



Análisis multivariado para la caracterización de zonas agroecológicas según factores edafoclimáticos en las fincas productoras de banano del Urabá Antioqueño.

Angela Marcela Portillo Rosero

Universidad del Valle
Facultad de Ingeniería, Escuela de Estadística
Santiago de Cali, Colombia
2015

Análisis multivariado para la caracterización de zonas agroecológicas según factores edafoclimáticos en las fincas productoras de banano del Urabá Antioqueño.

Angela Marcela Portillo Rosero

Trabajo de grado presentado como requisito parcial para optar al título de:
Estadístico(a)

Director:
David Arango
Codirector:
Javier Olaya,

Universidad del Valle
Facultad de Ingeniería, Escuela de Estadística
Santiago de Cali, Colombia
2015

Con cariño y gratitud dedico esta tesis:

A Dios por mostrarme día a día que con humildad, paciencia y sabiduría todo es posible...

A mis padres Edmundo y Nancy quienes impulsan mi vida y gracias a su apoyo y comprensión, he logrado terminar mi carrera profesional, que es para mí la mejor de las herencias...

A mi querido hermano Ruben, quien con su calor humano me motivo en los momentos más difíciles.

A mi sobrina por pintar mi mundo de colores.

Y a mis tias y prima por abrirme las puertas de su casa y sobre todo las de su corazón.

Este logro no es tan solo mío, sino de cada uno de ustedes.

Agradecimientos

Sin lugar a dudas, este trabajo no habría sido posible sin la excelente labor de mis directores de tesis David Arango y el profesor Javier Olaya. Sus comentarios y sugerencias contribuyeron enormemente no sólo al desarrollo del proyecto, sino que han tenido especial importancia en mi formación como profesional.

A los profesores de la escuela de Estadística por transmitir su valioso conocimiento. Y para finalizar lo más importante: agradezco el apoyo incondicional que he recibido en todo momento de mi familia reconociendo el inmenso esfuerzo que han hecho durante toda la carrera.

Resumen

La premisa de esta investigación parte de las ventajas de adecuar la metodología de zonificación agroecológica y las técnicas estadísticas multivariantes para caracterizar extensiones de tierra con base en aspectos medio ambientales y topográficos sobre las fincas productoras de banano en el Urabá Antioqueño. Las condiciones agroecológicas se caracterizaron a partir de suelo, clima y vegetación, definiendo inicialmente grupos de lotes internamente similares respecto a variables edafoclimáticas utilizando algoritmos mixtos, que incluyen el Análisis Factorial Múltiple (AFM) como paso previo a una clasificación usando Análisis de Conglomerados. Se logra conformar un total de 5 grupos homogéneos, en los que se evidenció que la variabilidad de las características climáticas y de vegetación es poco apreciable a lo largo de la región; sin embargo las variables referentes al suelo muestran tener mayor variabilidad dentro de los grupos, dado que sus características difieren entre cada clase, por lo que son consideradas muy influyentes en la clasificación de las unidades de manejo, se encontró también que la cualidad más distintiva en toda la zona es la acidez de los suelos debido a que su pH varía aproximadamente entre 5 y 6, característica que hace del Urabá una región ideal para la producción del banano, debido además a que se trata de suelos con buena capacidad de almacenamiento de agua y nutrientes por los altos contenidos de arcilla. Finalmente se concluye que las clases son conformadas a partir de su distancia espacial debido a que las características pueden ser muy similares por su cercanía.

Palabras clave: Zonificación Agroecológica, Análisis Factorial Múltiple, Análisis de conglomerados, Caracterización y Unidad de manejo.

Abstract

The premise of this research part of the advantages of adapting the methodology of agro-ecological zoning and multivariate statistical techniques to characterize areas of land-based environmental and topographical aspects of producing banana farms in Antioquia's Urabá. Agro-ecological conditions were characterized based on soil, climate and vegetation, like Batch groups initially defining internally regarding soil and climatic variables using mixed algorithms, including the Multiple Factorial (AFM) analysis as a prelude to a classification step using Cluster Analysis. This is accomplished by forming a total of 5 homogeneous groups, in which it was shown that the variability of climate and vegetation characteristics is of little importance throughout the region; however the variables concerning the ground shown to have greater variability within groups, since their characteristics differ between classes, which are considered very influential in the classification of management units, it was also found that the most distinctive quality in The whole area is the acidity of the soil because its pH varies from about 5 to 6 feature that makes an ideal Urabá banana production region and because it is good soil water storage capacity and nutrients by high clay content. Finally we conclude that classes are formed from spatial distance because the features can be very similar because of its proximity.

Keywords: Agro-ecological zoning, Multiple Factor Analysis, Cluster Analysis, Characterization and Management Unit.

Índice general

Agradecimientos	IV
Resumen	V
1. Introducción	1
2. Planteamiento del Problema	2
3. Objetivos	4
3.1. Objetivo general	4
3.2. Objetivos específicos	4
4. Justificación	5
5. Antecedentes	7
6. Marco teórico	9
6.1. El Banano	9
6.1.1. Ecofisiología del cultivo	9
6.1.2. Producción	10
6.2. Urabá Antioqueño	10
6.3. Zonificación Agroecológica	12
6.4. Zona Agroecológica	12
6.5. Unidad de Manejo (UM)	12
6.6. Agricultura Específica por Sitio AEPS	13
6.7. Sistema de información geográfica (SIG)	13
6.7.1. Modelo de datos Raster	13
6.8. Datos Climáticos	15
6.8.1. Clima	16
6.8.2. Precipitación	16
6.8.3. Temperatura	16
6.9. Datos del Índice de Vegetación Normalizada (NDVI)	16
6.9.1. Índice de vegetación normalizada	17

6.10. Datos de Suelo	17
6.10.1. El Suelo	17
6.10.2. Conductividad hidráulica	18
6.10.3. Textura	18
6.10.4. Diámetro medio ponderado (DMP)	18
6.10.5. Nivel Freático	18
6.10.6. Humedad gravimétrica	18
6.10.7. pH	18
6.11. Análisis Factorial Múltiple (AFM)	19
6.11.1. Factores extraídos por el AFM	22
6.11.2. Representaciones del AFM y ayudas a la interpretación	22
6.12. Análisis de Conglomerados (AC)	23
6.12.1. Medidas de similitud entre objetos	23
6.12.2. Métodos de clasificación	25
6.12.3. Determinación del número de conglomerados	28
6.12.4. Combinación de métodos factoriales y de análisis de conglomerados	29
6.12.5. Caracterización de los grupos	30
7. Metodología	31
7.1. Población objeto de estudio	31
7.2. Unidad experimental	31
7.3. Aspectos éticos	31
7.4. Muestra	32
7.5. Descripción de la base de datos	32
7.5.1. Revisión y selección de variables	33
7.6. Análisis exploratorio	35
7.7. Análisis multivariado	35
7.8. Herramienta computacional	39
8. Resultados	40
8.1. Análisis descriptivo	40
8.1.1. Análisis de las variables climáticas	40
8.1.2. Análisis de las variables del índice de vegetación normalizada (NDVI)	41
8.1.3. Análisis de las variables del Suelo	41
8.2. Análisis multivariado	42
8.2.1. Análisis factorial múltiple (AFM)	42
8.2.2. Análisis de Conglomerados	45
8.2.3. Caracterización de los conglomerados y descripción de variables	47
9. Conclusiones y Recomendaciones	52

Índice general	IX
A. Anexo: Descripción de las variables	54
B. Anexo: Análisis de correlación	56
C. Anexo: Análisis descriptivos para los grupos según variables edafoclimáticas	58
Bibliografía	60

Índice de Gráficas

6-1. Mapa del Urabá Antioqueño	11
6-2. Configuración del tamaño de un modelo raster	14
6-3. Sistema de coordenadas para un modelo raster	15
6-4. Matriz X conformada por la yuxtaposición de las tablas	20
7-1. Distribución de lotes productores de Banano	32
8-1. Histograma de los valores propios	43
8-2. Representación de los grupos en los planos factoriales 1-2 y 1-3	43
8-3. Representación de las variables sobre el primer plano factorial 1 - 2	44
8-4. Representación de las variables sobre el plano factorial 1 - 3	45
8-5. Dendograma e Histograma del índice de Nivel para variables edafoclimáticas	46
8-6. Índice de Davies Bouldin y Calinski-Harabasz para variables edafoclimáticas	46
8-7. Representación de los individuos sobre el primer plano factorial y la distribución de los lotes agrupados sobre el Urabá Antioqueño	47

Índice de Tablas

7-1. Variables climáticas seleccionadas	33
7-2. Variables del NDVI seleccionadas	34
7-3. Variables del Suelo seleccionadas	34
8-1. Estadísticas descriptivas variables de clima	40
8-2. Estadísticas descriptivas para el NDVI	41
8-3. Estadísticas descriptivas del Suelo	41
8-4. Valores propios de las variables edafoclimáticas	42
8-5. Caracterización a partir de valores test: Cluster 1	48
8-6. Cluster 2	49
8-7. Cluster 3	50
8-8. Cluster 4	50
8-9. Cluster 5	51
A-1. Descripción de variables de clima 1950-2000	54
A-2. Descripción de variables del NDVI 2000-2009	55
A-3. Descripción de variables del Suelo 2000-2009	55
B-1. Matriz de correlación para el NDVI	56
B-2. Matriz de correlación para variables de Suelo	56
B-3. Matriz de correlación variables climáticas	57
C-1. Análisis descriptivo grupo edafoclimático 1	58
C-2. Análisis descriptivo grupo edafoclimático 2	58
C-3. Análisis descriptivo grupo edafoclimático 3	59
C-4. Análisis descriptivo grupo edafoclimático 4	59
C-5. Análisis descriptivo grupo edafoclimático 5	59

1. Introducción

El cultivo del banano en Colombia ha sido históricamente importante por ser generador de divisas, empleo y un componente básico para la seguridad alimentaria del país. Sin embargo los factores no controlables como el clima y el suelo son una de las fuentes de variación más importantes y fundamentales para el funcionamiento adecuado de cualquier cultivo como parte esencial de la producción agrícola, por tanto en la mayoría de los casos el productor está obligado a aceptar las condiciones del medio que se le presenten. Por tal motivo se presenta gran complejidad al caracterizar un sistema productivo como lo es el del banano, al relacionar estos factores y realizar los grupos respectivos a fin de alcanzar la homogeneidad y las características por las cuales se asocian y se pueden diferenciar, entendiendo que estos aspectos son de gran interés para cualquier agricultor, ya que todo el tiempo necesita información que le ayude a tomar decisiones sobre el manejo apropiado de su siembra.

Por consiguiente, dado el enfoque de Zonificación agroecológica, (Cock 2007), puede resultar interesante evaluarse el cultivo de banano, teniendo en cuenta las condiciones temporales y espaciales, con la información disponible sobre algunos aspectos del clima y el suelo de fincas productoras de banano que se ubican en el Urabá antioqueño, en el que se considera que el clima es de gran importancia ya que influye directamente sobre el ciclo vegetativo afectando el desarrollo fisiológico y productivo (Rodríguez & Rodríguez 2002), y el suelo desempeña un papel relevante ya que la planta depende de las características físicas y del suministro y balance de los elementos minerales esenciales requeridos para su metabolismo, crecimiento y producción, aparte de cumplir la función de soporte vital (Gildardo & Martín 2006), factores que son propicios para caracterizar en términos edafoclimáticos las perspectivas en la que se encuentran los productores, es decir condiciones ambientales parecidas en la producción del banano mediante el uso de estos factores que más inciden o afectan el desarrollo del cultivo, donde los métodos estadísticos multivariados son de gran utilidad como herramienta para la exploración de las variables y caracterización de las unidades de manejo.

