primarios ígneos o hidrotermales o procesos secundarios sedimentarios. (Martínez & Valle, 2014)

El PCA tambien permite identificar las poblaciones relacionadas entre variables cuantitativas (geoquímica) con las categoricas (litologias). En la Figura 17 se muestra un biplot de correlación entre los tipos de litologias y los diversos elementos químicos.

Figura 17.

Correlación global entre las litologías y los elementos químicos

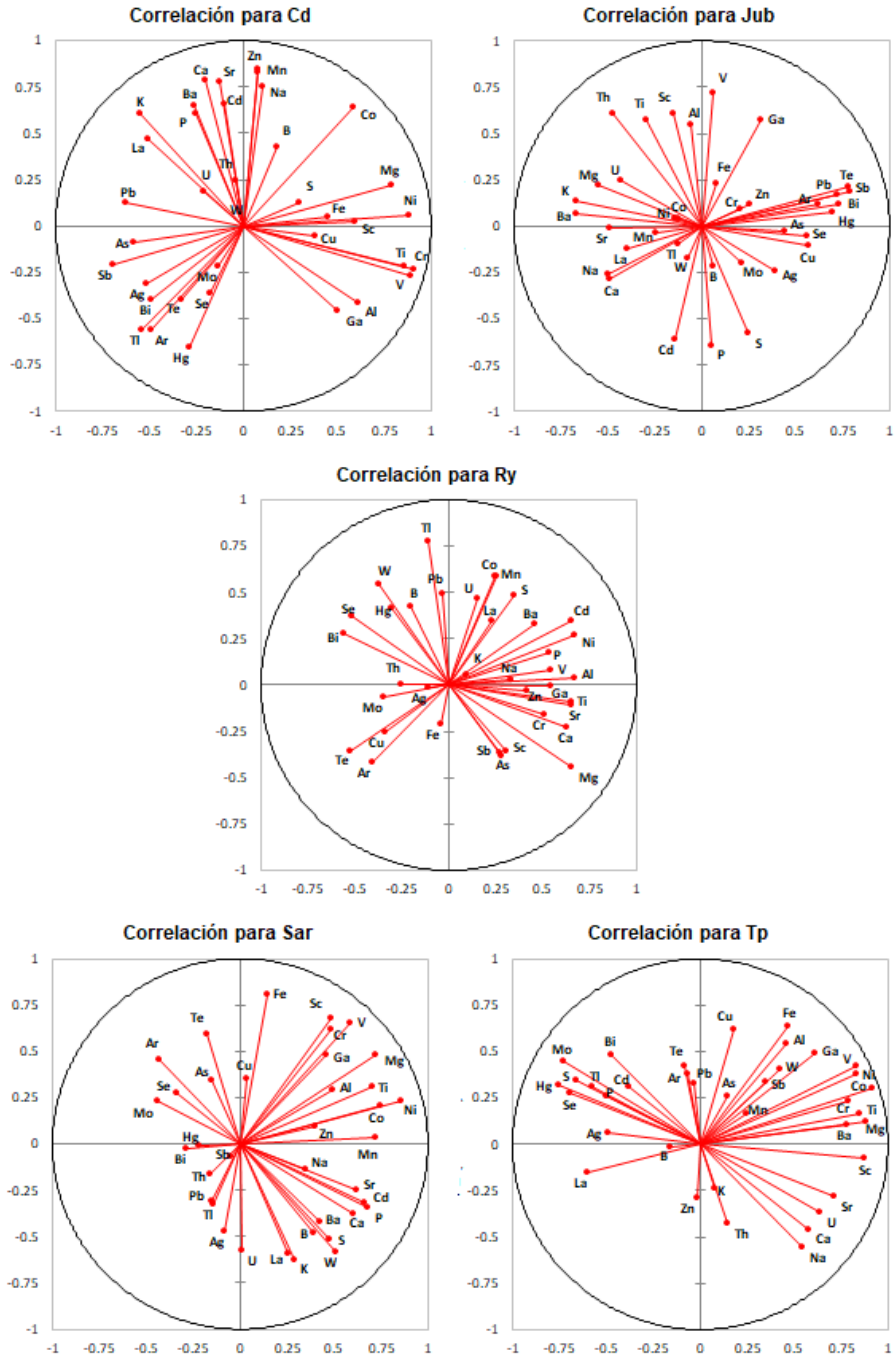


En la figura anterior se puede analizar en forma a priori las relaciones de algunos elementos con las litologias de la zona. Por ejemplo, las Tobas Turupamaba (Tp), tienen mas asociación con elementos químicos como el Al, Ga, Zn, Ba, K, La, Na, entre otros; pero no tiene afinidad cercana con el Fe, Cu, Mo, Se, Pb; esto ultimos tienen afinidad con las litologías Saraguro (Sar) y riolitas (Ry).

