
Aplicación de Análisis de Componentes
Principales y Conglomerados a datos de
Fluorescencia de Rayos X (FRX) en suelos



Presentado por:

Merydoreya Moreno Velásquez

LOS LIBERTADORES
FUNDACIÓN UNIVERSITARIA

Fundación Universitaria Los Libertadores
Facultad de Ingeniería y Ciencias Básicas
Especialización en Estadística Aplicada
Bogotá D.C, Colombia
2019

Aplicación de Análisis de Componentes Principales y Conglomerados a datos de Fluorescencia de Rayos X (FRX) en suelos

Presentado por:

Merydoreya Moreno Velásquez

en cumplimiento parcial de los requerimientos para optar al título de

Especialista en Estadística Aplicada

LOS LIBERTADORES
Quien asesora

FUNDACIÓN UNIVERSITARIA
John González Veloza

Profesor

Fundación Universitaria Los Libertadores
Facultad de Ingeniería y Ciencias Básicas
Especialización en Estadística Aplicada
Bogotá D.C, Colombia
2019

Notas de aceptación



LOS LIBERTADORES

FUNDACIÓN UNIVERSITARIA

Firma del presidente del jurado

Firma del jurado

Firma del jurado

Bogotá DC, noviembre de 2019.



LOS LIBERTADORES

FUNDACIÓN UNIVERSITARIA

Las directivas de la Fundación Universitaria Los Libertadores, los jurados calificadores y el cuerpo docente no son responsables por los criterios e ideas expuestas en el presente documento. Estos corresponden únicamente a los autores y a los resultados de su trabajo.

Índice General

Capítulo 1	2
Introducción	2
Capítulo 2	3
Planteamiento del problema	6
Objetivos	7
Objetivo general	7
Objetivos específicos.....	7
Justificación.....	7
Capítulo 3	7
Marco Teórico	7
Fluorescencia de Rayos X (FRX).....	7
Rayos X.....	7
Fluorescencia.....	7
Fluorescencia de Rayos X (FRX).....	7
Análisis Univariado.....	8
Análisis Multivariado.....	10
Plancha 5, Alta Guajira, Colombia.....	12
Localización	12
Suelos	13
Unidades litológicas	14
Capítulo 4	16
Marco Metodológico	16
Método de Estudio	16
Método	16
Enfoque	16
Fases.....	16
Selección de la base de datos	16
Análisis estadístico.....	17

Análisis de resultados.....	18
Capítulo 5.....	19
Análisis y Resultados	19
Análisis Descriptivo Univariado	19
Medidas de tendencia central, dispersión y forma.	19
Prueba de normalidad:.....	20
Análisis Descriptivo Multivariado	20
Matriz de correlaciones	20
Análisis de Componentes Principales (ACP).....	21
Conglomerados de K medias.....	22
Análisis de Cluster Jerárquico (CJ).....	25
Capítulo 6.....	26
Conclusiones y Recomendaciones	26
Conclusiones	26
Recomendaciones.....	26
Bibliografía	28

Índice de figuras

Figura 5-1. Matriz de Correlación de Pearson	20
Figura 5-2. Izquierda: círculo de correlaciones. Derecha: calidad de la representación de las variables en los componentes principales.	22
Figura 5-3. Izquierda: número óptimo de conglomerados. Derecha: Conglomerados de K medias / ACP.	23
Figura 5-4. Izquierda: dendograma de individuos para los metales analizados. Derecha: conglomerados obtenidos al cortar, en 3 clases, el árbol de individuos.	25

Índice de tablas

Tabla 5-1. Análisis descriptivo univariado de metales en suelos de la plancha 5, Alta Guajira, Colombia. ^a Datos en g por 100g, ^b Datos en mg por Kg.	19
Tabla 5- 2. Resultados del test de normalidad Lilliefors.	20
Tabla 5- 3. Componentes Principales.	21
Tabla 5- 4. Componentes principales 1 y 2: promedio variables por conglomerados. ^a Datos en g por 100g, ^b Datos en mg por Kg.	23
Tabla 5- 5. Cluster jerárquico: promedio variables por cluster. ^a Datos en g por 100g, ^b Datos en mg por Kg.	25

Aplicación de Análisis de Componentes Principales y Conglomerados a datos de Fluorescencia de Rayos X (FRX) en suelos

Resumen

Se realizó la categorización preliminar de los suelos de la de la Plancha 5, Alta Guajira, Colombia, muestreadas por el Servicio Geológico Colombiano, entre marzo y agosto de 2007. El ACP realizado conllevó la reducción de las dimensiones iniciales de 7 a 2; las dimensiones seleccionadas, 1 y 2, acumularon el 76,01 % de la varianza y correspondieron a aquellas con valores propios mayores a 1. Las variables que más contribuyeron al CP1 fueron Fe, Al, V, Zn y Mg, y las que más contribuyeron al CP2 fueron Si y Ca. En el análisis de conglomerados se encontró que las muestras se dividen en tres grupos: Conglomerado 1: presenta el contenido más alto de Si, más bajo de Mg, Ca, V y Zn e intermedio de Fe y Al; Conglomerado 2: presenta el contenido más alto de Ca, más bajo de Si, Fe y Al e intermedio de Mg, V y Zn. Conglomerado 3: presenta un contenido bajo de Ca, intermedio de Si y el más alto de Al, Fe, Mg, V y Zn. Se concluyó que la combinación de las técnicas ACP, Conglomerados K-Medias y Conglomerados Jerárquicos, es una herramienta efectiva para hacer una categorización preliminar de muestras de suelos en una zona de interés geoquímico, con base a los resultados de FRX.

Palabras clave: Fluorescencia de rayos X, Análisis de Componentes Principales, Análisis de conglomerados, suelos, unidades litológicas.

Capítulo 1

Introducción

En este documento se presenta el trabajo de grado titulado “Aplicación de Análisis de Componentes Principales y Conglomerados a Datos de Fluorescencia de Rayos X (FRX) en Suelos”.

Inicialmente se hace el planteamiento del problema de investigación resaltando la importancia de aplicar técnicas estadísticas multivariadas para hacer una categorización preliminar de los suelos muestreados en una zona de interés geoquímico para facilitar la posterior caracterización fisicoquímica y mineralógica de los mismos.

Con base a la pregunta de investigación se define como objetivo general: “Aplicar Análisis de Componentes Principales (ACP) y conglomerados para analizar resultados de Fluorescencia de Rayos X (FRX) con el fin de hacer una clasificación preliminar de suelos de origen colombiano según su composición química”.

En el marco teórico se incluyen los temas que son críticos para el desarrollo de la idea de investigación: a) Presentación general de la técnica de Fluorescencia de Rayos X (FRX), b) Explicación del análisis estadístico descriptivo univariado y multivariado aplicado y c) Descripción de la zona de estudio; localización, suelos y unidades litológicas.

En el marco metodológico se expone que se realizó una investigación descriptiva con un enfoque mixto la cual se desarrolló en tres etapas principales: selección de la base de datos, aplicación de técnicas estadísticas descriptivas univariadas y multivariadas y realización del análisis de resultados.

En el capítulo de análisis y resultados se presentó el análisis descriptivo univariado y multivariado realizado resultando en la obtención de tres grupos principales para categorizar las muestras de suelos estudiadas.