De acuerdo a lo anterior el propósito de esta investigación es analizar con la metodología del Análisis Multivariante de Conglomerados la estructura natural de grupos conformado por Unidades de Manejo del cultivo de Banano, caracterizadas por diversas variables ambientales relacionadas con su productividad y funcionamiento vegetal, buscando identificar la forma en la cual se pueden establecer analogías y diferencias entre los lotes donde se siembra o se cultiva este fruto tropical.

2. Planteamiento del Problema

Desde el punto de vista histórico, la agricultura ha jugado un papel muy importante en el proceso de desarrollo de las naciones, tanto en los países desarrollados como en muchas de las economías emergentes. En este sentido, el Banco Mundial (2008) destaca que la agricultura contribuye al desarrollo general de las naciones de tres formas: como actividad económica porque puede ser fuente de crecimiento para la economía nacional, como medio de subsistencia constituyendo el medio de vida del 86 % de la población rural mundial y como proveedor de servicios ambientales ya que retiene carbono, permite la ordenación de las cuencas hidrográficas y la preservación de la biodiversidad biológica, apoyando el desarrollo sostenible de los países.

Según Bonilla & E. (2000), Colombia es esencialmente agrícola y esta actividad constituye la economía de este país cumpliendo con funciones estratégicas en su desarrollo social dado que contribuye a la producción de alimentos, al abastecimiento de materias primas, a la generación de empleo y divisas, entre otras; Esto debido a que tiene un amplio potencial productivo que equivale aproximadamente al 36.2 % del territorio, comprendido por los sistemas tradicionales, en donde a la agricultura le corresponde el 19.3 %, característica que puede ser asociada a su condición de país tropical, además, dispone de una gran variedad de pisos térmicos que van desde los nevados hasta las sabanas de la Orinoquía en el oriente y del Caribe en el norte del país (SAC 2011). Por su parte el cultivo del Banano no es distante a este hecho, ya que el País presenta una dinámica importante en la producción de este, como evidencia, es el tercer producto agrícola de exportación en importancia, con una producción promedio de 1.876 cajas por hectárea y su mayor concentración se registra en el Urabá Antioqueño, seguido por el Magdalena y la Guajira, ocupando aproximadamente 48300 hectáreas es decir el 7 % de la superficie total plantada de cultivos frutícolas (AUGURA 2012).

A pesar de este destacable desempeño, gran parte de la agricultura en Colombia se encuentra en alto riesgo, ya que el manejo de los sistemas productivos se basa en recomendaciones generales, con labores bastante similares aún en distintas regiones geográficas, sin tener en cuenta las especificidades propias del sitio, del cultivo, de la agro-empresa y la heterogeneidad ambiental. Ese manejo inadecuado conlleva a un pobre entendimiento del proceso de producción, a degradación ambiental e ineficiente uso de los recursos disponibles, trayendo consigo baja productividad, altos costos de producción y dificultades para competir (Leiva 2003).

Debido a que la producción de cultivos depende de una serie de fenómenos y variables, muchas veces

poco entendidos y que interactúan de manera compleja, además la amplia variabilidad climática y del terreno, que son un factor clave para conocer los procesos dentro de los sistemas agrícolas, los agricultores cuentan con poca información disponible lo que dificulta que tanto las experiencias productivas locales como los criterios de expertos sean aplicables a todos los sitios de producción (Leiva 2003). Por lo que es de interés recopilar información disponible que permita caracterizar los cultivos de banano a partir de condiciones climáticas y del suelo, en el que se definan grupos homogéneos que podrán describir las relaciones existentes y una variabilidad específica presente en el sitio, en base a técnicas estadísticas como el análisis factorial múltiple que servirá para extraer la información más relevante a partir de la variabilidad que caracteriza a las unidades de manejo según las variables de edafoclimáticas que se miden y de esta manera llevar a cabo la clasificación con el análisis de conglomerados a partir de las componentes o factores extraídos del análisis factorial mencionado.

3. Objetivos

3.1. Objetivo general

Caracterizar y agrupar zonas agroecológicas de las fincas del Urabá Antioqueño bajo producción bananera según factores edafoclimáticos.

3.2. Objetivos específicos

1. Realizar un análisis exploratorio para las variables edafoclimáticas de las fincas bananeras.
2. Realizar un Análisis factorial múltiple para el conjunto de variables edafoclimáticas, que permita determinar los ejes factoriales que retienen la mayor variabilidad de la información recabada de las fincas productoras de banano.
3. Realizar un análisis de conglomerados para clasificar las unidades de manejo en grupos homogéneos de clima, NDVI y suelo según características ambientales similares.
4. Describir los grupos conformados a partir de las características más significativas encontradas en cada clúster.

4. Justificación

No todas las áreas geográficas con aparentes condiciones similares son aptas para la siembra de todas las especies, éstas tienen sus propias exigencias y rangos de tolerancia respecto de los diferentes factores ambientales y socio económicos lo que hace que se limite su área de distribución a aquellas zonas que presenten ventajas comparativas y competitivas; dichas condiciones se identifican en los cultivos de banano en el territorio colombiano, debido a la posición geográfica del país, la gran biodiversidad y el amplio rango de pisos térmicos que permite un gran desarrollo agropecuario, todo esto sumado las diferentes formas de manejo de las siembras por parte de los agricultores (Ríos & Echeverry 2004).

Al disponer de un conjunto de diversas unidades de manejo asociadas a variables que miden aspectos medio ambientales y que se encuentran relacionadas con la productividad del cultivo, surge la necesidad de conformar grupos con condiciones ambientales similares, buscando reducir la variabilidad observada de toda una población en el análisis generada a partir de distintos factores como el clima y el suelo que no son controlables por el agricultor (Teran & Villaneda 1998), pero están directamente relacionados con el rendimiento que puede obtener de su cultivo, facilitando el intercambio de información y de experiencias agrícolas entre ellos mismos bajo sitios que posean condiciones climáticas o de suelo parecidas. Sin embargo, al tener numerosas variables que actúan conjuntamente en este sentido, las técnicas multivariadas son herramientas idóneas para el análisis de la información; además teniéndose en cuenta que no solo la dispersión geográfica en donde se sitúan los lotes contribuirá a la complejidad de tratar este tipo de información, también las interrelaciones de las variables ya que en ocasiones es necesario reducir la dimensionalidad de la matriz de datos; realizando en primer lugar una análisis factorial como lo menciona en su estrategia (Lebart & Piron 1995), con el fin de clasificar u agrupar a partir de las coordenadas factoriales, permitiendo tener un marco común en el proceso de formación de conglomerados.

A demás, el análisis de conglomerados que permite agrupar objetos o individuos en base a las características que poseen en común, de tal forma que cada objeto o individuo dentro de un grupo sea muy parecido a los demás objetos que hay en su grupo, pero que estos a su vez sean muy diferentes a los que se encuentran en otros; expresado en términos de variabilidad (Peña 2002). Este tipo de análisis multivariado toma sentido cuando se dispone de grandes volúmenes de información y se analiza de forma conjunta permitiendo determinar subgrupos; en este caso en particular con este análisis se busca que las clasificaciones de lotes con presencia de banano, pueden

dar lugar a un análisis descriptivo de los grupos e interpretación de las características que los definen.

Por lo tanto, el presente trabajo pretende adecuar la metodología de Zonificación Agroecológica (ZAE) de la FAO y técnicas estadísticas multivariantes para caracterizar extensiones de tierra por medio de información cuantificada de clima, suelos y otros factores físicos, que se utilizan para predecir la productividad potencial del cultivo del banano de acuerdo a sus necesidades específicas de entorno y manejo, de tal forma que esta técnica le permita a los agricultores adquirir conocimientos sobre las características ambientales a los que esta fruta se encuentra sometido.

5. Antecedentes

Son diversos los trabajos de investigación que usan métodos para la caracterización de zonas agroecológicas o zonas con características similares, haciendo uso de variables edafoclimáticas y de manejo de los cultivos que influyen en su procedimiento por lo que han recurrido al uso de diferentes métodos estadísticos con el fin de planificar y explicar técnicas que permitan a los agricultores adquirir conocimientos de cultivos. En este orden de ideas se realizó una revisión bibliográfica de algunos trabajos que sirvieron como base metodológica y temática para el presente trabajo.

Teran & Villaneda (1998) desarrollaron una metodología objetiva de zonificación agroclimática mediante el uso del Sistema de Información Geográfica SIG. Consideraron la variación espacio-temporal de los elementos climáticos y espaciales del suelo y la vegetación prevaleciente en la región de La Mojana (Colombia). La zonificación se efectuó teniendo en cuenta los excesos de precipitación derivados del balance hídrico que se producen durante el periodo de lluvias, para ello aplicaron métodos multivariados como el análisis de componentes principales y análisis cluster sobre una matriz geográfica que generaron el cual representa la variación espacial sobre estos excesos de precipitación, con el fin de zonificar y poder entender el comportamiento dinámico del agua como elemento primordial de los ecosistemas presentes en la región estudiada.

El proposito de Ríos & Echeverry (2004), en su investigación, es caracterizar y tipificar los sistemas de producción óptimos para el cultivo de lulo a fin de contribuir a la planificación y organización de la producción en el eje cafetero, para llevar a cabo el desarrollo de dicho estudio integraron técnicas estadísticas como el análisis factorial de correspondencias múltiples y el análisis de agrupamiento jerárquico basándose en las características comunes de las fincas productoras de lulo con información proporcionada por las practicas del agricultor, aspectos generales del cultivo, características físicas, socio-económicas, de mercado y lógica productiva que fue recogida en los departamentos de Caldas, Quindío y Risaralda. Como resultado identificaron cinco clases o sistemas de producción muy ligados a su ubicación geográfica.

Casa & Castrignanó (2008), realizan un análisis de las relaciones espaciales entre las variables de suelo y del cultivo en un campo de trigo en Italia, con el fin de identificar las principales causas de la variación espacial en el rendimiento del cultivo y su calidad en un ambiente mediterráneo. Para el desarrollo de la investigación tienen en cuenta variables como las propiedades del suelo, el desarrollo de la biomasa vegetal, el índice de vegetación de diferencia normalizada (NDVI) y los

componentes del rendimiento del cultivo que fueron medidas siguiendo un esquema de muestreo de cuadrícula, posteriormente los autores proponen la aplicación de una metodología geoestadística multivariante que es un análisis factorial kriging (FKA) ya que permite el procesamiento simultáneo de varias capas de información espacial y temporal de las variables del cultivo y las propiedades del suelo, atribuyendo pesos a las propiedades que afectan a la variación del rendimiento del trigo. En sus resultados logran determinar los grupos de variables que se vieron afectados por procesos que actúan a diferentes escalas espaciales, también que para las propiedades del suelo y el cultivo la variabilidad temporal es generalmente mucho mayor que la variabilidad espacial y la definición de estabilidad de las zonas potenciales de alto y bajo rendimiento es muy incierta, finalmente concluyen que las estrategias de gestión adecuadas pueden ser ajustadas a través del espacio y el tiempo con el fin de diseñar prácticas agronómicas más eficientes.

Cortes (2010), analiza la variabilidad espacial y temporal de la precipitación para la ciudad de Manizales, tomando los datos de precipitación registrados durante el año 2008 para 11 estaciones meteorológicas distribuidas sobre dicha ciudad por lo que implementa una metodología multivariada como el análisis de componentes principales (ACP) y el agrupamiento particional, con el fin de encontrar zonas de variabilidad homogénea. Los resultados obtenidos por el autor fueron confrontados con características morfológicas y climáticas de cercanía entre estaciones, de manera que hubiera concordancia entre los resultados arrojados y las características físicas, logrando finalmente identificar las zonas de variabilidad homogénea en cuanto a magnitud e intensidad de las lluvias.