La presencia de mayor o menor concentracion de las covariables geoquimicas tambien pueden ser representadas su correlación mediante los circulos de correlación para cada tipo de litologia (variable categorica). En la siguiente ilustración se muestra para cada litologia las correlaciones (Figura 18) que hay entre cada elemento quimico (covariables geoquimicas)

Figura 18.

Correlación de la geoquímica en cada grupo de litología



En la figura anterior, se considera Potasio (K) y Aluminio (Al) quienes usualmente formar minerales félsicos como por ejemplo los feldespatos ($KAlSi_3O_8$), algunas micas [K (Mg, Fe^{2+}) (Al, Fe^{3+}) $Si_3O_{10}(OH, F)_2$], tiene correlación para las rocas ácidas como es el caso de la zona de estudio que son riolitas (Ry), tobas riolíticas ricas en cristales (Jub), rocas andesitas y lavas (Cd) ácidas que dependen de su composición mineralógica tipo félsico y con enriquecimiento de sílice.

4.4 Clasificación supervisada

La aplicación de machine learning como métodos de clasificación supervisada, permitirá determinar la precisión del método a utilizar para predecir las litologías encontradas en base al mapa geológico y su relación con las 36 covariables de geoquímica.

4.4.1 Análisis discriminante

Con este método, la varianza acumulada entre los valores propios entre los dos primeros componentes es de 78.6 %, valor aceptable lo que muestra que existe correlación entre todas las variables. Una primera aproximación de la relación de linealidad entre las covariables geoquímicas y las litologías, el Análisis Discriminante otorga una precisión del 79.39% como lo indica la matriz de confusión de la Tabla 7.

Tabla 7.

Matriz de confusión para los resultados de validación cruzada por Análisis Discriminante

de \ a	Cd	Sar	Ry	Jub	Tp	Total	% correcto
Cd	**10**	2	0	11	0	**23**	43.48%
Sar	1	**15**	8	29	0	**53**	28.30%
Ry	1	8	**4**	53	0	**66**	6.06%
Jub	22	19	4	**718**	13	**776**	92.53%
Tp	0	0	0	26	**12**	**38**	31.58%
Total	**34**	**44**	**16**	**837**	**25**	**956**	**79.39%**

Nota: Cd, andesitas, lavas y brechas; Sar, andesitas y dacitas; Ry, riolitas; Jub, tobas riolíticas ricas en cristales; Tp, tobas turupamba]

La Ry (riolita) es la litología que a posteriori pierde la mayor parte de su codificación en los puntos iniciales, la mayoría de estos datos pasan a ser parte de la litología JUB que comprende las tobas riolíticas ricas en cristales).

Esta primera aproximación, se tratará de mejorar con otros algoritmos de clasificación para conocer su precisión y sus nuevas litologías para cada punto muestreado.

4.4.2 Árbol de decisión y bosque aleatorio

Como parte del machine learning, estos métodos son unos de los utilizados para la clasificación supervisada. Se busca el mejor método de algoritmo de clasificación con la mayor precisión. A continuación, se presenta en una tabla la puntuación de precisión entre los métodos analizados y que fueron indicado en el capítulo de Marco Teórico.

Tabla 8.

Porcentaje de precisión entre los métodos probados

Método	Precisión del método
Chaid	86.51 %
Chaid exh	93.41 %
Cart	85.15 %
Quest	81.17 %
Bosque aleatorio (RDF)	81.38 %

De acuerdo con la Tabla 8 el algoritmo chaid exhaustivo es el que tiene alta puntuación, por consiguiente, con este método se procede a realizar el análisis de clasificación para determinar la predicción de los datos.

4.4.2.1 Métodos de clasificación chaid exhaustivo. La base de datos contiene 956 datos, de estos se obtiene una base de entrenamiento y una base de prueba que corresponde al 20 % de los datos iniciales. Obteniendo una base de entrenamiento de 765 datos (Tabla 9) y una base de prueba con 191 datos (Tabla 10). Esto permite realizar una validación cruzada y conocer la precisión del método. Los resultados obtenidos se muestran en las tablas siguientes.