Finalmente se presentan las conclusiones y recomendaciones del estudio realizado, siendo la principal conclusión que la combinación de las técnicas de ACP y análisis de conglomerados es una herramienta efectiva para hacer una categorización preliminar de las muestras de suelos en una zona de interés geoquímico, con base a los resultados de FRX.

Capítulo 2

Planteamiento del problema

Dos de las técnicas multivariadas más empleadas para analizar resultados químicos son el Análisis de Componentes Principales (ACP) y el análisis de conglomerados: el primero permite reducir la dimensionalidad de los datos facilitando la interpretación de la información contenida en los datos y el segundo permite categorizar las muestras en una serie de clases de manera que los objetos similares se encuentren en la misma clase.

La técnica de Fluorescencia de Rayos X es un método moderno de análisis químico de gran aplicación en diferentes campos como la geología y la geoquímica. Permite la cuantificación no destructiva y rápida de varios elementos simultáneamente sin procedimientos complejos de preparación de muestra. Aun cuando no es necesario aplicar técnicas estadísticas complejas para hacer el análisis de los resultados obtenidos mediante esta técnica, hay múltiples tareas analíticas en las que aplicar técnicas como el ACP y el análisis de conglomerados puede ayudar a superar los métodos tradicionales de procesamiento de datos, como por ejemplo en la caracterización de suelos con base a su composición elemental.

El suelo no es un medio simple y estático, es complejo y dinámico y sus propiedades distintivas se adquieren lentamente a través del tiempo, bajo la acción combinada de los factores y procesos del medio donde se localiza (IGAC, 2009). En la zona de estudio seleccionada, Alta Guajira, Colombia, los factores que más han influenciado la formación de los suelos han sido el clima y el material parental. En este territorio se encuentran ubicadas varias unidades litológicas (Grupo Macuira, Formación Jarara, Formación de Uitpa, entre otras) correspondientes a rocas sedimentarias y metamórficas, que originan suelos heterogéneos con una composición química variada.

Teniendo en cuenta la complejidad de los suelos y la heterogeneidad que estos pueden presentar en términos de su composición elemental en una misma zona de estudio, se hace indispensable emplear una herramienta estadística que permita hacer una categorización preliminar de los suelos con base a los resultados obtenidos mediante FRX para facilitar su caracterización.

Pregunta de investigación

¿Es viable emplear Análisis de Componentes Principales (ACP) y conglomerados para analizar resultados de Fluorescencia de Rayos X (FRX) con el fin de hacer una clasificación preliminar de suelos de origen colombiano según su composición química?

Objetivos

Objetivo general

Aplicar Análisis de Componentes Principales (ACP) y conglomerados para analizar resultados de Fluorescencia de Rayos X (FRX) con el fin de hacer una clasificación preliminar de suelos de origen colombiano según su composición química.

Objetivos específicos

- Seleccionar una base de datos que incluya resultados de FRX para muestras de suelos de origen colombiano y delimitar la zona de estudio.
- Aplicar ACP y conglomerados a la zona de estudio seleccionada para categorizar las muestras según su composición química.
- Analizar cada categoría obtenida tomando como referencia las unidades litológicas de la zona de estudio.

Justificación

El Análisis de Componentes Principales y Conglomerados puede facilitar la interpretación de los resultados de Fluorescencia de Rayos X obtenidos en estudios de caracterización geoquímica de suelos de origen Colombiano. Los estudios de suelos establecen lineamientos que permiten a los planificadores y especialistas, adelantar programas y proyectos que orienten la ocupación, manejo y utilización adecuada de las tierras de una manera más precisa (IGAC, 2009). Se escogió la Alta Guajira como área de estudio porque a pesar de ser una zona semidesértica (IGAC, 2012) presenta condiciones geológicas favorables para la formación de ciertos tipos de rocas y minerales de interés industrial (Ingeominas, 2009).

Capítulo 3

Marco Teórico

Fluorescencia de Rayos X (FRX)

Rayos X

Hay tres interacciones principales de los rayos X con la materia: Fluorescencia, Dispersión de Compton y Dispersión de Rayleigh. Cuando los rayos X caen sobre un material, una fracción pasa a través de la muestra, otra es absorbida y produce radiación fluorescente, y la restante es dispersada de regreso. La dispersión puede ocurrir con pérdida o sin pérdida de energía. La primera es llamada dispersión de Compton y la segunda dispersión de Rayleigh (Brouwer, 2003).

Fluorescencia

La fluorescencia es la emisión de radiación (luz visible, UV o RX) por parte de una sustancia, en respuesta a una excitación, que cesa casi inmediatamente al desaparecer su causa. La excitación es suministrada por una radiación incidente de mayor energía, radiación primaria, que la emitida por la sustancia, radiación secundaria (Brouwer, 2003).

Fluorescencia de Rayos X (FRX)

El análisis de fluorescencia de rayos X es uno de los métodos populares de análisis químico. Permite la cuantificación no destructiva y rápida de varios elementos simultáneamente sin procedimientos complejos de preparación de muestra. FRX ha encontrado una amplia aplicación en geoquímica, ciencia de materiales, arqueología, etc. (Panchuk V., 2018).

El tiempo de medición depende del número de elementos que se va a determinar y la precisión requerida, y varía entre segundos y 30 minutos. El tiempo de análisis después de la medición es solo de unos pocos segundos (Brouwer, 2003).

Los sistemas de espectrómetros pueden ser divididos en dos grandes grupos: sistemas por dispersión de energía (FRX-ED) y sistemas por dispersión de longitud de onda (FRX-WD). Los elementos que se pueden analizar y sus niveles de detección dependen principalmente del sistema usado. El rango elemental para FRX-ED va desde Sodio hasta Uranio (Na a U). Para FRX-WD es aún más amplio, desde Berilio hasta Uranio (Be a U). El rango de concentraciones va desde (sub) ppm hasta 100%. En general, los elementos con números atómicos altos tienen mejores límites de detección que los elementos ligeros (Brouwer, 2003).

La Fluorescencia de Rayos X (FRX) es una técnica espectroscópica que utiliza la emisión secundaria o fluorescente de radiación X generada al excitar una muestra con una fuente de radiación X (Martinez, Gil, & Barón, 2006). En la mayoría de los casos la fuente es un tubo de rayos X. Los elementos presentes en la muestra emitirán fluorescencia de rayos X con energías discretas que son características para cada elemento. Mediante la medición de las energías de la radiación emitida por la muestra es posible determinar cuáles elementos están presentes. Este paso es llamado Análisis Cualitativo. Mediante la medición de las intensidades de las energías emitidas es posible determinar cuánto de cada elemento está presente en la muestra. Este paso es llamado Análisis Cuantitativo (Brouwer, 2003).

Análisis Estadístico Descriptivo

Análisis Univariado

Medidas de posición

- **Media aritmética:** se define como la suma de todas las observaciones de una variable dividida entre el número de ellas. Dada su definición, la media aritmética es muy sensible a los valores extremos de la variable (Castillo & Guijarro, 2006).
- **Mediana:** su característica principal es que divide un conjunto de valores ordenados en dos grupos iguales; la mitad de los números tendrá valores que son menores que la mediana y la otra mitad alcanzará valores mayores (Monroy, 2008). Aunque la media es la medida de localización central más empleada, en algunas ocasiones se prefiere la mediana porque este no es sensible a valores extremos. (Anderson, Sweeny, & Williams, 2008).