Recientemente, (Dorado 2013), realiza una investigación a partir de información proporcionada por agricultores y registros meteorológicos de los cultivos de plátano ubicados en once departamentos de Colombia, con el objetivo de caracterizar zonas productoras en cuanto a condiciones climáticas y de suelo en base a métodos estadísticos multivariados; debido a la naturaleza de los datos emplea un análisis de componentes principales para cada grupo de variables (cuantitativa y cualitativas), en el que encontró que para el caso de las variables climáticas los tres primeros factores logran explicar el 86.83 % de la inercia total, mientras que para el suelo cuenta con un conjunto de datos que contienen escalas mixtas por lo que realiza un análisis de componentes principales para variables categóricas, como resultado obtiene un modelo factorial en el que sus tres primeros ejes contienen el 71,69 % de la variabilidad total; posteriormente realiza una agrupación de las parcelas con la aplicación del método de clasificación jerárquica de Ward combinado con el algoritmo de k medias, de lo que obtiene una conformación de 10 grupos climáticos y 5 grupos de suelos; finalmente describe las variaciones en rendimiento y manejo dentro de cada grupo agroecológico, distinguiendo que algunas muestran ser más compatibles con altos rendimientos pero que además todas contienen producciones poco exitosas y muy exitosas de acuerdo a la distribución general del grupo, también logra evidenciar que en las variables de manejo la zona puede influir tal como la elección de la variedad o la decisión de tener el plátano como mono cultivo.

6. Marco teórico

En este capítulo se presentan los conceptos de los aspectos más destacados del cultivo del banano en Colombia, además las técnicas estadísticas multivariadas como el Análisis de componentes principales y el Análisis de conglomerados que se utilizaran para llevar a cabo el desarrollo del estudio.

6.1. El Banano

Su nombre científico es *Musa Sapientum* son plantas herbáceas rizomatosas que tienen una altura entre 3 y más de 7 m. Son de origen Asiático, no toleran las heladas; tanto las especies silvestres como las cultivadas muestran un amplio rango de adaptación al suministro de agua y temperatura. Cada retoño de banano produce una serie de hojas, hasta que en la parte superior nace directamente una inflorescencia que emerge del centro de los pseudo tallos en posición vertical semejando un enorme capullo púrpura o violáceo, para iniciar el llenado de la fruta. Este proceso es guiado internamente y no es influenciado directamente por el fotoperiodo o la temperatura. Cada planta produce una serie de racimos, uno de cada rebrote lateral. Luego que la fruta alcanza la madurez, es cosechada y el pseudo tallo muere, todo o parte de este es cortado y un nuevo racimo emergerá de la siguiente generación proveniente del rebrote previamente seleccionado. Brotes laterales producen puyones (rebrotos), que juntos con los padres forman la planta. Usualmente tres generaciones son visibles (Madre, Hija y Nieta), este proceso puede repetirse por 50 generaciones o más; el elemento perenne en una planta de banano es el rizoma superficial o subterráneo, que posee meristemas a partir de los cuales nacen entre 200 y 500 raíces fibrosas, que pueden alcanzar una profundidad de 1.5 m y cubrir 5 m de superficie. Cada rebrote produce 40 ± 10 hojas antes de producir partes reproductivas (Guarín 2011).

6.1.1. Ecofisiología del cultivo

El cultivo se realiza con éxito en diferentes ambientes, tanto semi-áridos como subtropicales. En los trópicos el cultivo se ve más favorecido debido a los regímenes de temperatura óptimos, de alrededor de 24 °C, y a la abundancia y distribución uniforme de las lluvias. Una plantación de banano requiere una temperatura promedio de 25 °C con una precipitación anual entre 2000 y 5000 mm y una altitud hasta 300 msnm. Se requieren de suelos profundos, bien drenados, que retengan humedad, siendo mejores los de textura franca arcillosa a franco arenoso, arcillas livianas y drenadas

con un pH de entre 6.0 y 7.5. Entre los factores que más influyen en el desarrollo del cultivo están la temperatura, el nivel nutricional del suelo, humedad y duración del día. La producción de banano está directamente relacionada con el peso del racimo y con el número de plantas por unidad de área; e inversamente relacionada con la longitud del tiempo requerido para la formación de los frutos. El tamaño del racimo está relacionado al número de manos, número de dedos o bananas por mano y por el tamaño de cada fruta. El tamaño de las plantas y el peso de los racimos se ha relacionado al número y tamaño de las hojas funcionales. Las mayores producciones se producen cuando a la floración hay de 10 a 12 hojas funcionales con un adecuado suministro de nitrógeno. El peso máximo de los racimos se alcanza antes del invierno donde los días son más cortos (Espinal & Peña 2005).

6.1.2. Producción

En el país existen dos tipos de banano: banano de exportación y banano criollo o de consumo interno. Colombia ha tenido una larga tradición como productora y exportadora de banano tipo Cavendish Valery; en relación al banano criollo (común y murrapo) o de consumo interno, se produce principalmente en el Valle del Cauca, Tolima y Antioquia y tiene un área cosechada y una producción significativamente menores al de exportación. Para el año 2003 la producción de este banano alcanzó 136644 Tn, lo que representó un 5.0 % del total de frutas producidas y el 7.5 % (13964 ha) del total de hectáreas de frutas del país. Como proporción del renglón bananero total, el banano común participó con el 8.9 % de la producción y el 24.9 % de la superficie total del cultivo. En relación al mercado de exportación, la Cadena de banano ha logrado mantener su posición como exportadora neta, después del café y las flores, con valores de exportación que han oscilado entre 400 y 444 millones de dólares desde 1991. En el 2003 ocupó el 1.06 % (42000 Ha) del total del área agrícola sembrada en el país y alcanzó el 5.7 % (1399623 Tn) de la producción agrícola total (Espinal & Peña 2005).

6.2. Urabá Antioqueño

La región del Urabá está localizada en la parte noroccidental del departamento de Antioquia, limita al norte y noroeste con el Mar caribe (Océano Atlántico) donde se encuentra el Golfo de Urabá, al oriente con el departamento de Córdoba, al sur con el departamento del Chocó y al occidente con el departamento del Chocó. Tiene una extensión de $11,664 \text{ Km}^2$ (18.6 % del total del departamento), con 507.794 ha cubiertas de bosques (IGAC 2007); posee alturas sobre el nivel del mar que oscilan entre 0 y 3200 m (máxima en el Nudo Paramillo, Mutatá), pero predominan las alturas entre 0 y 200 m, con 70 % de su territorio en planicies y terrenos ondulados.

En su vasta geografía comparte ecosistemas estratégicos con los departamentos de Córdoba y Chocó, y hace parte a la vez de las regiones Andina, Caribe y Pacífico, tiene áreas costeras y marinas. La confluencia de todos estos elementos le confieren al territorio de Urabá unas características únicas

que favorecen su riqueza, biodiversidad y pluriculturalidad.

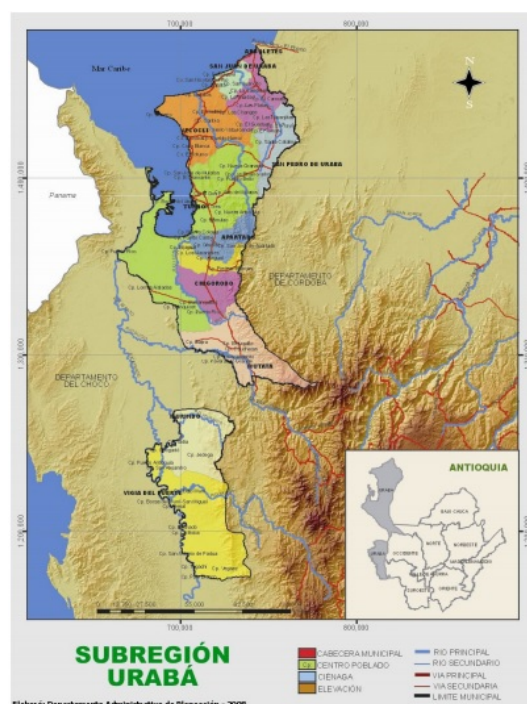


Figura 6-1.: Mapa del Urabá Antioqueño

En su organización político administrativa, comprende once municipios, los cuales, dadas las diferentes dinámicas económicas, ambientales, culturales y niveles de articulación, les definen características especiales, agrupándose en tres zonas claramente determinadas:

- Zona Norte: Arboletes, Necoclí, San Juan de Urabá y San Pedro de Urabá
- Zona Centro: Apartadó, Carepa, Chigorodó, Mutatá y Turbo
- Zona Atrato Medio: Murindó y Vigía del Fuerte

La economía de la zona se basa en :

- La ganadería extensiva, que se desarrolla en grandes latifundios, en las áreas planas y de poca pendiente de los valles de los ríos San Juan y Mulatos y sobre la planicie costera en Arboletes.
- Una economía campesina en pequeñas parcelas desarrollada sobre la planicie costera de San Juan de Urabá y sobre las partes altas de las cuencas de los ríos San Juan y Mulatos (coco, banano y maíz), hasta donde han sido desplazados los campesinos por la expansión del latifundio.

- El desarrollo incipiente de la pesca y el turismo de mar especialmente en los municipios de Arboletes y Necoclí.

La zona centro es la que tiene mayor dinámica económica y demanda de productos generada por la concentración de población. El eje está marcado por la agricultura comercial (actividad bananera), hasta el punto que ha llegado a convertirse en el símbolo de la región (Ruiz & Suaza 2011).

6.3. Zonificación Agroecológica

Se refiere a la división de la superficie de tierra en unidades más pequeñas, que tienen características similares relacionadas con la aptitud de tierras, la producción potencial y el impacto ambiental. Esta Metodología (ZAE) permite hacer un mejor uso del suelo, identificando las zonas más apropiadas para el desarrollo de las actividades agropecuarias ya que estas se ven influenciadas por interacciones ecológicas tales como altitud, tipos de suelos, régimen de lluvias, temperatura, etc (FAO 2005).

6.4. Zona Agroecológica

Es una unidad cartográfica de recursos de tierras, definida en términos de clima, fisiografía y suelos que tiene un rango específico de limitaciones y potencialidades para el uso de tierras (FAO 2005).

Otros autores como Carbonell & Isaacs (2001) definen las zonas agroecológicas como sitios que poseen combinaciones de clima y características de suelos similares además de tener la misma capacidad biofísica para la producción agrícola a largo plazo con factores generalmente estables; Los parámetros particulares para definirlos normalmente son asignados de acuerdo a los requerimientos climáticos y edáficos de los cultivos, también teniendo en cuenta los sistemas de manejo en que se desarrollan, su caracterización es útil como punto de referencia de las recomendaciones diseñadas para mejorar la situación existente en el uso de la tierra, incrementando la producción.

6.5. Unidad de Manejo (UM)

Se define como un lote o área con características homogéneas y además un manejo uniforme dado por el agricultor. Existen debido a razones naturales como las diferencias de suelos, climas, topografías, cualidades fenológicas del cultivo (edad y variedad) o debido a la intervención del agricultor que puede implementar diversas prácticas (Cock & Estrada 2006).

En una misma finca puede existir más de una unidad de manejo de acuerdo a la variabilidad que presenten sus lotes, un solo cultivo puede diferenciarse en términos de variedad, parámetros del suelo, edades de las plantas y prácticas agrícolas, cada una de ellas son constituidas individualmente por tener características homogéneas las cuales se les asocia una sola respuesta de producción.