Tabla 9.

Matriz de confusión para la base de entrenamiento

Litología observada	Litología pronosticada					Total	Precisión %
	Cd	Sar	Ry	Jub	Tp		
Cd	**14**	0	1	3	0	18	77.778
Sar	0	**30**	2	11	0	43	69.767
Ry	1	3	**48**	5	0	57	84.211
Jub	3	12	9	**586**	7	617	94.976
Tp	0	0	0	6	**24**	30	80.000
Total	**18**	**45**	**60**	**611**	**31**	**765**	**91.765**
Precisión de la clasificación							**91.76 %**

Tabla 10.

Matriz de confusión para la base de prueba (Validación)

Litología observada	Litología pronosticada					Total	Precisión %
	Cd	Sar	Ry	Jub	Tp		
Cd	**2**	0	0	2	1	5	40.000
Sar	0	**4**	4	2	0	10	40.000
Ry	0	1	**6**	2	0	9	66.667
Jub	4	5	12	**132**	6	159	83.019
Tp	0	0	0	7	**1**	8	12.500
Total	**6**	**10**	**22**	**145**	**8**	**191**	**75.916**
Precisión de la clasificación							**75.92 %**

La precisión para los dos subconjuntos es alta entre el 75.92 % a 91.76 %. La litología Saraguro (Sar) es la que tiene baja precisión en comparación con las otras litologías, algunos de esos puntos pasan a ser parte de la litología Jubones (Jub), y de la misma forma algunas litologías de la Jub pasan ser parte de la Saraguro (Sar).

La clasificación supervisada permite también obtener una nueva predicción de litologías desde el punto de vista determinístico; en la tabla 11 se muestra la matriz de confusión para toda la base de datos, el cual proporciona una precisión del 93.4 %.

Tabla 11.

Matriz de confusión con la base general.

Litología observada	Litología pronosticada					Total	Precisión %
	Cd	Sar	Ry	Jub	Tp		
Cd	**20**	2	0	1	0	23	86.9
Sar	0	**33**	5	15	0	53	62.2
Ry	0	1	**56**	9	0	66	84.8
Jub	4	4	6	**761**	1	776	98.0
Tp	0	0	0	15	**23**	38	60.5
Total	**24**	**40**	**67**	**801**	**24**	**956**	**93.410**
Precisión de la clasificación							**93.41 %**

A pesar de que los datos de litologías fueron obtenidos del mapa geológico regional, la clasificación a posteriori es buena con una puntuación del 93.41 %, la litología de más predominio es la que corresponde la Jub (tobas riolíticas ricas en cristales). Ciertos datos pasan a ser parte de otras categorías que puede deberse a la vecindad entre los contactos respectivos de cada punto muestreado y/o a la escala en la cual está reflejada la geología regional de la zona de estudio. Con estos nuevos valores de litologías se realizó un mapa geológico predictivo que se lo indica en el Anexo A.

Capítulo cinco

Análisis de resultados

El zinc, magnesio, bario, hierro, aluminio son elementos químicos con altas concentraciones en la zona de estudio, los mismos que tiene una distribución asimétrica positiva. Al correlacionar algunos elementos como el Cu y Pb que pueden estar relacionados a una zona a priori de una mineralización de pórfidos, entre estos dos elementos tienen una correlación > +0.8. El lantanio (La) al ser una tierra rara y no tiene afinidad con sulfuros de cobre, su valor de correlación con el Cu es -0.9 (neagtivo).

Otro análisis se da entre los elementos litófilos torio (Th) y uranio (U) que tienen una correlación positiva, pero estos con los elementos calcófilos como el molibdeno (Mo), plata (Ag), azufre (S) tiene una correlación negativa.

La correlación entre Talio (Tl) y Manganeso (Mn) es alta de 0.725, al ser elementos trazas presentes en las rocas tipo volcánicas félsicas, litología común en la zona de estudio; además el Talio está ampliamente distribuido en la corteza terrestre, aunque en concentraciones muy bajas, y se encuentra asociado con otros metales pesados en piritas y blendas y en los nódulos de manganeso en el lecho de los océanos.