Medidas de dispersión

- **Varianza:** corresponde a la suma de cuadrados de las distancias entre la media y cada elemento de la población, dividida entre el número total de observaciones de la población N (Levin & Rubin, 2004). Cuando se tienen datos de una muestra y no de toda la población se calcula la varianza muestral; la suma de cuadrados de las distancias entre la media y cada elemento de la muestra se divide entre el número total de observaciones de la muestra menos 1, $n - 1$. La varianza se representa como s^2 cuando se refiere a la varianza muestral y σ^2 cuando se trata de la varianza de una población (Monroy, 2008).
- **Desviación estándar:** se define como la raíz cuadrada positiva de la varianza. Continuando con la notación adoptada para la varianza muestral y la varianza poblacional, se emplea s para denotar la desviación estándar muestral y σ para la desviación estándar poblacional (Newbold, Carlson, & Thorne, 2008).
- **Coefficiente de variación:** es la relación de la desviación estándar de una muestra con respecto al valor absoluto de su promedio, se denota C_v (Selvamuthu & Das, 2018).

Medidas de forma

- **Coefficiente de asimetría:** indica la simetría de la distribución de una variable respecto a la media sin hacer una evaluación gráfica. Si la distribución es simétrica, esto es, cuando a la derecha y a la izquierda de su media aritmética existe el mismo número de valores de la variable, a la misma distancia de la media y con la misma frecuencia, este coeficiente es nulo, siendo positivo o negativo si la distribución es asimétrica positiva o asimétrica negativa, respectivamente (Castillo & Guijarro, 2006).
- **Coefficiente de curtosis:** es una medida numérica que fue diseñada para analizar el grado de concentración que presenta un conjunto de datos alrededor de la zona central de la distribución de frecuencias sin necesidad de generar el gráfico. Distribución mesocúrtica: el grado de concentración de un conjunto de datos al redor de sus valores centrales es el mismo que presenta una distribución normal. Distribución Leptocúrtica: el grado de concentración de un conjunto de datos al redor de sus valores centrales es mayor que el presentado por una distribución normal. Distribución Platicúrtica: el grado de concentración de un conjunto de datos al redor de sus valores centrales es menor que el presentado por una distribución normal (Tellez & Lemus, 2014).

Pruebas de normalidad:

- **Prueba de Kolmogorov – Smirnov (KS):** conocida como prueba KS, es una prueba de significación estadística para verificar si los datos de la muestra proceden de una distribución normal. Se emplea para variables cuantitativas continuas y cuando el tamaño muestral es mayor a 50 (Romero, 2016). Esta prueba compara la distribución acumulativa de los datos con la distribución acumulativa normal esperada, y basa su p valor en la mayor discrepancia (Oztuna, Elhan, & Tüccar, 2006). Si el valor p es mayor que el nivel de significancia escogido no se rechaza la hipótesis nula y se concluye que la población se distribuye normalmente.
- **Prueba de Kolmogorov – Smirnov – Lilliefors (KSL):** es una modificación a la prueba Kolmogorov – Smirnov. La prueba KS es apropiada cuando los parámetros de la distribución hipotética son completamente conocidos. Sin embargo es difícil especificar dichos parámetros cuando la distribución es desconocida (Mohd & Bee, 2011); aplicar la prueba bajo estas circunstancias puede conllevar errores de interpretación (Lilliefors, 1967). La prueba Lilliefors difiere de la prueba KS porque estima los parámetros de la población que son desconocidos. La tabla de valores de las dos pruebas son diferentes, lo que resulta en diferentes decisiones (Mendes & Pala, 2003).

Análisis Multivariado

Análisis de Componentes Principales (ACP)

El propósito del análisis de componentes principales es reducir un espacio de dimensión p a un nuevo espacio de dimensión d , donde d es mucho menor que p . Con el ACP, se transforman los datos en un nuevo conjunto de coordenadas o variables que son una combinación lineal de las variables originales. Además, las observaciones en el nuevo espacio de componentes principales no están correlacionadas. Se espera obtener información y comprensión de los datos al analizar las observaciones en el nuevo espacio (Sánchez, Cruz, Posada, Torrijos, & Osorio, 2016).

La información de un conjunto de datos dado corresponde al total de la variación que este contiene. El objetivo del ACP es identificar las direcciones (o componentes principales) a través de los cuales la variación en los datos es máxima. En otras palabras, el APC reduce la dimensionalidad de los datos multivariados a dos o tres componentes principales, que se pueden visualizar gráficamente, con una pérdida mínima de información (Kassambara, 2017).

Análisis de Conglomerados:

Es un método para dividir un grupo de objetos en una serie de clases de manera que los objetos similares se encuentren en la misma clase. Como en el APC, los grupos no se suelen conocer antes de realizar el análisis matemático y no se realiza ningún supuesto sobre la distribución de las variables. El análisis de conglomerados busca objetos que se encuentren próximos en el espacio de las variables. Como en el ACP, se ha de tomar la decisión de si se estandarizan o no los datos. La estandarización de los datos significará que todas las variables se miden en una escala común de manera que una variable no domina a las otras (Miller & Miller, 2002).

Existen dos grandes tipos de análisis de conglomerados: no jerárquicos y jerárquicos. Los **métodos jerárquicos** buscan agrupar cluster para formar uno nuevo o separar alguno ya existente para dar origen a otros dos de forma que se maximice una medida de similaridad o se minimice alguna distancia. Los **métodos no jerárquicos** están diseñados para la clasificación de individuos (no de variables) en k grupos. El procedimiento es elegir una partición de los individuos en k grupos e intercambiar los miembros de los clusters para tener una partición mejor (Fernández, 2011).

- **Conglomerados de K-Medias:** es uno de los métodos no jerárquicos más empleados. Este método clasifica objetos en múltiples grupos (ej., conglomerados), de manera que los individuos en un mismo conglomerado sean tan similares como sea posible (ej., alta similaridad intra clase) mientras que los objetos de diferentes grupos sean tan diferentes como sea posible (ej., baja interclase similaridad). En conglomerados de K-Medias, cada conglomerado es representado por su centro (ej., centroide) el cual corresponde al promedio de los puntos asignados al conglomerado (Kassambara, Practical Guide to Cluster Analysis in R, 2017).

La idea básica detrás del conglomerado de K-Medias consiste en definir conglomerados de manera que la variación total intra conglomerados (conocida como la variación total dentro del conglomerado) sea minimizada. Hay muchos algoritmos K-Medias disponibles. El algoritmo estándar es el algoritmo de Hartigan-Wong (1979), el cual define la variación total dentro de los conglomerados como la suma de los cuadrados de las distancias Euclidianas entre los ítems y los correspondientes centroides (Kassambara, Practical Guide to Cluster Analysis in R, 2017).