6.6. Agricultura Específica por Sitio AEPS

Según la definición de Cenicaña (Centro de Investigación de la Caña de Azúcar de Colombia), la agricultura específica por sitio (AEPS) es el arte de realizar las prácticas agronómicas requeridas por una especie vegetal de acuerdo con las condiciones espaciales y temporales del sitio donde se cultiva, para obtener de ella su máximo rendimiento potencial.

Particularmente consiste en una tecnología que se basa en obtener datos georreferenciados de los lotes para lograr un mayor conocimiento de los factores que contribuyen a la variabilidad que se presenta a nivel de lote y así delimitar Zonas de Manejo (ZM) que han sido definidas como subregiones dentro de los lotes que expresan una combinación homogénea de factores determinantes en la producción de cultivos (Anselin & Lowengberg-DeBoer 2004).

6.7. Sistema de información geográfica (SIG)

Entendido como procedimiento de análisis espacial, consiste en un proceso de obtención, almacenamiento y análisis de datos que permite generar información útil para la toma de decisiones. La principal característica de este sistema de información en particular, es que está diseñado para trabajar con datos referenciados con respecto a coordenadas espaciales o geográficas. Un dato georeferenciado o dato espacial se puede definir como un elemento ubicado en el espacio mediante un sistema de coordenadas, el cual puede ser descrito a través de una serie de atributos y que además permite conocer su relación con respecto a otros elementos. Un conjunto de datos espaciales asociado a una localización específica, y que sea de utilidad para la toma de decisiones, se puede denominar información geográfica. El SIG posee una base de datos integrada en que los aspectos geométrico, topológico y descriptivos, deben existir como uno solo. Esto significa que si se altera la parte geométrica automáticamente se altera la base descriptiva. Esta integración entre los datos geométricos y descriptivos que permiten un complejo análisis espacial diferencia a los SIG de los programas de diseño, ingeniería, arquitectura y también cartográficos (Moreira 1996). Se distinguen tres modelos de Sistemas de Información Geográfica: vectoriales, raster y datos orientados a objetos. Se describe a continuación solamente el modelo que es de interés, debido a que los datos obtenidos están almacenados en ese formato.

6.7.1. Modelo de datos Raster

Se fundamenta en la división del área de estudio en una matriz de celdas (o píxeles) organizadas en filas y columnas, generalmente cuadradas. Cada una de estas celdas recibe un único valor que se considera representativo para toda la superficie abarcada por la misma, el cual puede contener información asociada a la temperatura, precipitación, suelo, vegetación, entre otras. Los rásteres son fotografías aéreas digitales, imágenes de satélite, imágenes digitales o incluso mapas escaneados.

La matriz de datos, que puede contener tres tipos de datos:

- Valores numéricos en caso de que la variable representada sea cuantitativa.
- Identificadores numéricos en caso de que se trate de una variable cualitativa. Estos identificadores se corresponden con etiquetas de texto que describen los diferentes valores de la variable cualitativa.
- Identificadores numéricos únicos para cada una de las entidades representadas en caso de que la capa raster contenga entidades (puntos, líneas o polígonos).

El tamaño de celda determina el grosor o la delgadez con la que aparecerán los patrones o las entidades en el ráster. Cuanto más pequeño sea el tamaño de celda, más suave o más detallado será el ráster. Sin embargo, cuanto mayor sea el número de celdas, más tiempo tardará en procesar, aumentándose a su vez la demanda de espacio de almacenamiento. Si el tamaño de una celda es demasiado grande, se podría perder información o los patrones sutiles podrían oscurecerse (ver figura 6-2).

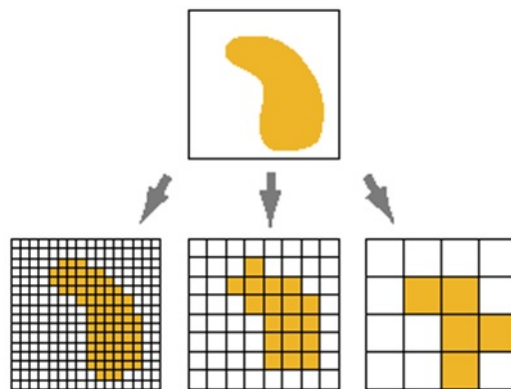


Figura 6-2.: Configuración del tamaño de un modelo raster

La ubicación de cada celda se define por la fila o columna en la que se ubica dentro de la matriz ráster, esencialmente, esta queda representada por un sistema de coordenadas cartesianas, en las que las filas de la matriz son paralelas al eje x y las columnas al eje y del plano cartesiano. Los valores de fila y columna parten de 0 (ver figura 6-3).

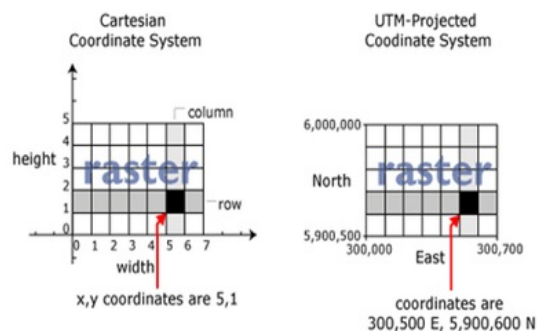


Figura 6-3.: Sistema de coordenadas para un modelo raster

6.8. Datos Climáticos

Las estaciones meteorológicas en Colombia a menudo no están tan cerca de los campos donde la mayoría de las especies de frutales tropicales se cultivan, por el cual la información climática proporcionada por estas estaciones rara vez representan el clima de los centros de producción individual en gran parte; debido a la gran variación de altitud de la Región. Por lo tanto, la generación de las condiciones climáticas, paisaje topográfico en cada estación se obtiene a partir de las coordenadas (latitud y longitud) (Jiménez & Barreto 2011). Con esta información espacial, es posible extraer los datos actuales del clima obtenidos de la base de datos WorldClim, desarrollado por Hijmans & Jarvis (2005), disponible en <http://www.worldclim.org>, que se presenta en la forma de archivos raster y fue desarrollada a partir de datos de estaciones meteorológicas de diferentes fuentes y es hasta la fecha la de mayor resolución existente, a 30 arco-segundos de resolución espacial (aproximadamente 1 km en el Ecuador) y con cobertura global para el análisis climático. ***Son promedios climáticos de largo plazo que contienen datos de precipitación mensual total, temperaturas mensuales mínimas, máximas y promedios del periodo comprendido entre 1950 y 2000 y por tanto son útiles para representar el clima de grandes y pequeñas regiones en diferentes áreas geográficas.*** La Precipitación anual total se calcula como la suma de todos los valores mensuales y los promedios anuales de temperatura como los promedios de los valores mensuales. A partir de las variables mensuales, se derivan una serie de índices bioclimáticos que en su mayoría, se encuentran altamente relacionados con el crecimiento, desarrollo fisiológico y biológico de las especies de plantas tanto cultivadas como silvestres. Para tal efecto, calcularon 19 variables que representan tendencias anuales y trimestrales, estacionalidades y factores limitantes o extremos, como se muestra en el **ANEXO A**, tabla **A-1**).

6.8.1. Clima

Es el conjunto fluctuante de las condiciones atmosféricas, caracterizado por los estados y evoluciones del tiempo, durante un periodo espacio-temporal de una región dada, y controlado por la interacción entre los diferentes componentes del sistema climático. Debido a que el clima se relaciona generalmente con las condiciones predominantes en la atmósfera, esto se describe a partir de variables atmosféricas como la temperatura y la precipitación (Pabón J. & J. 2002).

6.8.2. Precipitación

Es la cantidad de agua que cae a la superficie terrestre y proviene de la humedad atmosférica, ya sea en estado líquido (llovizna y lluvia) o en estado sólido (escarcha, nieve o granizo). La precipitación es uno de los procesos meteorológicos más importantes para la hidrología, y junto a la evaporación constituyen la forma mediante la cual la atmósfera interactúa con el agua superficial en el ciclo hidrológico del agua.

6.8.3. Temperatura

Según la OMM (1996), la temperatura es la condición que determina la dirección del flujo neto de calor entre dos cuerpos y permite expresar el grado de calentamiento o enfriamiento de los cuerpos. Es una de las magnitudes más utilizadas para describir el estado de la atmósfera, por información meteorológica se sabe que la temperatura del aire varía entre la noche y el día, entre una estación y otra, entre las ubicaciones geográficas, la cual se mide a través de un instrumento denominado termómetro. Además tiene asociadas diferentes escalas de medida, grados Celcius, Fahrenheit y Kelvin.

6.9. Datos del Índice de Vegetación Normalizada (NDVI)

Los datos de NDVI se extrajeron a partir de imágenes del satélite TERRA MODIS, producto MOD13Q1. La resolución espacial de estos es de 7.5 arco-segundos (aproximadamente 250 metros en el Ecuador), y su frecuencia es de 16 días, esto significa que cada 16 días hay un nuevo dato de NDVI resultado de la medición del satélite, y que es publicado por la NASA, descargado y post-procesado por el CIAT. El período para el cual se extrajeron los datos de NDVI fue del 2000 al 2009. Dado que el objetivo de tener datos de NDVI para un análisis de zonificación es lograr capturar patrones temporales en los datos espaciales, se usaron por tanto las variables que representan tendencias centrales y de dispersión en los datos (ver **ANEXO A**, Tabla **A-2**).

6.9.1. Índice de vegetación normalizada

Un Índice de Vegetación puede ser definido como un parámetro calculado a partir de los valores de la reflectancia a distintas longitudes de onda, y es particularmente sensible a la cubierta vegetal. Estos índices no miden directamente productividad ni disponibilidad forrajera, pero tienen una estrecha relación con estas variables, lo que permite utilizarlos para análisis regionales. Es decir, que los cambios del índice de vegetación permiten decir si un determinado mes o año es mejor o peor respecto al valor histórico. En otras palabras este índice permite identificar la presencia de vegetación verde en la superficie y caracterizar su distribución espacial así como la evolución de su estado a lo largo del tiempo. Esto está determinado fundamentalmente por las condiciones climáticas. El NDVI es un índice no dimensional, y por lo tanto sus valores van de 0 a 1. En la práctica los valores que están por debajo de 0,1 corresponden a los cuerpos de agua y a la tierra desnuda, mientras que los valores más altos son indicadores de la actividad fotosintética de las zonas de matorral, el bosque templado, la selva y la actividad agrícola (Meneses 2011).

6.10. Datos de Suelo

El Centro Internacional de Agricultura Tropical (CIAT) y la corporación Biotec desarrollaron una metodología para caracterizar el suelo y el terreno de las fincas o lotes de productores, llamada RASTA (Evaluación Rápida de Suelo y Terreno), permitiendo que los agricultores con materiales de campo puedan evaluar e identificar su suelo de una forma simple con el fin de conocer sus recursos y tomar las mejores decisiones para el manejo de los mismos sin recurrir a complicados y costosos sistemas de caracterización y clasificación considerando que la mayoría de los agricultores no cuentan ni con el conocimiento profundo ni con los recursos para evaluar su suelo.

Este método se divide en tres partes: En la primera, se puede medir características del suelo (forma del terreno, pendiente, color, textura, estructura, potencial de Hidrógeno (pH), pedregosidad, capas endurecidas, moteados, resistencia al rompimiento y presencia de carbonatos); la segunda parte consta de una serie de observaciones de campo; que permitirá identificar, en la tercera parte, cuatro propiedades del suelo, como son: la materia orgánica, el drenaje, la profundidad efectiva y la presencia de sales proporcionando la medición de factores relacionados con las propiedades del suelo tanto físicas como químicas (ver **ANEXO A**, Tabla **A-3**), el cual tienen un efecto importante sobre el crecimiento de la planta (Cock & Estrada 2006).