En el análisis de componentes principales se identificó que la Tobas Turupamaba (Tp), no tienen relación con los elementos químicos Fe, Cu, Mo, Se, Pb porque estos formar parte de la litología Saraguro (Sar) y Riolitas (Ry). Siguiendo este análisis entre las covariables geoquímicas y las litologías, considerando los elementos Potasio (K) y Aluminio (Al) que usualmente forman minerales félsicos como por ejemplo los feldespatos ($KAlSi_3O_8$), algunas micas [$K\ (Mg,\ Fe^{2+})\ (Al,\ Fe^{3+})\ Si_3O_{10}(OH,\ F)_2$]. El potasio y aluminio tiene correlación para las rocas ácidas como el caso de la zona de estudio que son riolitas (Ry), tobas riolíticas ricas en cristales (Jub), rocas andesitas y lavas (Cd) ácidas.

El análisis discriminante muestra una relación lineal entre las covariables geoquímicas y las litologías otorgando una precisión del 79.33 %, con la finalidad de mejorar la clasificación de predicción se realizó un análisis más detallado, para lo cual se utilizó los árboles de decisión mediante el método *chaid exh* el cual permitió determinar nuevas clasificaciones

litológicas dando un puntaje de 93.41 %, en el cual la litología de más predominio corresponde a la tobas riolíticas ricas en cristales de la Jub; la litología Sar pierde datos pasando a ser parte de la nueva litología de la Jub. Las tobas de turupamba (Tp) toma datos de la Jubones posible por su afinidad litológica.

Se buscó el mejor método (algoritmo de clasificación), es decir el método que presenta mayor certeza, siendo el chaid exhaustivo con el 93.41 % de certeza. Para corroborar dicha afirmación se realizó una validación cruzada obteniendo una certeza del 75.92 %.

Conclusiones

En este estudio se analizaron 1016 datos, a los cuales se les aplicó el análisis exploratorio de datos (EDA) obteniendo como base final 956 datos, mismos que fueron analizados mediante el análisis multivariable concluyendo que este método permite simplificar los datos con la mínima perdida de información. Además, mediante el análisis multivariante se puede predecir el grado de relación que existe entre las variables.

Se concluye que la mayor concentración de elementos químicos se encuentra en la litología Jubones misma que prevalece en sector Suroeste de la zona Tambo-Azuay.

Al aplicar el árbol de decisiones se concluye el método que mayor eficacia presenta es el Chaid exhaustivo, con un 93.41 % de certeza.

Mediante el árbol de decisiones se obtuvo el nuevo modelo geológico con la clasificación de la tabla por objetos (a priori y a posteriori) se concluye que existe una modificación de los elementos químicos en los diferentes tipos de litologías a piori vs posteriori porque existe una depuración de datos; es decir se presenta una reclasificación dentro de las observaciones entre los distintos tipos de litologías.

Recomendaciones

Antes de empezar a aplicar la estadística descriptiva se recomienda realizar el análisis exploratorio de datos para obtener datos limpios, es decir que no existan duplicados.

Al manejar una base de datos bastante grande se recomienda hacer uso del análisis multivariable, misma que ayuda a la simplificación de información para una mejor comprensión de esta.

Para mejorar la predicción litológica se deberá realizar simulaciones geoestadísticas para luego aplicar el análisis multivariable, con ello subir el porcentaje de precisión y conocer los nuevos valores en sitios no muestreados.

Referencias

Acosta, D. (2014). *Perfil de los clientes que aceptan una tarjeta de crédito de un banco vía call center utilizando el algoritmo Chaid Exhaustivo*. Obtenido de http://repositorio.lamolina.edu.pe/bitstream/handle/UNALM/2274/E13-A23-T.pdf?sequence=1&isAllowed=y

Alperin, M. (2013). *Introducción al análisis estadístico de datos geológicos*. Obtenido de http://naturalis.fcnym.unlp.edu.ar/repositorio/_documentos/sipcyt/bfa003805.pdf

Barbosa, P., Oliveira, T., Silva, J., 2010. (s.f.). Regionalized classification of multivariate geochemical data from Jacupiranga Alkaline Complex (Ribeira de Iguape Valley/Sao Paulo, Brazil). *Revista Brasileira de Geociencias*, 40(2): 212-219.

Barnett, R. M. (2015). Imputación multivariante de variables geológicas muestreadas de manera desigual. 47(7), 791–817.

Berlanga, V., Rubio, M., & Vilà, R. (01 de 08 de 2013). *Cómo aplicar árboles de decisión en SPSS.* Obtenido de https://revistes.ub.edu/index.php/REIRE/article/viewFile/reire2013.6.1615/7229

Breiman, L. (1996). *Statistics Department, University of California. (Berkele) CA 94720. .* California: Ross Quinlan.