- **Conglomerados Jerárquicos:** consiste en clasificar los datos en grupos con estructura arborescente de dependencia, de acuerdo con diferentes niveles de jerarquía. Partiendo de tantos grupos iniciales como individuos se estudian, se trata de conseguir agrupaciones sucesivas entre ellos de forma que progresivamente se vayan integrando en clusters que, a su vez, se unirán entre sí en un nivel superior formando grupos mayores que posteriormente se juntarán hasta llegar al cluster final que contiene todos los casos analizados. La representación gráfica de estas

etapas de formación de grupos, a modo de árbol invertido, se denomina dendograma (Fernández, 2011).

Todas estas agrupaciones se toman en función de similaridad multivariante (o de su contrario, la distancia) proporcionada por el conjunto de variables estudiadas, ya que en cada nivel de jerarquía se unen los dos cluster más cercanos. En esta línea, como paso previo a un análisis de clusters es muy importante la elección de una adecuada métrica de similaridad o disimilaridad. A partir de una tabla inicial de $n \times p$ datos es preciso calcular una matriz de distancias entre individuos $n \times n$. La distancia euclídea es la más sencilla y utilizada, también se emplea en el análisis de componentes principales cuyos factores son muchas veces datos previos para entrar en un análisis de clusters (Fernández, 2011).

Existen varios métodos de unión de algoritmos de clasificación jerárquica. Uno del más empleados es el método aglomerativo de Ward. **Método de Ward** (método de mínima varianza): cuando se unen dos conglomerados, con independencia del método utilizado, la varianza aumenta. El método de Ward une los casos buscando minimizarla varianza dentro de cada grupo. Para ello se calcula, en primer lugar, la media de todas las variables en cada conglomerado. A continuación, se calcula la distancia entre cada caso y la media del conglomerado, sumando después las distancias entre todos los casos. Posteriormente se agrupan los conglomerados que generan menos aumentos en la suma de las distancias dentro de cada conglomerado. Este procedimiento crea grupos homogéneos y con tamaños similares (Fernández, 2011).

Plancha 5, Alta Guajira, Colombia

Localización

El Departamento de la Guajira, se encuentra situado en el extremo Norte de Colombia, en la parte más septentrional de la América del Sur, enmarcada entre las coordenadas geodésicas $10^{\circ} 23'$ y $12^{\circ} 28'$ Latitud Norte y $71^{\circ} 06'$ y $73^{\circ} 39'$ de longitud Oeste (IGAC, 1996).

Este departamento limita por el Norte y el Este con el mar Caribe y la República de Venezuela, por el sur con el Departamento el César (franja de territorio en litigio), y por el Oeste con el Departamento del Magdalena (franja de territorio en litigio) y el mar Caribe y ocupa una superficie total de 20.848 km^2 (IGAC, 1996).

La zona de estudio corresponde a la **Plancha 5** del Mapa índice establecido por el IGAC. Esta plancha se encuentra localizada en la región natural Alta Guajira, en el extremo

peninsular. Esta región es semidesértica, de escasa vegetación, con presencia de algunas serranías que no sobresalen de los 865msnm tales como Jarara y Simarúa (IGAC, 1996).

Suelos

A pesar de mostrar propiedades comunes, el conjunto de suelos de la Alta Guajira es variado y contrastante en muchos aspectos. Por ejemplo, hay suelos poco o nada evolucionados como los aluviales en los fondos de los arroyos o al pie de los grandes macizos, los depósitos eólicos litorales o continentales, y los suelos truncados de las mesas litorales desarrollados sobre materiales calcáreos consolidados (INGEOMINAS, 2009). Horizontes con acumulaciones de carbonato de calcio son frecuentes en muchos de los suelos de la zona, lo mismo que los colores rojizos que indican procesos de rubeficación (liberación del hierro de los minerales primarios), la dispersión de partículas de óxido de Fe y su progresiva oxidación o hidratación (Cordoba & Cortes, 1979).

Génesis de suelos:

El desarrollo genético de los suelos está influido por una serie de factores (clima, relieve, material parental, organismos y tiempo) y procesos (ganancias y pérdidas) cuya combinación e interacción dan lugar a la diferenciación progresiva de horizontes que, a medida de que pasa el tiempo, construyen el perfil del suelo (IGAC, 2009).

Los principales factores que han determinado la formación de los suelos de la Alta Guajira son el clima y el material parental.

- **Clima:** como en la mayor parte del mundo, en la Alta Guajira el clima ha sido y es el factor más importante en la génesis de suelos. La deficiencia de precipitación, las altas temperaturas, la pérdida excesiva de agua por evapotranspiración, los vientos fuertes y constantes y la incidencia directa de la energía solar, han contribuido a la formación de los suelos típicos de la zona, cuyas características más sobresalientes (Cordoba & Cortes, 1979) son: moderada a alta salinidad, deficiencia de humedad, abundancia de carbonatos, tendencia a la compactación de los materiales, erosión intensa, presencia de capas delgadas o espesas de arenas transportadas por el viento, materia orgánica escasa y pH básico (INGEOMINAS, 2009).
- **Material Parental:** para poder comprender la naturaleza y origen de los suelos deben conocerse las principales características físicas y químicas de los minerales primarios (cuarzo, ortoclasa, plagioclasa, mica moscovita, hornblenda, augita, olivino, magnetita y apatita). La naturaleza química o elementos que constituyen un

mineral, indica la contribución que el mineral podrá hacer al estado nutritivo del suelo a que da origen; la fuente principal de potasio en el suelo es la ortoclasa y la plagioclasa lo es de calcio y sodio; las micas dan potasio, magnesio y hierro. El material parental se ha considerado como un factor de formación pasivo en la génesis y evolución de los suelos y constituye la materia prima sobre la cual actúan los factores dinámicos de clima y organismos; sin embargo, su constitución mineral afecta la evolución y las características de los suelos resultantes.

Unidades litológicas

Las unidades litológicas que se encuentran en la Alta Guajira se describen de acuerdo al tipo de roca así: unidades metamórficas, ígneas y sedimentarias (INGEOMINAS, 2009). En la Plancha 5 de la Alta Guajira, predominan las unidades metamórficas y sedimentarias:

Metamórficas:

- **Grupo Macuira:** aflora en el municipio de Uribia, localmente en la serranía de Macuira, en el cerro Hualahuálohu, en las serranías de Jarara y Simarúa y en el municipio de Dibulla, cuenca del río Palomino. Está constituido por el Neis de Uray, compuesto por neises de hornblenda y plagioclasa, y los esquistos de Jutúruhu de almandino, plagioclasa, cuarzo y biotita (IGAC, 2009).
- **Metasedimentitas de Alas:** aflora en el suroeste de la serranía de Jarara, municipio de Uribia; están constituidas por filitas, mármoles con restos de fósiles, metaareniscas arcósicas y metaconglomerados (IGAC, 2009).
- **Formación Jarara:** localizada en el municipio de Uribia, aflora en la parte occidental de las serranías de Macuira, Jarara y Simarúa y se extiende como una franja alargada en dirección noreste. Constituida principalmente por filitas que gradan a esquistos moscovíticos y sedimentitas cuarzo-arenosas metamorfoseadas y localmente cuarcitas, neises y rocas hornbléndicas (IGAC, 2009).
- **Formación Ipanaruhu:** Cinturón alargado de 14 km de largo y entre 1 y 3 km de ancho, que conforman la parte más occidental del área de Simarúa; aflora en las lomas de Ipanaruhu y Maíz el municipio de Uribia. Esta constituida principalmente por filitas compuestas por clorita, sericita, cuarzo, limolitas y localmente se encuentran lentes de calizas y fragmentos de fósiles (IGAC, 2009).