6.10.1. El Suelo

Desde el punto de vista agrícola, el suelo es la capa de material fértil que recubre la superficie de la Tierra que sirve de medio natural para el crecimiento de las plantas y a partir de la cual obtienen sostén, nutrimentos y agua. Desde una perspectiva ambiental, existen varias definiciones

que incorporan su papel fundamental en los procesos ecosistémicos, debido a las funciones y servicios que realiza tales como la regulación y la distribución del flujo de agua o como amortiguador de los efectos de diversos contaminantes.

6.10.2. Conductividad hidráulica

Es un parámetro clave en la descripción de los procesos de infiltración y redistribución de agua en el suelo. Esta puede ser determinada en el laboratorio y en el campo. La conductividad hidráulica es una propiedad muy importante de los medios porosos, que indica la movilidad del agua dentro del suelo y depende del grado de saturación y la naturaleza del mismo (Gabriels & Pinto 2011).

6.10.3. Textura

En el suelo existen diversas partículas, entre las cuales las más importantes difieren por su tamaño y se clasifican como **arcillas** que son suelos de textura fina o pesada, los suelos **limosos** que son de textura media y los suelos **arenosos** que son de textura gruesa o liviana; la textura es la proporción que hay de cada una de ellas en el suelo, y se expresa en porcentaje (%). Influye en procesos de retención y almacenamiento de agua y oxígeno, en la fertilidad, la porosidad y el drenaje, entre otros (Cock & Estrada 2006).

6.10.4. Diámetro medio ponderado (DMP)

Cuando las partículas individuales se agrupan, toman el aspecto de partículas mayores y se denominan agregados, esta medida (DMP) expresa la estabilidad y tamaño de los agregados del suelo.

6.10.5. Nivel Freático

Altura que alcanza la capa acuífera subterránea más superficial, es decir que es la acumulación de agua subterránea que se encuentra a una profundidad relativa bajo el nivel del suelo en donde las aguas subterráneas tienen una presión igual a la atmosférica.

6.10.6. Humedad gravimétrica

La relación de la masa de agua contenida en los poros de los suelos o rocas respecto a la masa sólida de las partículas en ese material, expresada como un porcentaje.

6.10.7. pH

Es la medida de la acidez (1-5), neutralidad (5-7) o alcalinidad (más de 7) del suelo e influye en sus propiedades físicas, químicas y biológicas, por lo cual puede limitar o favorecer el crecimiento

de ciertos cultivos (Cock & Estrada 2006).

6.11. Análisis Factorial Múltiple (AFM)

El AFM desarrollado por Escofier & Page (1992), permite el tratamiento simultáneo de tablas de datos en las que un mismo conjunto de individuos se describe a través de varios grupos de variables de distinta naturaleza (cualitativa o cuantitativa). Las diferentes tablas equilibran su influencia que sobre el análisis global pueda ejercer cada una de ellas. El procedimiento consiste en comparar las nubes de individuos a través de las matrices de productos escalares (matriz de covarianzas o de correlación, según la transformación más apropiada de los datos), equilibrando su influencia.

Notación: En la aplicación de métodos multivariados de análisis de datos, es usual contar con individuos descritos por variables cuantitativas y cualitativas, estas últimas medidas en escala nominal u ordinal. La matriz de datos, denotada por Y , se considera estructurada en tres subtablas: Y_c para las variables cuantitativas, Y_q para las variables nominales y Y_o para las variables ordinales; la estructura de la matriz Y es (Bécue 2008):

$$Y = [Y_c, Y_q, Y_o]$$

Si se requiere considerar grupos de variables en cada una de las subtablas, se induce una estructura de grupos. Es decir: $Y_c = [Y_{c1}, \dots, Y_{cJ_c}]$ conjunto de variables cuantitativas; $Y_q = [Y_{q1}, \dots, Y_{qJ_q}]$ conjunto de variables nominales y $Y_o = [Y_{o1}, \dots, Y_{oJ_o}]$ conjunto de variables ordinales.

A partir de la matriz Y se construye la matriz X , cuya estructura es:

$$X = [X_c, Z_q, Z_o]$$

Donde

- X_c es ahora la matriz de datos transformados, es decir contiene las estandarizaciones de las variables cuantitativas de la matriz Y_c de la forma, para $j = 1, \dots, J_c$.

$$X_j = \frac{(Y_j - \bar{Y}_j)}{\sigma_{Y_j}}$$

- Z_q y Z_o son matrices en la forma de una tabla disyuntiva completa (TDC), la cual representa la codificación de variables cualitativas, en este caso variables nominales u ordinales. El termino general de esta matriz indicadora es $z_{ikj} = 1$ en presencia de la categoría y 0 en ausencia, los términos i, k, j indican:

- i:** un individuo o unidad estadística, en general $i = 1, \dots, I$
- k:** una variable cuantitativa o categórica, en general $k = 1, \dots, K$
- j:** un conjunto de variables, en general $j = 1, \dots, J$

Finalmente la matriz X, toma la siguiente estructura:

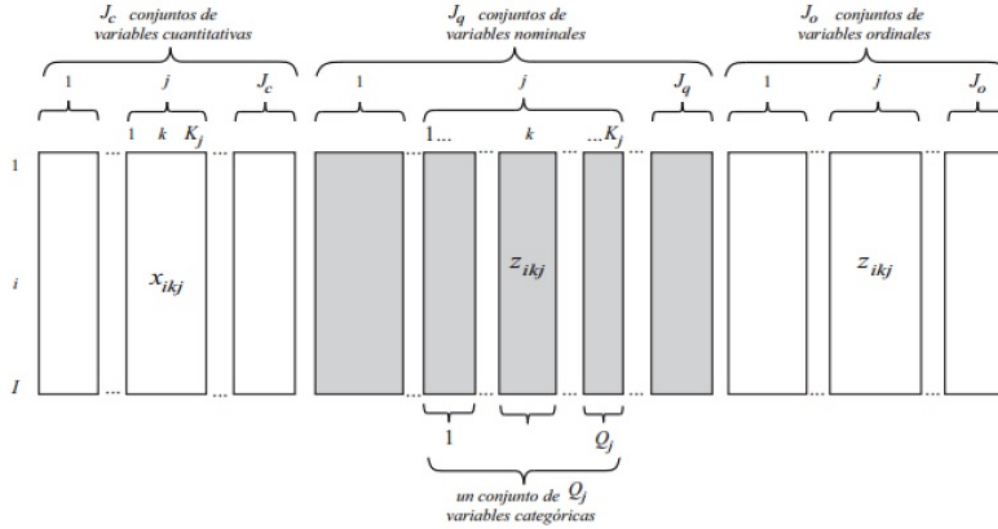


Figura 6-4.: Matriz X conformada por la yuxtaposición de las tablas

Donde, I son los individuos o unidades estadísticas, J_c número y conjuntos de variables cuantitativas, J_q número y conjuntos de variables nominales, J_o número y conjuntos de variables ordinales. Además, K_j es el número total de categorías en un conjunto j , con Q_j variables en el conjunto. Donde $I \times K_j$ es la tabla asociada a un conjunto j y $K = \sum_{j \in J} K_j$.

Para los J_c grupos o conjuntos de variables cuantitativas, los K_j son tanto el número de columnas como el número de variables. Para un grupo j con Q_j variables cualitativas (nominales u ordinales) el número total de columnas K_j representa el número total de categorías (modalidades) asumidas a través de las Q_j variables.

Ponderación de los grupos de variables

El método se desarrolla en dos etapas: inicialmente se realiza un análisis de componentes principales (ACP) por separado para cada subtabla, reteniendo en cada uno de los análisis, el primer valor propio, denotado por λ_1^j .

El papel que puede desempeñar un grupo de variables en un análisis de conjunto interviene dos elementos (Escofier & Page 1992):

1. El número de variables del grupo (cuanto más importante es el número, mayor es la influencia del grupo).
2. La estructura del grupo; cuanto más fuerte es la estructura del grupo, es decir cuanto más relacionadas están las variables su influencia será más determinante en la construcción de los principales factores de variabilidad.

En la segunda etapa se realiza un ACP para la tabla completa (X, M, D) en la cual las variables evaluadas sobre una misma condición se ponderan por el inverso del primer valor propio $\frac{1}{\lambda_1^2}$ extraído del ACP de cada subtabla (condición) y se estructuran así:

- X , es la matriz de datos transformados:

$$X = (X_c, Z_q, Z_o)$$

- M , es la matriz de métrica:

$$M = \begin{pmatrix} M^{J_c} & 0 & 0 \\ 0 & M^{J_q} & 0 \\ 0 & 0 & M^{J_o} \end{pmatrix}$$

Donde M^{J_c} es la matriz de métrica para los conjuntos de variables cuantitativas conformada por J_c bloques, la matriz M^{J_c} tiene la siguiente forma:

$$M^{J_c} = \begin{pmatrix} M_1 & 0 & \dots & 0 \\ 0 & \ddots & & 0 \\ \vdots & & M_j & \vdots \\ 0 & 0 & \dots & M_{j_c} \end{pmatrix}$$

Cada matriz $M_j = \frac{1}{\lambda_1^2} I_{K_j}$, donde I_{K_j} , es la matriz idéntica de orden $K_j \times K_j$.

M_{J_q} es la matriz de métrica para los conjuntos de variables nominales definida por J_q bloques, la matriz M_{J_q} tiene la forma:

$$M^{J_q} = \begin{pmatrix} M_1 & 0 & \dots & 0 \\ 0 & \ddots & & 0 \\ \vdots & & M_j & \vdots \\ 0 & 0 & \dots & M_{j_q} \end{pmatrix}$$

Cada matriz es de la forma $M_J = \frac{1}{\lambda_1^j} \text{diag}(\frac{I_K}{I_{Q_j}})$, donde I_K es el número de individuos que seleccionan la categoría k , I es el número de individuos y Q_j es el número de variables cualitativas (nominales) en el grupo j . La definición de la matriz de métrica M_{J_o} , para los conjuntos de variables cualitativas (ordinales) definida por J_o bloques, es similar a la definida para las variables nominales.

- D es la matriz de pesos, donde I_I es la matriz idéntica de tamaño $I \times I$, (Escofier & Page 1992).

$$\frac{1}{I} I_I$$

6.11.1. Factores extraídos por el AFM

Del análisis global expresado como un ACP ponderado, la matriz métrica M , la matriz de pesos D y las coordenadas del individuo i sobre el primer eje factorial, se obtienen los valores de la primera componente principal para individuos, el cual se definen como:

$$F_s = XM u_s$$

Para variables se define como:

$$G_s = X' D v_s$$

donde M es la matriz diagonal de pesos de variables, D es la matriz diagonal de pesos de individuos, u_s es el vector propio obtenido por diagonalización de la matriz $XM X' D$ y v_s es el vector propio obtenido por diagonalización de la matriz $X' D X M$.

6.11.2. Representaciones del AFM y ayudas a la interpretación

Según Escofier & Page (1992) el método AFM presenta:

- Los resultados clásicos de un ACP: la representación de individuos, variables y las ayudas a la interpretación como calidad de la representación y las contribuciones de individuos y variables a la formación de un eje.

- Los resultados específicos de la estructura tres vías. Entre estos se destacan: la representación de las condiciones (grupos de variables) como puntos, la importancia de los factores comunes en cada grupo, la representación simultánea de las nubes de individuos y las ayudas a la interpretación para evaluar la calidad de las representaciones desde un punto de vista global.

Las representaciones se generan en tres espacios: R^J, R^I, R^{I_2} , en los cuales se localiza la nube de individuos, variables y condiciones respectivamente. Entre otras ayudas a la interpretación el AFM permite extraer factores comunes al conjunto de las condiciones, a algunas condiciones y los factores específicos de una condición.