CLIRSEN,SIGAGRO-MAGAP. (noviembre de 2011). ¨gestion de geoinformatica en las areas de influencia de los proyectos estrategicos nacionales¨. chaucha, Azuay, ecuador.

Cuadras, C. (02 de 02 de 2007). *NUEVOS MÉTODOS DE ANÁLISIS MULTIVARIANTE*. Obtenido de http://www.est.uc3m.es/esp/nueva_docencia/getafe/estadistica/analisis_multivariante/doc_generica/archivos/metodos.pdf

De la Fuente Fernandez, S. (2011). *Análisis Discriminante*. Obtenido de https://www.fuenterrebollo.com/Economicas/ECONOMETRIA/SEGMENTACION/DISCRIMINANTE/analisis-discriminante.pdf

De la Fuente, S. (2011). *Componentes Principales*. Obtenido de https://www.estadistica.net/Master-Econometria/Componentes_Principales.pdf

Diaz, L. G. (2007). *ESTADÍSTICA MULTIVARIADA: INFERENCIA Y MÉTODOS*. Obtenido de http://ciencias.bogota.unal.edu.co/fileadmin/Facultad_de_Ciencias/Publicaciones/Imagenes/Portadas_Libros/Estadistica/Estadistica_Multivariada_Inferencia_y_Metodos/Estadistica_multivariada_inf..pdf

Dunkley, P & Gaibor, A. (1997). *Informe No.2, Proyecto de Desarrollo Minero y Control Ambiental, Programa de Información Cartográfica y Geológica: Geology of the Western Cordillera of Ecuador between 2-3°S. CODDGEM-BGS.* Quito, Ecuador.

Espinoza, J. (2015). “EL PAISAJE RURAL EN LA PARROQUIA CHAUCHA”. Cuenca , Azuay, Ecuador .

Galton F. (1877). *Typical laws of heredity. Proceedings of the Royal Institution 8.*

Grunsky E & Caritat P. (2020). State-of-the-art analysis of geochemical data for mineral. *Geochemistry: Exploration, Environment, Analysis, 20*, 217 - 232. Obtenido de https://geea.lyellcollection.org/content/geochem/20/2/217.full.pdf

Grunsky, E. C. (2010). *The interpretation of geochemical survey data. Geochemistry: Exploration, Environment and Analysis.*

Grunsky, E., & Smee, B. (1999). *The differentiation of soil types and mineralization from multi-element geochemistry using multivariate methods and digital topography.* Journal of Geochemical Exploration 67.

Hungerbuhler D, et al. (Enero de 1997). Neogene stratigraphy and Andean geodynamics of southern Ecuador. *ScienceDirect/ ELSEVIER, 57*, 75-124. Obtenido de https://sci-hub.do/10.1017/S0016756800010499

ILION,SISTEMS. (2019). *GAD,Chaucha*. Obtenido de http://chaucha.gob.ec/azuay/?p=122

INSTITUTO GEOGRÁFICO MILITAR (IGM) DEL ECUADOR. (18 de 11 de 2002). *Hoja 7. Geológica de Cuenca (versión actual). Escala:1, 100000.* Obtenido de INSTITUTO DE INVESTIGACION GEOLOGICO Y ENERGETICO Copyright 2018: https://drive.google.com/file/d/1JAPCFuGB7q6XWjrcBxTAIy1kwwhZ9lm6/view

Jenny, H. (1941). *Factors of Soil Formation. A System of Quantitative Pedology.* New York: McGraw Hill. Obtenido de file:///C:/Users/Pc%20one/Downloads/HANS_JENNY__1941_FACTORS_OF_SOIL_FORMATION_-_A_System_of_Quantitative_Pedologypdf.pdf

Litherland, M; et al. (1994). The metamorphic Belts of Ecuador. *Institute of Geological Sciences, 11 de Overseas memories*.

Martínez R., et al. (2009). EL COEFICIENTE DE CORRELACION DE LOS RANGOS DE SPEARMAN CARACTERIZACION. *Revista Habanera de Ciencias Médicas, Vol. 8*(Núm. 2). Obtenido de https://www.redalyc.org/pdf/1804/180414044017.pdf

Martínez, J., & Valle, A. (2014). *Las tierras raras: un sector estratégico para el desarrollo tecnológico de China*. Obtenido de https://dusselpeters.com/CECHIMEX/CuadernosdelCechimex20146.pdf

McBratney, A., et al. (2003). On digital soil mapping. *ELSEVIER / SCIENCE DIRECT, 117*, 3. Obtenido de https://sci-hub.mksa.top/10.1016/s0016-7061(03)00223-4

Misión Belga. (1986). *Informe Final, Estudio de Yacimiento de Cobre Porfídico de Chaucha. CODIGEM.* Quito, Ecuador.