Sedimentarias:

- **Formación de Siamaná:** Aflora en la serranía de Macuira, cerca de las poblaciones de Parajimaruhu, Ichipa y Anteshiruhu; en la serranía de Jarara cerca al cerro Karariapiau y en la base del cerro Jimol, Uitpa y en el arroyo Koushotpa; en la serranía de Simarúa al sureste del municipio de Taparajín y en la serranía de Cosinas al oeste y noroeste de la población de la Flor de la Guajira. Constituida por calizas arenosas, calizas arrecifales, calizas margosas, margas y shales arcillosos.

- **Formación de Uitpa:** Afloramiento de esta unidad se presentan en los alrededores de la serranía de Jarara, al noroeste de la población de Anteshiruhu, al norte del ares de Simarúa, al este de Uitpa, y en un carretable que une esta población con Huinkuaru. Constituida por shales arcillosos, shales y en menor proporción por capas de arenisca y calizas arenosas; en la zona de Jarara está constituida por lodolitas limosas yesíferas y cantidades menores de limolitas, areniscas, margas y algunas capas delgadas de calizas.

- **Formación Jimol:** Aflora en los cerros cercanos al carretable que une las poblaciones de Uitpa y Siamaná y Siamaná - Castilletes, al este de la serranía de Jarara, en la población de Anteshiruhu, entre las serranías de Jarara Y Macuira. Constituida por calizas arenosas, areniscas calcáreas y cantidades menores de arcillolitas; a lo largo de la parte sur de la cuenca y hacia la bahía de Cosinetas, se presenta una secuencia de hasta 10m de espesor de calizas muy puras, algáceas y arrecifales con cabezas de corales y restos de conchas.

Capítulo 4

Marco Metodológico

Método de Estudio

Método

Se realizó una investigación descriptiva: con base a los resultados obtenidos del análisis estadístico descriptivo univariado y multivariado realizados, se establecieron características comunes para hacer una categorización preliminar, según su composición elemental, de los individuos objeto de estudio.

Enfoque

Enfoque mixto: Se integraron datos cuantitativos y cualitativos para hacer una discusión conjunta de los resultados obtenidos en el análisis estadístico realizado. Lo anterior con el fin de hacer una categorización preliminar de los suelos presentes en la zona de estudio.

Fases

Este estudio se desarrolló en tres etapas principales: Selección de la base de datos, aplicación de técnicas estadísticas descriptivas univariadas y multivariadas y la realización del análisis de resultados tomando como referencia las unidades litológicas de la zona de estudio.

Selección de la base de datos

Teniendo en cuenta que el principal objetivo de este estudio era aplicar Análisis de Componentes Principales (ACP) y conglomerados para analizar resultados de

Fluorescencia de Rayos X (FRX) con el fin de categorizar suelos de origen colombiano según su composición geoquímica, se hizo la búsqueda de una base de datos libre que incluyera este tipo de datos.

Base de datos

Nombre: Geoquímica base de datos.

Fuente: Servicio Geológico Colombiano.

Descripción: Contiene la Base de Datos de Geoquímica con datos hasta el año 2010, colocada en el Geoportal del SGC - Gestión de Información - Productos Geocientíficos - Bases de Datos.

Archivo: gq_AN_2010.mdb

Link: <https://www2.sgc.gov.co/sgc/mapas/Paginas/Bases-Datos.aspx>

Datos empleados para el estudio

Proyecto: Alta Guajira.

Zona de estudio: Plancha 5, Uribia, Alta Guajira, Guajira, Colombia.

Fecha muestreo: Entre marzo y agosto de 2007.

Escala Muestreo: 1: 25000.

Técnica Analítica de medición: Fluorescencia de Rayos X.

Tipo de muestras: Suelos.

Análisis estadístico

Análisis descriptivo univariado

Se llevó a cabo un análisis descriptivo univariado estándar para describir el contenido de elementos mayores, menor y trazas en las muestras de suelo estudiadas. Los estadísticos empleados fueron:

- Mínimo y máximo.
- Mediana y promedio (medidas de posición).
- Desviación estándar y coeficiente de variación (medidas de dispersión).
- Coeficiente de asimetría y curtosis (medidas de forma).
- Prueba de Lilliefors (Test de normalidad).

Análisis descriptivo multivariado

Con el fin de evaluar si era posible categorizar los individuos objeto de estudio en grupos que presentarán una composición química similar se emplearon las siguientes técnicas de análisis estadístico multivariado:

- Análisis de Componentes Principales (ACP).
- Análisis de conglomerados K-medias
- Análisis de conglomerados jerárquico.

Análisis de resultados

Teniendo en cuenta que uno de los principales factores determinantes de la composición química de los suelos es el material parental del cual se originan, se tomaron como referencia las unidades litológicas de la zona de estudio para interpretar los resultados obtenidos.

Capítulo 5

Análisis y Resultados

Análisis Descriptivo Univariado

Medidas de tendencia central, dispersión y forma.

Se llevó a cabo un análisis estadístico estándar para describir el contenido de elementos mayores, menores y trazas en las muestras de suelo estudiadas.

	Si ^a	Al ^a	Fe ^a	Mg ^a	Ca ^a	V ^b	Zn ^b
n	95	95	95	95	95	95	95
Mínimo	28.24	5.16	1.07	0.38	0.28	19	21
Máximo	84.57	19.94	10.31	1.94	25.13	94	181
Mediana	64.2	9.75	3.76	0.9	1.39	92	86
Promedio	61.96	10.11	4.12	0.96	5.39	94	95
s	14.15	2.85	1.6	0.37	6.69	37.23	36.7
Cv (%)	22.84	28.19	38.87	38.68	124.10	39.80	38.68
Coefficiente asimetría	-0.54	0.59	1.33	0.56	1.23	0.53	0.56
Coefficiente curtosis	-0.67	0.67	2.84	-0.37	0.31	-0.04	-0.43

Tabla 5-1. Análisis descriptivo univariado de metales en suelos de la plancha 5, Alta Guajira, Colombia. ^a Datos en g por 100g, ^b Datos en mg por Kg.

Al evaluar las concentraciones promedio de los 7 metales estudiados se encontró que Si, Al, Fe y Ca correspondieron a elementos mayores con concentraciones superiores al 1.0%, el Mg correspondió a un elemento menor con concentraciones entre 0.5 y 1.0% y tanto el V como el Zn con concentraciones inferiores al 0.5% correspondieron a elementos trazas. Se obtuvo un coeficiente de variación alto para todos los metales; el Ca es el elemento con mayor variación.

Todos los metales, exceptuando el Si, presentaron asimetría positiva. El Si, Mg, V y Zn presentaron una distribución Platicúrtica, por el contrario, el Al, Fe, y Ca presentaron una distribución Leptocúrtica. Los valores de los coeficientes de asimetría y curtosis dieron

indicios de que los datos no se distribuían normalmente; con el fin de corroborar este hecho se realizó la prueba de normalidad Lilliefors.