6.12. Análisis de Conglomerados (AC)

El análisis de conglomerados tiene por objeto agrupar observaciones o variables con características lo más parecidas posibles al interior de cada grupo y lograr la mayor diferencia entre ellos, es decir que particiona los elementos en grupos homogéneos en función de las similitudes y este análisis toma sentido cuando se dispone de un conjunto de datos que se sospecha contienen alta variabilidad. El análisis de conglomerados estudia tres tipos de problemas (Peña 2002):

1. **Partición de los datos:** se dispone de un conjunto de datos que se sospechan son heterogéneos y se desea dividirlos en un número de grupos, de tal manera que: cada elemento pertenezca a uno y solo uno de los grupos, todo elemento quede clasificado y cada grupo sea internamente homogéneo.
2. **Construcción de jerarquías:** se desea estructurar los elementos de un conjunto de forma jerárquica por su similitud, esto implica que los datos se ordenan en niveles, de manera que los niveles superiores contienen a los inferiores. La jerarquía construida permite obtener también una partición de los datos en grupos.
3. **Clasificación de variables:** cuando se dispone de muchas variables, es interesante en ocasiones hacer un estudio exploratorio para dividir las variables en grupos lo que puede orientar a plantear los modelos formales para reducir la dimensión de los datos.

6.12.1. Medidas de similitud entre objetos

El objetivo del análisis de conglomerados es agrupar objetos similares, por lo que se establece un indicador para evaluar las distancias y similitudes (di-similaridades) que son medidas de asociación que reflejan su lejanía o su proximidad entre los objetos (Díaz 2007). La estrategia más común consiste en medir la equivalencia en términos de la distancia entre los pares de objetos. Los objetos con distancias reducidas entre ellos son más parecidos entre sí que aquellos que tienen distancias mayores y se agruparán, por tanto, dentro del mismo clúster (Peña 2002).

Distancia Euclidiana: es la raíz cuadrada de las diferencias de las componentes elevada al cuadrado. En general si tenemos p variables $X_1, X_2, \dots, X_p$ medidas en una muestra de n individuos, la observación para el sujeto i puede ser escrita como $x_{i1}, x_{i2}, \dots, x_{ip}$ y la observación de los datos para el sujeto j $x_{j1}, x_{j2}, \dots, x_{jp}$ y se denota por:

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$$

Distancia Mahalanobis: incorpora directamente el procedimiento de estandarización. Es apropiada cuando se disponen de variables intercorrelacionadas pues ajusta equitativamente variables por intercorrelaciones y pesos. X_i es el vector de observaciones de la variable k -ésima, X_j el vector de mediciones de la variable j -ésima y se denota:

$$d_{ij} = (X_i - X_j) \Sigma^{-1} (X_i - X_j)$$

donde Σ es la matriz de varianzas y covarianzas de las variables consideradas y X_i y X_j , son matrices fila ($1 \times p$)

Distancia Manhattan: también llamado City Block, usa diferencias absolutas en vez de diferencias al cuadrado. Asume que las variables no están correlacionadas, de estarlo los grupos no serian validos.

$$d_{ij} = \sum_{p=1}^n |(x_{ip} - x_{jp})|$$

Distancia de Gower: esta distancia es aplicable cuando se dispone de un conjunto de datos mixtos, en los que a partir de los individuos se han observado tanto variables cualitativas como cuantitativas y se define como $d_{ij}^2 = 1 - S_{ij}$, donde

$$S_{ij} = \frac{\sum_{k=1}^p \delta_{ijk} s_{ijk}}{\sum_{k=1}^p \delta_{ijk}}$$

S_{ij} es el coeficiente de similaridad de Gower entre el individuo i y el individuo j ; s_{ijk} es una medida parcial de similitud entre los individuos i y j al ser comparados con base a la variable k -ésima y δ_{ijk} es un ponderador o peso de la comparación entre los individuos i -ésimo y j -ésimo en la k -ésima variable.

- Para la variables binarias de presencia-ausencia, $s_{ijk} = 1$, si $x_{ik} = x_{jk}$ y $s_{ijk} = 0$ para divergencias, cuando x_{ik} no es igual a x_{jk} . Además $\delta_{ijk} = 0$ para dobles ausencias y $\delta_{ijk} = 1$ en los demás casos.
- Para las variables nominales $s_{ijk} = 0$ si x_{ik} no es igual a x_{jk} y $s_{ijk} = 1$ si $x_{ik} = x_{jk}$
- Si la variable es cuantitativa entonces:

$$s_{ijk} = 1 - \frac{x_{ik} - x_{jk}}{R_k}$$

Donde x_{ik} y x_{jk} son los valores de la k -ésima variable observados en los individuos i -ésimo y j -ésimo, respectivamente. R_k es el recorrido de la k -ésima variable, bien pudiendo tomarse el recorrido muestral. $\delta_{ijk} = 1$ en todos los casos. Sin importar el tipo de variable, δ_{ijk} valdrá cero cuando falta al menos uno de los dos valores involucrados en la comparación.

Una similaridad alta, es decir cercana a 1, indicara gran homogeneidad entre los individuos; por el contrario, una similaridad cercana a cero (0) indica que los individuos son diferentes (Valencia 2013).

6.12.2. Métodos de clasificación

Una vez que se ha obtenido la matriz de similaridades, el siguiente paso es aplicar una regla que permita identificar a los sujetos o variables que están más próximos o alejados entre ellos con el fin de conformar los grupos. Las técnicas de clasificación habituales se distinguen de acuerdo a: (i) la formación de grupos exclusivos o no exclusivos; (ii) formación de grupos jerárquicos y no jerárquicos; (iii) métodos directivos o iterativos, entre otros (Pérez 2007), donde las técnicas más comunes son las clasificaciones jerárquicas y no jerárquicas.

Métodos jerárquicos

El proposito en este método es agrupar conglomerados para formar un nuevo o separar uno ya existente para dar origen a otros dos, de tal forma que, si sucesivamente se va efectuando este proceso de aglomeración o división, se minimice alguna distancia o se maximice alguna medida de similitud. Una característica importante de estos métodos es que al localizar un objeto, es decir al agruparlo con otros o dejarlo solo, no se puede volver a mover en el transcurso del proceso. *En esta categoría los métodos se subdividen en aglomerativos y disociativos:* (Gallardo 2011).

Métodos aglomerativos: los algoritmos aglomerativos que se utilizan tienen siempre la misma estructura y sólo se diferencian en la forma de calcular las distancias entre grupos. Su finalidad es partir con tantos grupos como individuos haya, luego se van formando grupos de forma ascendente hasta que al final queda uno englobado en un solo conglomerado; estos algoritmos requieren menos tiempo de cálculo y son los más usados (Gallardo 2011). Su estructura es:

1. Comenzar con tantas clases como elementos, n . Las distancias entre clases son las distancias entre elementos originales.
2. Seleccionar los dos elementos más próximos en la matriz de distancias y formar con ellos una clase.

3. Sustituir los dos elementos utilizados en (2) para definir la clase por un nuevo elemento que represente la clase construida. Las distancias entre este nuevo elemento y los anteriores se calculan con uno de los criterios que comentamos a continuación.
4. Volver a (2) y repetir (2) y (3) hasta que tengamos todos los elementos agrupados en una clase única.

De acuerdo a Peña (2002), algunos criterios para definir distancias entre grupos son:

Encadenamiento simple o vecino más próximo: agrupa objetos basándose en la distancia mínima o la regla del vecino más próximo es decir la distancia entre los dos nuevos grupos es la menor de las distancias entre grupos antes de la fusión. Este método no funciona adecuadamente cuando los conglomerados no están bien definidos. Se ha comprobado que este criterio tiende a producir grupos alargados, que pueden incluir elementos muy distintos en los extremos.

Encadenamiento completo o vecino más alejado: la distancia entre los dos nuevos grupos es la mayor de las distancias entre grupos antes de la fusión. Tiende a producir conglomerados muy compactos y de forma esférica.

Media de grupos: la distancia entre los dos nuevos grupos es la media ponderada entre las distancias entre grupos antes de la fusión. Este criterio no es invariante ante transformaciones monótonas.

Método del centroide: se aplica generalmente solo con variables continuas. La distancia entre dos grupos se hace igual a la distancia euclídea entre sus centros, donde se toman como centros los vectores de medias de las observaciones que pertenecen al grupo. Cuando se unen dos grupos se pueden calcular las nuevas distancias entre ellos sin utilizar los elementos originales.

Método de Medianas: en el método de agrupación de medianas, los dos conglomerados (u objetos) que se combinan tienen idéntica ponderación en el cálculo del nuevo centroide combinado, independientemente del tamaño de cada uno de los conglomerados (objetos). Esto permite que a la hora de caracterizar los conglomerados resultantes, los conglomerados pequeños tengan la misma importancia que los conglomerados grandes.

Método de Ward: propuesto por Ward y Wishart, es un procedimiento jerárquico en el cual, en cada etapa, se unen los dos clusters para los cuales se tenga el menor incremento en el valor total de la suma de los cuadrados de las diferencias, o sea, la distancia euclídea al cuadrado entre cada individuo del cluster k a su centroide, y se denota como sigue (?):

$$E_k = \sum_{i=1}^{n_k} \sum_{j=1}^p (x_{ij}^k - \bar{x}_j^k)^2 = \sum_{i=1}^{n_k} \sum_{j=1}^n (x_{ij}^k)^2 - n_k \sum_{j=1}^n (\bar{x}_j^k)^2$$

$$E = \sum_{k=1}^h E_k$$

Donde x_{ij}^k es el valor de la j-ésima variable sobre el i-ésimo individuo, del k-ésimo cluster, suponiendo que dicho cluster posee n_k individuos. Luego $\bar{x}_j^k$ es la media de la variable j-ésima dentro del cluster k. El proceso comienza con m grupos, cada uno de los cuales está compuesto por un solo individuo, por lo que cada individuo coincide con el centro del grupo y por lo tanto en este primer paso se tendría $E_k = 0$. El objetivo del método de Ward es encontrar en cada etapa aquellos dos grupos cuya unión proporcione el menor incremento en la suma total de errores E_k . Luego, E es la suma de cuadrados de los errores para todos los grupos, es decir, si se supone que hay h grupos.

Métodos disociativos: estos métodos parten de considerar el conjunto de individuos en uno solo y sucesivamente se va separando en varios pequeños grupos hasta llegar a tantas agrupaciones como casos han sido tratados (Gallardo 2011).

Métodos no jerárquicos

Se caracteriza porque a diferencia del jerárquico, se conoce a priori el número de K grupos que se desea. El procedimiento consiste en elegir una partición de los individuos en K grupos e intercambiar los miembros de los clusters para tener una partición mejor, de tal forma que se maximiza la homogeneidad de los sujetos asignados a un mismo grupo y la heterogeneidad entre los distintos conglomerados. No establece relaciones de orden o jerárquicas entre dichos grupos (Peña 2002).

Método de las k-medias: se supone una muestra de n elementos con p variables. El objetivo es dividir esta muestra en un número de grupos prefijado, K . Este algoritmo requiere de cuatro etapas (Peña 2002):

1. Seleccionar K puntos como centros de los grupos iniciales, esto puede hacerse:
 - Asignando aleatoriamente los objetos a los grupos y tomando los centros de los grupos así formados.
 - Tomando como centros los K puntos más alejados entre sí.
 - Construyendo unos grupos iniciales con información a priori y calculando sus centros, o bien seleccionando los centros a priori.
2. Calcular las distancias euclídeas de cada elemento a los centros de los K grupos, y asignar cada elemento al grupo de cuyo centro esté más próximo. La asignación se realiza secuencialmente

y al introducir un nuevo elemento en un grupo se recalculan las coordenadas del nuevo centro del grupo.