Nordberg, G. (s.f.). *Metales propiedades químicas y toxicas* . Obtenido de https://www.insst.es/documents/94886/162520/Cap%C3%ADtulo+63.+Metales+propiedades+qu%C3%ADmicas+y+toxicidad

Oleas, R. (1999). *Geostatistics for engineers and earth scientists. Kluwer Academic Publishers, Norwell.* Massachusetts, USA.

Peralta, R. (2018). *Determinación de dispersión geoquímica de Pb en sedimentos de afluentes del área de incidencia del proyecto minero Loma Larga*. Obtenido de file:///C:/Users/Pc%20one/Downloads/13950.pdf

Pino, P. (2017). *Evaluación del riesgo crediticio mediante árboles de clasificación y bosques aleatorios*. Obtenido de http://bibing.us.es/proyectos/abreproy/91149/fichero/MEMORIATFG.pdf

Poldervaart, A. (15 de July de 1955). Chemistry of the Earth Crust. In: Poldevaart, A. (ed.). Crust of the Earth. A Symposium. Geol. Soc. Am. Spec. *Paper 62*, 119-144. Recuperado el 2021, de Crust of the Earth. A Symposium: https://watermark.silverchair.com/spe62-fm.pdf?token=AQECAHi208BE49Ooan9kkhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAAkowggJGBgkqhkiG9w0BBwagggI3MIICMwIBADCCAiwGCSqGSIb3DQEHATAeBglghkgBZQMEAS4wEQQMdJqmD5oEwMlyv-aAAgEQgIIB_VQyjXRYs5-YIgIc5aSD7k3K_Gf57cdz9Vd7GpQUldLyV3

Pratt, W., et al. (1997). *Informe N º 1; Proyecto de Desarrollo Minero y Control Ambiental, Programa de Información Cartográfica y Geológica: Mapa escala 1: 200.000. Geology of the Cordillera Occidental of Ecuador between 3° S and 4° S. CODIGEM - BGS.* Quito, Ecuador.

Reiman, C., Filmozer, P., & Dutter, R. (2008). Statistical data analysis explained: Applied enviromental statistics with R. 343 p.

Rodríguez, M. (28 de 10 de 2009). *Análisis exploratorio de datos*. Obtenido de http://rua.ua.es/dspace/handle/10045/12075?mode=full

Ronov, A., & Yarosheysky, A. (1976). A new model for the chemical structure of the Earth's Crust. *Geochem, 13*, 89-121.

Sarango, S. (2013). *Elementos Calcófilos*. Obtenido de https://es.scribd.com/doc/124261987/Elementos-Calcofilos-docx

Servicio geológico Mexicano. (2017). *Magma.* Obtenido de https://www.sgm.gob.mx/Web/MuseoVirtual/Informacion_complementaria/Magma.html#:~:text=La%20fase%20pegmat%C3%ADtica%20tiene%20lugar,originando%20yacimientos%20filonianos%20de%20pegm%C3%A1titas.

Snelling, N. (1970). *K-Ar Determinations on Samples from Ecuador. (Int. Rep. Institute of Geological Sciences, London).*

Stevens, J. (1984). Valores atípicos y puntos de datos influyentes en el análisis de regresión. Boletín psicológico . *95(2)*, 334–344.

Touchette, P. E., MacDonald, R. F., & Langer, S. N. (1985). A scatter plot for identifying stimulus control of problem behavior. Journal of Applied Behavior Analysis. 18 (4), 343–351.