La alta variación y asimetría de los datos está relacionada con la heterogeneidad del territorio estudiado; en la zona predominan los paisajes de lomerío estructural erosional y planicie fluvio-marina. En estos paisajes los suelos presentan diferente composición química según su litología: hay suelos provenientes de areniscas, conglomerados, sedimentos aluviales mixtos y estratos calcáreos y margosos con presencia de sales (IGAC, 2009).

Prueba de normalidad:

Variable	Si	Al	Fe	Mg	Ca	V	Zn
D	0.1193	0.0820	0.1112	0.0847	0.2619	0.0834	0.1219
p-valor	0.0020	0.1198	0.0056	0.0894	2.200E-16	0.0993	0.0014

Tabla 5- 2. Resultados del test de normalidad Lilliefors.

Al aplicar el test de Lilliefors con un nivel de significancia de 0.05, se encontró que solamente Al, Mg y V presentaron una distribución normal.

Análisis Descriptivo Multivariado

Matriz de correlaciones

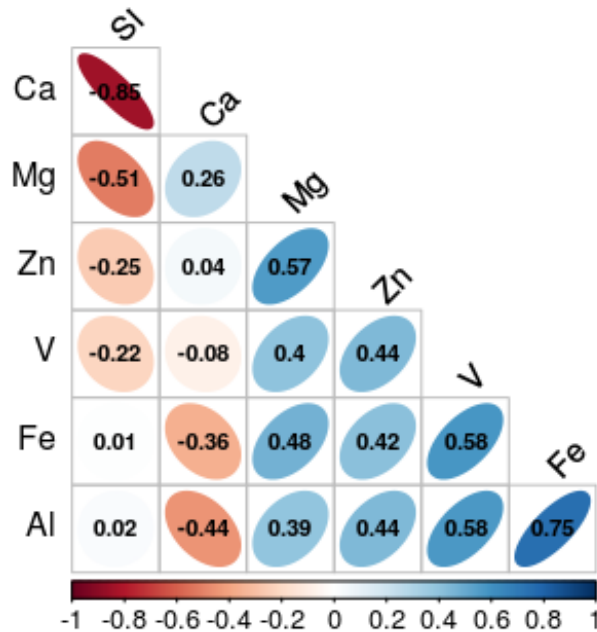


Figura 5-1. Matriz de Correlación de Pearson

Según los coeficientes de correlación de Pearson (r), se evidenció una correlación fuerte negativa entre Si-Ca, y fuerte positiva entre Al-Fe; una correlación positiva media entre Mg, Zn, V, Fe y Al y media negativa entre Si-Mg, Si-Zn, Si-V, Ca-Mg, Ca-Fe y Ca-Al; finalmente una correlación débil, casi nula, entre Si-Fe, Si-Al, Ca-Zn y Ca-V. Estas correlaciones se confirmaron mediante el análisis de componentes principales.

De los factores que determinan la formación de los suelos del departamento de la Guajira, uno de los que más afecta la composición química de los mismos es el material parental. El material parental aporta los minerales primarios (cuarzo, ortoclasa, plagioclasa, mica moscovita, mica biotita, hornblenda, augita, olivino, magnetita y apatita) que forman los suelos; la naturaleza química o elementos que constituyen un mineral, indican la contribución que el mineral podrá hacer al estado nutritivo del suelo al que da origen (IGAC, 2009).

La fuerte correlación negativa Si-Ca, la correlación media negativa entre Si y Ca con respecto a los demás metales estudiados, y la correlación positiva intermedia entre Mg, Zn, V, Fe y Al, permitió suponer que se tenían tres grupos de suelos de distinto origen: el primero posiblemente derivado de silicatos primarios como cuarzo, anfíboles o piroxenos que presentan un mayor contenido de Si; el segundo de rocas carbonatadas con un mayor contenido de Ca como las calizas, dolomitas o margas; y el tercero de micas o agregados de arcillas con óxidos de Fe, V, Zn y Mg.

Análisis de Componentes Principales (ACP)

		CP 1	CP 2	CP 3	CP 4	CP 5	CP 6	CP 7
Contribución de cada variable	Si	2.53	37.26	6.13	0.79	12.49	0.01	40.79
	Al	21.43	6.32	1.74	2.67	39.86	19.39	8.59
	Fe	22.24	4.23	2.57	14.11	0.18	56.62	0.04
	Mg	17.20	8.33	8.40	22.34	28.68	14.95	0.10
	Ca	0.50	42.45	1.85	0.09	0.94	3.80	50.37
	V	19.28	0.02	30.21	33.94	14.91	1.62	0.02
	Zn	16.83	1.39	49.10	26.04	2.94	3.60	0.10
Valor Propio		3.12	2.20	0.61	0.48	0.30	0.23	0.05
Varianza (%)		44.56	31.45	8.76	6.92	4.31	3.24	0.76
Σ Varianza (%)		44.56	76.01	84.77	91.69	96.00	99.24	100

Tabla 5- 3. Componentes Principales.

El ACP conllevó la reducción de las dimensiones iniciales de 7 a 2; las dimensiones seleccionadas, 1 y 2, acumularon el 76,01 % de la varianza y correspondieron a aquellas con valores propios mayores a 1. Las variables que más contribuyeron al CP1 fueron Fe, Al, V, Zn y Mg, y las que más contribuyeron al CP2 fueron Si y Ca.

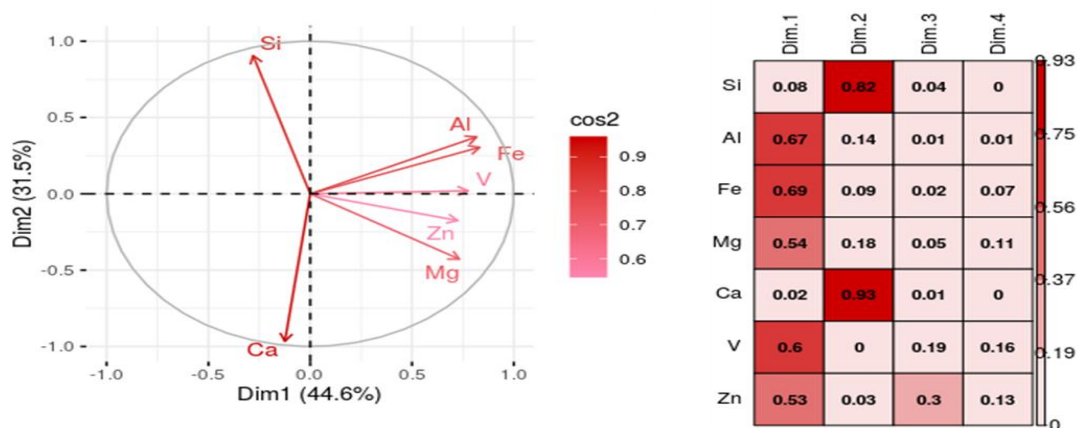


Figura 5-2. Izquierda: círculo de correlaciones. Derecha: calidad de la representación de las variables en los componentes principales.

En el círculo de correlaciones se obtuvieron resultados similares a los obtenidos en la matriz de correlación de Pearson, figura 4-1: las variables Mg, Zn, V, Fe y Al se correlacionaron positivamente, entre estas variables las más correlacionadas fueron Al y Fe; también se encontró que las variables Si y Ca se correlacionaron negativamente entre sí y con respecto a las demás variables.