3. Definir un criterio de optimilidad y comprobar si reasignando alguno de los elementos mejora el criterio.
4. Si no es posible mejorar el criterio de optimalidad, terminar el proceso.

El criterio de homogeneidad que se utiliza en el algoritmo de k-medias es minimizar la *suma de cuadrados dentro de los grupos (SCDG)* para todas las variables y se define como:

$$SCDG = \sum_{k=1}^K \sum_{j=1}^p \sum_{i=1}^{n_k} (x_{ijk} - \bar{x}_{jk})^2$$

donde x_{ijk} es el valor de la variable j en el elemento i del grupo k y $\bar{x}_{jk}$ la media de esta variable en el grupo. Este criterio es equivalente a la suma ponderada de las varianzas de las variables en los grupos, ya que puede escribirse como (Peña 2002):

$$\min(SCDG) = \min\left(\sum_{k=1}^K \sum_{j=1}^p n_k * s_{jk}^2\right)$$

donde n_k es número de elementos del grupo k y s_{jk}^2 es la varianza de la variable j en dicho grupo.

6.12.3. Determinación del número de conglomerados

En la aplicación habitual es necesario seleccionar el número de grupos sin embargo, no existe un proceso preciso de selección. Según Diaz (2007) para el caso del método jerárquico, los procedimientos heurísticos son los más usados, en el caso más simple, un árbol de clasificación que recibe el nombre de dendograma, es cortado por inspección subjetiva en diferentes niveles, donde puede obtenerse a partir de ello la conformación de los distintos grupos. Las líneas verticales representan los grupos unidos y la posición de la línea en la escala indica las distancias en las que se unen los grupos.

Para los métodos no jerárquicos se puede utilizar índices que miden tanto calidad como estabilidad de los clústers, aún con estas alternativas no se conoce una manera estándar para tomar esta decisión y lo mejor es valerse de relaciones teóricas o conocimientos de los investigadores de la temática específica que tengan indicios sobre el número de grupos, complementando el juicio que se tiene sobre esta cantidad (Dorado 2013).

Algunos índices de validación para determinar el número de conglomerados tomados de Valencia (2013) son:

- Trazar un gráfico que compare el número de conglomerados con la relación entre la varianza total de los grupos y la varianza entre ellos. El punto del gráfico donde se presente un recodo o doblez marcado indicará el número apropiado de grupos.
- Definir algún tipo de conceptualización intuitiva de la relación teórica de los datos.
- El índice de **Calinski y Harabasz** denominado como el estadístico pseudo-F consiste en evaluar en un rango de posibles números de cluster para luego inspeccionar el pico más alto el cual determinara el valor óptimo.

$$pseudo - F = \frac{traza(S_B)}{traza(S_W)} * \frac{n_p - 1}{n_p - k}$$

donde S_B es la matriz de suma de cuadrados y de productos cruzados entre los grupos, S_W es la matriz de de suma de cuadrados y matriz productos cruzados interior a los grupos, n_p es el número de observaciones y k es el número de clúster.

- El índice de **Davies Bouldin**:

$$S_{DB} = \frac{1}{k} \sum_k^{i=1} \max_{j=1, \dots, k; i \neq j} \frac{d_i + d_j}{d(c_i, c_j)}$$

donde k denota el número de clúster, i, j dos clúster, d_i, d_j son las distancias promedio de todos los individuos dentro del clúster a su respectivo centroide y $d(c_i, c_j)$ es la distancia entre centroides. Los valores bajos de Davies Bouldin corresponden a grupos que son compactos y cuyos centros están lejos el uno del otro; por consiguiente, el número de grupos que minimiza DB se toma como el número óptimo de las agrupaciones.

6.12.4. Combinación de métodos factoriales y de análisis de conglomerados

Algoritmo de Clasificación mixta: La propuesta desarrollada por Lebart & Piron (1995), aprovecha las ventajas del método de Ward y las del K-medias, combinándolos de la siguiente manera (texto tomado de (Pardo & Del Campo 2007)):

1. **Clasificación inicial:** Si la cantidad de individuos por clasificar es muy alta, es probable que la clasificación jerárquica no se pueda ejecutar directamente. Entonces se efectúa esta primera etapa, la cual busca obtener rápidamente y a bajo costo una partición de los individuos en s clases homogéneas, donde es mucho mayor que el número de clases deseado en la población, y menor que la cantidad de individuos. Se emplea el algoritmo de agregación alrededor de centros móviles (K-medias).

2. **Agregación jerárquica con el método de Ward:** Se efectúa una clasificación ascendente jerárquica donde los elementos terminales del árbol son las s clases de la partición inicial o los individuos directamente. El árbol correspondiente se construye según el criterio de Ward, el cual une en cada paso de agregación las dos clases que incrementen lo menos posible la inercia intraclases.
3. **Corte del árbol:** El árbol o dendograma que resume el procedimiento de clasificación permite ver la estructura de clases de los individuos que son objeto de análisis. Mediante un gráfico de índices de nivel será fácil observar los cambios de inercia más grandes (saltos) y decidir el número de clases K .
4. **Consolidación de la clasificación:** La partición obtenida en el paso anterior no es óptima siempre, debido a la estructura de particiones anidadas del Dendograma obtenido. Para mejorarla se utiliza de nuevo un procedimiento de agregación alrededor de centros móviles (K -medias), utilizando los centros de gravedad de las clases obtenidas al cortar el árbol como centros iniciales.

6.12.5. Caracterización de los grupos

Para seleccionar las variables más características de cada grupo, se mide la desviación entre los valores relativos a la clase y los valores globales, utilizando los valores test (Lebart & Piron 1995).

Para una **variable continua** X , con media $\bar{X}_j$, el valor test asociado a la media $\bar{X}_{jk}$ del grupo k esta definida de la siguiente forma:

$$t_{jk}(X) = \frac{\bar{X}_{jk} - \bar{X}_j}{\sqrt{\frac{n-n_k}{(n-1)n_k} S_{x_j}}}$$

donde S_X es la desviación estándar de la variable X_j en todo el conjunto de datos, n la cantidad de individuos y n_k cantidad de individuos en la clase k .

7. Metodología

Para dar cumplimiento a los objetivos planteados en el proyecto, se propone la aplicación de métodos multivariados como el análisis factorial múltiple y el análisis de conglomerados que permitan caracterizar y distinguir aspectos importantes del sistema agrícola del cultivo del banano con la información disponible de suelo, índice de vegetación y clima de las distintas unidades de manejo distribuidas en el Urabá Antioqueño; investigación que está a cargo del Centro Internacional de Agricultura Tropical (CIAT).

7.1. Población objeto de estudio

Son todos los cultivos de banano representados en unidades de manejo, ubicados en el Uraba antioqueño. Se pretende que la investigación proporcione resultados que describan zonas donde se establecen estos cultivos, por tal razón no se le da importancia a la ubicación geográfica de las parcelas donde se tomó la información, sino la ubicación ecológica que las distinguen unas de otras.

7.2. Unidad experimental

Se definen las unidades experimentales como el "lote o Unidad de manejo"; estos deben cumplir con los siguientes criterios de inclusión:

- Lotes de parcelas dedicadas al cultivo del banano, pertenecientes al Urabá.
- Lotes que el agricultor ha autorizado caracterizar y ha compartido información de producción.

7.3. Aspectos éticos

Para la recolección de datos el productor manifiesta estar de acuerdo en compartir la información de la experiencia en su cultivo y autoriza que se practique la caracterización del suelo por el método RASTA; este método no causa afectación sobre el cultivo, ni sobre el terreno. La información proporcionada permanece confidencial para cada productor.

7.4. Muestra

Debido a que no se tiene un marco muestral no es posible aplicar métodos de muestreo, por lo que se define una investigación Cuasi-Experimental en la que se busca que la información capturada represente solo a los productores del banano.

7.5. Descripción de la base de datos

Debido a las diferencias que existen entre las resoluciones espaciales e incluso de los tipos de datos climáticos, de NDVI y de suelos, el CIAT uniformizó los datos tanto en formato como en resolución, para lograr un óptimo resultado con cualquier procedimiento estadístico que se realice, estableciendo que la resolución espacial para el estudio sería de 3 arco-segundos (90m), la cobertura sería de toda el área donde hay fincas y el formato de trabajo para todos los datos es el raster (grillas).

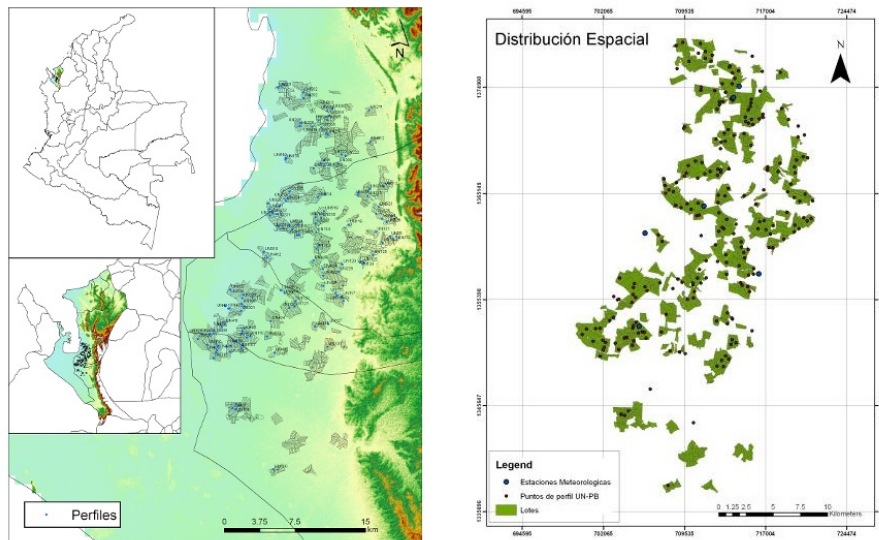


Figura 7-1.: Distribución de lotes productores de Banano

«Para obtener una superficie que describiera el gradiente espacial de cada una de las variables de suelo, usaron el método de interpolación Thin Plate Spline (TPS) desarrollado por Hutchinson (1984), que fue aplicado a los datos de los estudios de suelos de 296 fincas bananeras presentes en la zona usando cada una de las variables como variables dependientes y ajustaron su distribución a una función multivariada usando la altitud, la latitud y la longitud, dicha función fue proyectada sobre toda el área bajo análisis, lo cual permitió determinar de una manera aproximada los patrones de distribución espacial de las variables, obteniendo de esta manera, un total de 78638 diferentes puntos (pixeles) que contienen los valores para cada una de estas

características.» Este proceso de uniformización de los datos, permite la aplicación de un método estadístico para realizar la zonificación. La figura **8-1** muestra la distribución espacial de las fincas productoras de banano en el Urabá antioqueño.

7.5.1. Revisión y selección de variables

En base a estudios anteriores y las necesidades relacionadas a la fisiología del cultivo producido en las fincas del Urabá, se procede a la selección de las variables que entran a participar en la clasificación de las unidades de manejo resultante de la combinación del análisis factorial y el de conglomerados, a fin de no involucrar información redundante en los cálculos, además de la inspección de los coeficientes de variación de cada una de la variables y el cálculo de los coeficientes de correlación.