Tukey, J. (1977). *Exploratory data analysis.* Obtenido de http://www.ru.ac.bd/wp-content/uploads/sites/25/2019/03/102_05_01_Tukey-Exploratory-Data-Analysis-1977.pdf

U.S. Geological Survey (USGS). (2006). FGDC digital cartographic standard for geologic map symbolization (post script implementation), Survey Techniques and Methods. (Rep. 11-A2, U.S. Geol. Surv., Reston, Va.). Obtenido de https://www.fgdc.gov/standards/projects/geo-symbol/index_html

U.S. Geological Survey (USGS). (2006). *FGDC digital cartographic standard for geologic map symbolization (post script implementation), Survey Techniques and Methods Rep. 11-A2, U.S. Geol. Surv., Reston, Va.* Obtenido de file:///C:/Users/Pc%20one/Downloads/FGDC-GeolSymFinalDraft.pdf

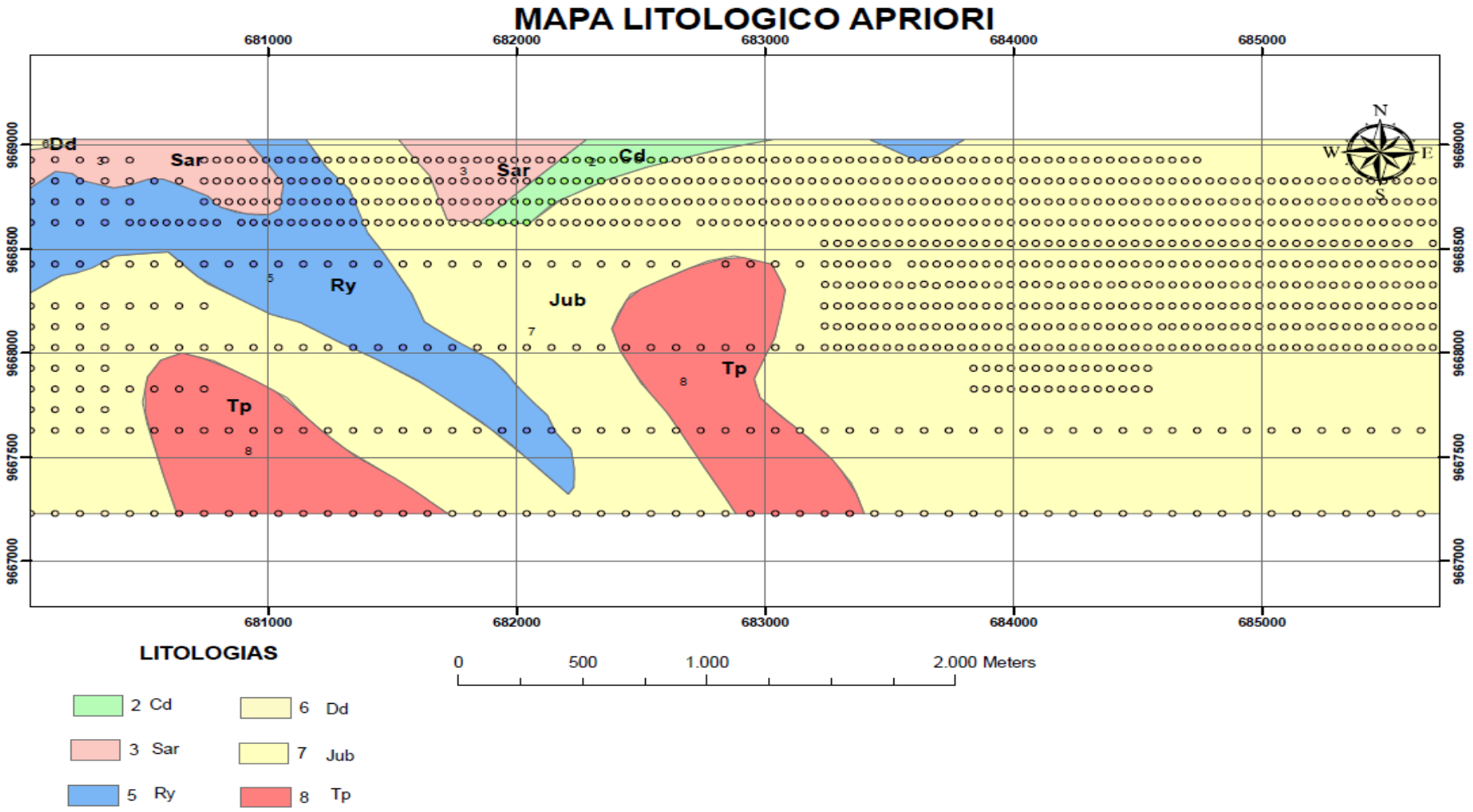
Vásquez, E. (2017). Obtenido de http://repositorio.unap.edu.pe/bitstream/handle/UNAP/8858/Vasquez_Choque_Edwin_Aderly.pdf?sequence=1&isAllowed=y

Williamson, D. F. (1989). The Box Plot: A Simple Visual Method to Interpret Data. Annals of Internal Medicine. 110(11), 916.