El valor de coseno cuadrado, \cos^2 , se utilizó para estimar la calidad de la representación de las variables en cada componente principal: las variables mejor representadas en la CP1 fueron Al, Fe, Mg y V, y las mejor representadas en la CP2 fueron Si y Ca. El Zn correspondió a la variable con la menor calidad de representación en los componentes principales 1 y 2, sin embargo, entre estas dos dimensiones la suma de los cosenos cuadrados fue mayor a 0.50.

Conglomerados de K medias.

Se formaron conglomerados de K medias tomando como referencia el plano factorial y las coordenadas obtenidas para cada individuo mediante el ACP.

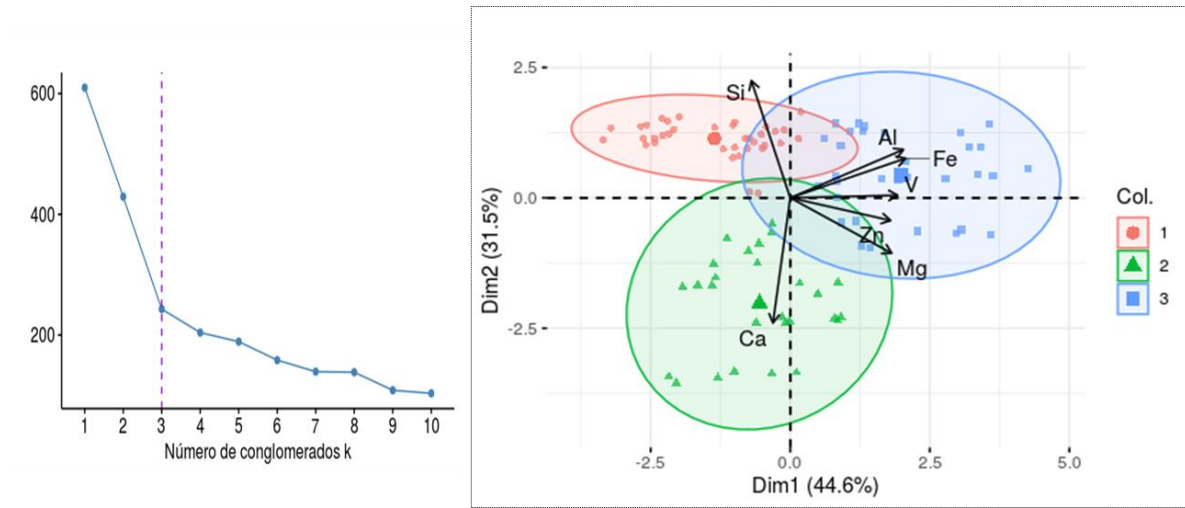


Figura 5-3. Izquierda: número óptimo de conglomerados. Derecha: Conglomerados de K medias / ACP.

Al graficar la varianza dentro de los conglomerados en función de k, se observó que estos tienen poca varianza a partir del tercero, motivo por el cual se decidió hacer el análisis de k medias con k=3, lo cual confirmó que los individuos objeto de estudio se subdividían en tres grupos de acuerdo a su composición elemental.

El eje 1 del plano factorial obtenido del ACP se relacionó con la variación de la concentración de Mg, Zn, V, Fe y Al, la cual aumentó de izquierda a derecha. El eje 2 se asoció con la variación de la concentración de Si y Ca; el Si aumentó de abajo hacia arriba y el Ca presentó un comportamiento opuesto.

Grupos	n	Variables						
		Si ^a	Al ^a	Fe ^a	Mg ^a	Ca ^a	V ^b	Zn ^b
1	36	74.81	9.18	3.43	0.66	0.89	71	69
2	27	45.39	7.96	3.13	1.03	14.95	87	94
3	32	61.49	12.97	5.72	1.23	2.38	124	125

Tabla 5- 4. Componentes principales 1 y 2: promedio variables por conglomerados. ^a Datos en g por 100g, ^b Datos en mg por Kg.

Con el fin de explicar las características de los suelos categorizados en cada conglomerado, se tomaron como referencia las unidades litológicas de la Plancha 5, Alta Guajira,

Colombia. Dichas unidades se encuentran descritas en el mapa geológico del departamento de la Guajira, escala 1:250000, versión 2, Ingeominas, 2002.

Conglomerado 1:

Incluyó el 37,9 % de los individuos estudiados; presentó el contenido más alto de Si, más bajo de Mg, Ca, V y Zn e intermedio de Fe y Al. Este grupo contiene suelos posiblemente derivados de rocas metamórficas del Grupo Macuira y la Formación Jarara. El Grupo Macuira está constituido por el Neis de Uray, compuesto por neises de hornblenda y plagioclasa, y los Esquistos de Jutúruhu de almadino, plagioclasa, cuarzo y biotita (IGAC, 2009). La formación Jarara está constituida principalmente por filitas y sedimentitas cuarzo-arenosas metamorfoseadas y localmente cuarcitas, neises y rocas hornbléndicas. (IGAC, 2009).

Conglomerado 2:

Incluyó el 28,4 % de los individuos estudiados; presentó el contenido más alto de Ca, más bajo de Si, Fe y Al e intermedio de Mg, V y Zn. Este grupo contiene suelos posiblemente derivados de rocas sedimentarias de la Formación Siamaná y la Formación de Jimol y rocas metamórficas de la unidad Metasedimentitas de Alas. La formación Siamaná está constituida por calizas arenosas, calizas arrecifales, calizas margosas, margas y shales arcillosos (IGAC, 2009). La Formación de Jimol está constituida por calizas arenosas, areniscas calcáreas y cantidades menores de arcillolita. Las Metasedimentitas de Alas están constituidas, entre otros, por mármoles con restos de fósiles (IGAC, 2009).

Conglomerado 3:

Incluyó el 33,7% de los individuos estudiados; presentó un contenido bajo de Ca, intermedio de Si y el más alto de Al, Fe, Mg, V y Zn. Este grupo contiene suelos posiblemente derivados de rocas sedimentarias de la Formación Uitpa mezclados con óxidos de hierro de la Formación Jimol. La formación Uitpa constituye la principal fuente de arcillas en la Alta Guajira ya que posee altos niveles de arcillas puras (Ingeominas, 2009). Adicionalmente en esta Formación fueron encontradas dentro de sus niveles arcillosos, abundantes manifestaciones de glauconita, un mineral del grupo de las micas (Ingeominas, 2009). En la formación Jimol es relativamente común la presencia de zonas superficiales de intensa oxidación de hierro en algunos niveles arenosos (Ingeominas, 2009).

Análisis de Cluster Jerárquico (CJ)

Se realizó un análisis de conglomerados jerárquicos partiendo de la información obtenida en el ACP.