Apéndice

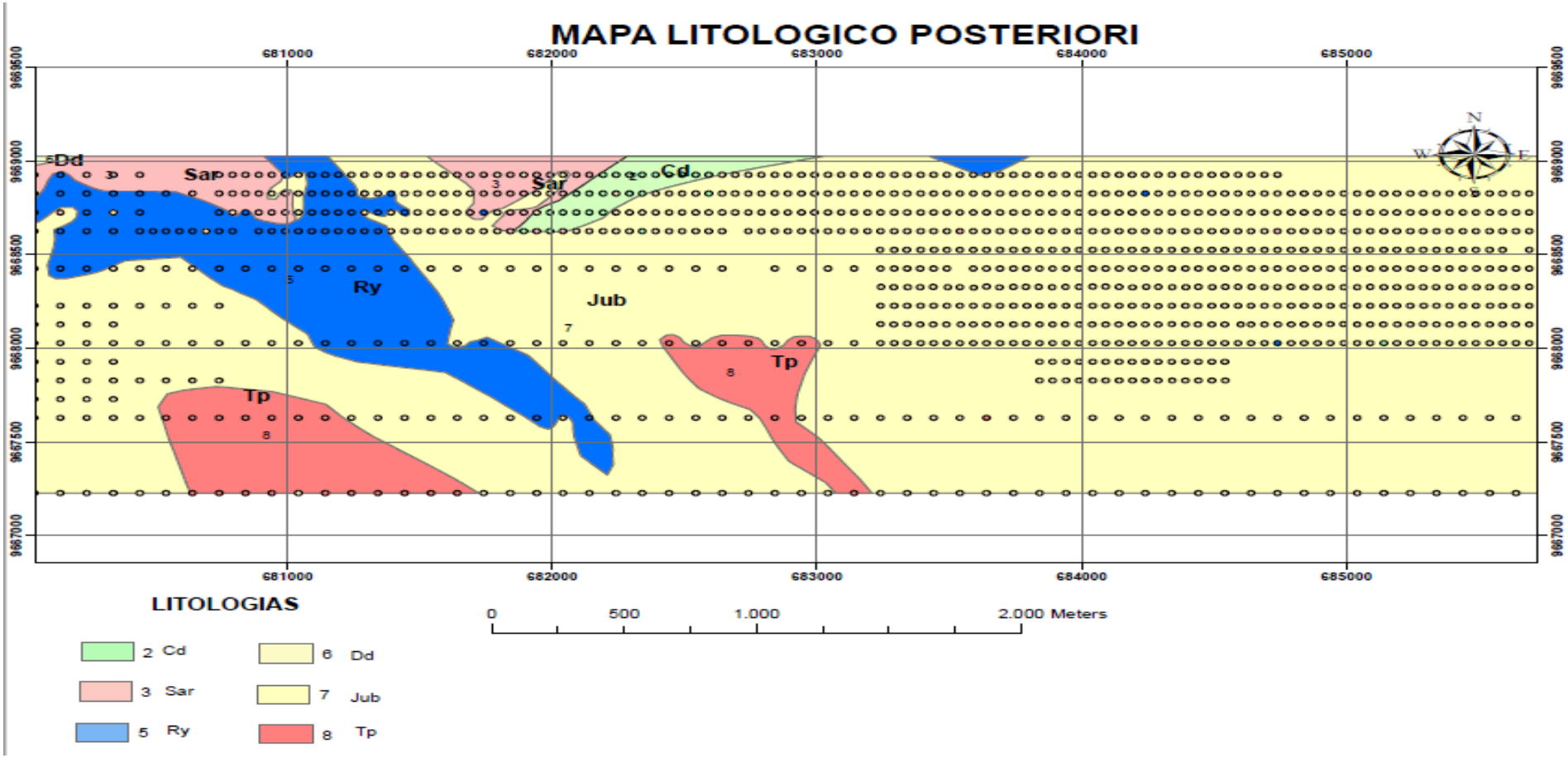
Apéndice A. Mapa litológico a priori

Figura 19. *Mapa litológico a priori*



Apéndice B. Mapa geológico a posteriori

Figura 20. *Mapa litológico realizado con la clasificación a posteriori*



Fuente: Sergio Rojas