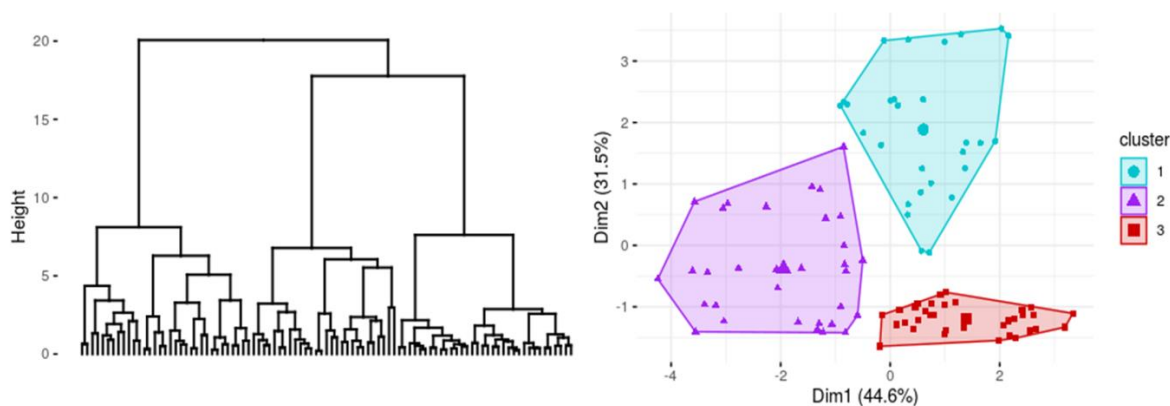


Figura 5-4. Izquierda: dendograma de individuos para los metales analizados. Derecha: conglomerados obtenidos al cortar, en 3 clases, el árbol de individuos.

En el dendograma de individuos se evidenció que estos se dividían en 3 conglomerados principales, motivo por el cual se decidió cortar el árbol en 3 clases. Los conglomerados formados presentan características similares a los obtenidos por K-medias, hecho que corrobora que los suelos de la zona se dividen en tres grandes grupos: a) con un contenido más alto de Ca, intermedio de Mg, V y Zn y bajo de Si, Fe y Al, b) con un contenido más alto de Al, Fe, Mg, V y Zn, intermedio de Si y bajo de Ca, c) con un contenido más alto de Si, intermedio de Fe y Al, y bajo de Mg, Ca, V y Zn.

Grupos	n	Variables						
		Si ^a	Al ^a	Fe ^a	Mg ^a	Ca ^a	V ^b	Zn ^b
1	28	46.27	8.02	3.18	1.03	14.29	86	89
2	33	61.34	12.87	5.67	1.22	2.77	123	126
3	34	75.49	9.15	3.37	0.64	0.61	71	69

Tabla 5- 5. Cluster jerárquico: promedio variables por cluster. ^a Datos en g por 100g, ^b Datos en mg por Kg.

Capítulo 6

Conclusiones y Recomendaciones

Conclusiones

- La combinación de las técnicas ACP, Conglomerados K-Medias y Conglomerados Jerárquicos, es una herramienta efectiva para hacer una categorización preliminar de muestras de suelos en una zona de interés geoquímico, con base a los resultados de FRX.
- Las muestras de suelos de la Plancha 5, Alta Guajira, Colombia, muestreadas por el Servicio Geológico Colombiano, entre marzo y agosto de 2007, en el marco del Proyecto de la Alta Guajira, se pueden dividir en 3 grupos; Conglomerado 1: presenta el contenido más alto de Si, más bajo de Mg, Ca, V y Zn e intermedio de Fe y Al; Conglomerado 2: presenta el contenido más alto de Ca, más bajo de Si, Fe y Al e intermedio de Mg, V y Zn. Conglomerado 3: presenta un contenido bajo de Ca, intermedio de Si y el más alto de Al, Fe, Mg, V y Zn.
- La similaridad de las muestras incluidas en cada conglomerado tiene sentido geoquímico y se puede explicar con base a las unidades litológicas de la zona de estudio.

Recomendaciones

- Contrastar los resultados obtenidos mediante FRX con datos de Difracción de Rayos X (DRX) con el fin de determinar las fases mineralógicas presentes en los suelos de la zona estudiada.

Bibliografía

- Anderson, D., Sweeny, D., & Williams, T. (2008). *Estadística para administración y economía*, S.A. México: Cengage Learning Editores.
- Brouwer, P. (2003). *Theory of XRF: Getting acquainted with the principles*. Países Bajos: PANalytical B.V.
- Castillo, I., & Guijarro, M. (2006). *Estadística descriptiva y calculo de probabilidades*. Madrid, España: Pearson Educación, S.A.
- Cordoba, H., & Cortes, A. (1979). *Los Suelos de la Alta y Media Guajira, sus Características y Aptitud de Uso*. Bogotá: IGAC.
- Fernández, S. (2011). *Análisis Conglomerados*. Madrid: Universidad Autonoma de Madrid.
- IGAC. (1996). *Diccionario Geográfico de Colombia*. Bogotá: IGAC.
- IGAC. (2009). *Estudio General de Suelos y Zonificación de Tierras, Escala 1:100000, Departamento de la Guajira*. Bogotá: IGAC.
- INGEOMINAS. (2009). *Proyecto de Invetigación: Cartografía e Historia Geológica de la Alta Guajira*. Bogotá: INGEOMINAS.
- Kassambara, A. (2017). *Practical Guide to Cluster Analysis in R*. STHDA.
- Kassambara, A. (2017). *Practical Guide to Principal Component Methods in R*.
- Levin, R., & Rubin, D. (2004). *Estadística para administración y economía*. Mèxico, D.C.: Pearson Educación.
- Lilliefors, H. (1967). On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. *Journal of American Satistical Association*, 309-402.
- Martinez, D., Gil, O., & Barón, A. (2006). Espectrometría de Fluorescencia de Rayos X. *Revista Colombiana de Física*, 790-793.
- Mendes, M., & Pala, A. (2003). Type I Error Rate and Power of Three Normality Tests. *Pakistan Journal of Information and Technology* , 135-139.
- Miller, J., & Miller, J. (2002). *Estadística y Quimiometría para Química Analítica*. Madrid: Pearson Educación, S.A.

- Mohd, N., & Bee, Y. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov - Smirnov, Lilliefors and Anderson - Darling tests. *Journal of Statistical Modeling and Analytics*, 21-33.
- Monroy, S. (2008). Estadística Descriptiva. México: Instituto Politécnico Nacional.
- Newbold, P., Carlson, W., & Thorne, B. (2008). *Estadística para administración y economía*. Madrid: Pearson Prentice Hall.
- Oztuna, D., Elhan, A., & Tüccar, E. (2006). Investigation of four Different Normality Test in Terms of Type 1 Error Rate and Power under Different Distributions. *Turkish Journal of Medical Sciences*, 171-176.
- Panchuk V., Y. I. (2018). Application of chemometric methods to XRF - data - A tutorial review. *Analytica Chimica Acta*, 19-32.
- Romero, M. (2016). Pruebas de bondad de ajuste a una distribución normal. *Revista Enfermería del Trabajo*, 105-114.
- Sánchez, A., Cruz, V., Posada, M., Torrijos, T., & Osorio, A. (2016). Estudio del Análisis de Componentes principales en bases de datos de calidad del aire. *Research in Computing Science*, 9-19.
- Selvamuthu, D., & Das, D. (2018). *Introduction to Statistical Methods, Design of Experiments and Statistical Quality Control*. Nueva Deli: Springer.
- Tellez, C., & Lemus, D. (2014). Estadística descriptiva y probabilidad con aplicaciones en R. Bogotá, D.C., Colombia: Fundación Universitaria los Libertadores.