■ Variables climáticas

Se cuenta con 19 variables bioclimáticas y son obtenidas a partir de la georeferenciación del lote (**ANEXO A**, tabla **A-1**); según Worldclim estas variables son derivadas del valores mensuales de la temperatura y la precipitación, representando tendencias anuales, estacionalidades y factores limitantes del sitio georeferenciado, con ellas se realizó una selección bajo criterios agronómicos como aquellos más relacionados con la fisiología y productividad del cultivo en cuestión. Además una análisis de correlación se llevó a cabo con el fin de enriquecer estadísticamente la selección previa de estas variables, que expresa el grado de asociación entre ellas, en el **ANEXO B**, tabla **B-3** se puede apreciar la matriz de correlaciones en la que variables de los coeficientes cercanos a +1 indicarían correlaciones fuertes y directas, igual sucede pero de forma indirecta cuando el coeficiente tienda a tomar valores cercanos a -1, mientras que valores próximos a 0 indicaría que no hay relación entre variables.

Nombre	Descripción	Unidad de medida	Clasificación
Bio 1	Temperatura media anual	Grados centígrados	Climatica
Bio 4	Estacionalidad de la temperatura	Porcentaje	Climatica
Bio 5	Temperatura máxima del mes más caliente	Grados centígrados	Climatica
Bio 7	Rango de temperatura anual	Grados centígrados	Climatica
Bio 12	Precipitación total anual	Milímetros	Climatica
Bio 14	Precipitación del mes más seco	Milímetros	Climatica
Bio 15	Estacionalidad de la precipitación	Porcentaje	Climatica

Tabla **7-1.**: Variables climáticas seleccionadas

Teniendo en cuenta lo anterior y que el ciclo productivo del sitio es de un periodo de cosecha, se seleccionaron las condiciones secas ya que según los expertos las sequías prolongadas provocan la caída de hojas lo que posiblemente reduciría el rendimiento en el cultivo, tal es el caso de las variables Bio 5 y Bio 14 que describen condiciones extremas, también se incluyeron las de

largo plazo como la temperatura promedio anual (Bio 1) y la precipitación anual (Bio 12), y aquellas relacionadas con la variación estacional de las temperaturas y precipitaciones Bio 4, Bio 7 y Bio 15, que describen de forma general el clima a lo largo del año (ver Tabla 7-1).

■ Variables del NDVI

Se involucran también variables relacionadas con el índice de vegetación normalizada, las cuales representan tendencias centrales y de dispersión en los datos, tales como el promedio, el rango medio anual y el rango total, que permiten identificar la cantidad de vegetación presente en el sitio.

Nombre	Descripción	Unidad de medida	Clasificación
nv mean	Promedio historico del NDVI	Porcentaje	NDVI
nv mrange	Rango medio anual del NDVI	Porcentaje	NDVI
nv range	Rango total historico del NDVI	Porcentaje	NDVI

Tabla 7-2.: Variables del NDVI seleccionadas

■ Variables del Suelo

Las variables de suelo que se miden en base al procedimiento metodológico o guía práctica RASTA no se utilizan en su totalidad para el estudio, se toman en cuenta 6 de ellas para ser incluidas en el análisis ya que afectan el desarrollo de una forma directa o indirecta sobre el comportamiento en el cultivo, es decir aquellas relacionadas con el rendimiento o productividad del cultivo del banano.

Nombre	Descripción	Unidad de medida	Clasificación
chid mean	Conductividad hidráulica promedio	cm/hora	Suelo
diam pond	Diámetro medio ponderado	Milímetros	Suelo
nive frea	Nivel freático	cm	Suelo
perc arc	Porcentaje de arcilla	Porcentaje	Suelo
perc aren	Porcentaje de arena	Porcentaje	Suelo
ph	Potencial de hidrogeno	Concentración por iones	Suelo

Tabla 7-3.: Variables del Suelo seleccionadas

Según los criterios agronómicos la adecuada relación entre estos componentes determina la capacidad de hacer crecer las plantas y la disponibilidad de suficientes nutrientes para ellas, por tanto las variables de interés son: el diámetro medio ponderado (diam_pond), nivel freático (nive.frea), porcentaje de arcilla (perc_arc), porcentaje de arena (perc_aren), ph y la conductividad hidráulica promedio (chid_mean). De modo general estas variables determinan propiedades físicas y químicas del terreno como la textura, estructura, permeabilidad, porosidad, consistencia y profundidad. Cabe notar que la conductividad hidráulica por método barrenado y la de laboratorio son variables que miden la misma característica por tanto fueron

promediadas generando así la nueva variable(chid_mean) a fin de no perder información ni tampoco introducir información redundante en el análisis.

7.6. Análisis exploratorio

Para dar solución al primer objetivo planteado se realiza un análisis exploratorio para el conjunto de variables edafoclimáticas seleccionadas para conocer propiedades y condiciones importantes del terreno y del cultivo del banano; teniendo en cuenta la naturaleza de los datos, se utiliza medidas de centramiento como el promedio y como indicadores de variabilidad, la desviación estándar y el coeficiente de variación.

7.7. Análisis multivariado

En base a los objetivos planteados, como descripción estadística de los datos edafoclimáticos de las unidades de manejo se utilizaron los algoritmos mixtos donde se conjugan los métodos factoriales y de clasificación para describir grupos de individuos o UM de los cultivos de banano bajo condiciones ambientales similares.

Para justificar la implementación del análisis factorial y el de conglomerados, el primer paso a seguir fue realizar un análisis de correlaciones que determina el grado de asociación entre las variables, en el **ANEXO B**, la tabla **B-1**, tabla **B-2** y la tabla **B-3**, muestran la matriz de correlaciones para cada grupo de variables y en las que se logra evidenciar la existencia de muchas correlaciones estadísticamente significativas superiores $|\rho| > 0,8$. De este modo se procedió a desarrollar la metodología del análisis factorial múltiple que se constituye como un pretratamiento a la clasificación de los individuos o unidades de manejo, que transforma los datos originales en variables continuas no correlacionadas, como propone (Lebart & Piron 1995), utilizando las coordenadas factoriales del AFM, el cual lleva a cabo inicialmente un análisis de componentes principales (ACP) a cada grupo de variables, para extraer de estos el primer valor propio asociado a la primera componente principal.

Teniendo en cuenta que la naturaleza de los grupos de variables son de escala continua, en donde X_1 es la matriz de variables climáticas de dimensión 78638×7 representando en cada fila i la unidad de manejo y cada columna la variable climática j ; X_2 , corresponde al grupo de variables del NDVI de dimensión 78638×3 y X_3 , es la matriz de variables continuas del suelo de dimensión 78638×6 , de tal manera que:

$$X_1 = \begin{pmatrix} 26,6 & 21,08 & \cdots & 33,04 \\ 26,5 & 20,9 & \cdots & 32,99 \\ \vdots & \vdots & \ddots & \vdots \\ 27,7 & 31,4 & \cdots & 47,9 \end{pmatrix} \quad X_2 = \begin{pmatrix} 26,6 & 21,08 & \cdots & 33,04 \\ 26,5 & 20,9 & \cdots & 32,99 \\ \vdots & \vdots & \ddots & \vdots \\ 27,7 & 31,4 & \cdots & 47,9 \end{pmatrix} \quad X_3 = \begin{pmatrix} 0,75 & 0,023 & 0,058 \\ 0,76 & 0,025 & 0,061 \\ \vdots & \vdots & \vdots \\ 0,79 & 0,028 & 0,069 \end{pmatrix}$$

Análisis de componentes principales por separado

Se realiza un Análisis de componentes principales por cada grupo de variables, este método se desarrolló a partir de la matriz de datos estandarizados $Z_{78638 \times 7}$, tal como se hizo para el conjunto de variables del NDVI y suelo, para extraer los valores característicos (λ_i) asociados a los ejes o factores.

Las componentes principales se escribieron como una combinación lineal de las variables originales, es decir:

$$\begin{pmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{178638} \end{pmatrix} = \begin{pmatrix} z_{11} & z_{12} & \cdots & z_{17} \\ z_{21} & z_{22} & \cdots & z_{27} \\ \vdots & \vdots & \ddots & \vdots \\ z_{786381} & z_{786382} & \cdots & z_{786387} \end{pmatrix} * \begin{pmatrix} u_{11} \\ u_{12} \\ \vdots \\ u_{17} \end{pmatrix}$$

El cálculo de cada componente se hizo por medio de multiplicadores de Lagrange, de lo cual interesa el cálculo del primer valor propio.

$$Y_1 = u_{11}(Bio1) + u_{12}(Bio4) + u_{13}(Bio5) + u_{14}(Bio7) + u_{15}(Bio12) + u_{16}(Bio14) + u_{17}(Bio15)$$

$$Y_2 = u_{21}(Bio1) + u_{22}(Bio4) + u_{23}(Bio5) + u_{24}(Bio7) + u_{25}(Bio12) + u_{26}(Bio14) + u_{27}(Bio15)$$

$$\vdots$$

$$Y_7 = u_{71}(Bio1) + u_{72}(Bio4) + u_{73}(Bio5) + u_{74}(Bio7) + u_{75}(Bio12) + u_{76}(Bio14) + u_{77}(Bio15)$$

Del mismo modo se realizó un ACP para el conjunto de variables del índice de vegetación y para el suelo ya que son de escala continua.

Análisis de componentes principales global

Después de cada análisis factorial, se tiene en cuenta en cada caso el primer valor propio, con el fin de llevar a cabo la siguiente ponderación:

$$\frac{1}{\lambda_1^1} = \frac{1}{6,384}, \quad \frac{1}{\lambda_1^2} = \frac{1}{1,967} \quad y \quad \frac{1}{\lambda_1^3} = \frac{1}{3,584}$$

Donde cada λ_1^j , es el primer valor propio de cada ACP realizado para cada grupo $j = 1, 2, 3$, clima, NDVI y suelo, respectivamente. Así, se equilibra el peso de los grupos en la obtención del primer factor de manera que el peso de cada grupo es el mismo y no dependerá del número y calidad de las variables empleadas. Para la implementación del ACP global se incluye todas la variables edafoclimáticas de tal manera que las matrices (X, M, D), tengan la siguiente estructura:

- X: es la matriz de datos transformados $X = (Z_1, Z_2, Z_3)$, que contiene los valores estandarizados de las variables de clima (Z_1), de NDVI (Z_2) y suelo (Z_3) calculados a partir de la siguiente expresión:

$$\frac{x_{ijk} - \bar{x}_{kj}}{\sigma_{kj}}$$

donde $i = 1, 2, \dots, 78638$, individuos, $k = 1, \dots, 16$, variables y $j = 1, 2, 3$ grupos de variables.

$$X = \begin{pmatrix} -1,56 & \cdots & -1,15 & \cdots & 1,88 \\ -1,57 & \cdots & -1,05 & \cdots & 1,83 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 2,42 & \cdots & -0,3 & \cdots & -1,56 \end{pmatrix}_{78638 \times 16}$$

- M: es la matriz métrica $M = (M^{J_{c1}}, M^{J_{c2}}, M^{J_{c3}})$, donde $M^{J_{cj}} = \frac{1}{\lambda_1^j} I_{K_j}$, siendo el K el número de variables en el grupo $j = 1, 2, 3$, la matriz identica I de orden $K_j \times K_j$.

$$M^{J_c} = \begin{pmatrix} 0,156 & 0 & \cdots & 0 \\ 0 & \ddots & & 0 \\ \vdots & & 0,508 & \vdots \\ 0 & 0 & \cdots & 0,278 \end{pmatrix}_{16 \times 16}$$

- D: es la matriz de pesos $D = \frac{1}{I} I_I$, donde I_I , es la matriz identica de tamaño $I \times I$.

$$D = \begin{pmatrix} 0,0000127 & 0 & \cdots & 0 \\ 0 & \ddots & & 0 \\ \vdots & & 0,0000127 & \vdots \\ 0 & 0 & \cdots & 0,0000127 \end{pmatrix}_{78638 \times 78638}$$

El cálculo de las componente del ACP global se obtuvo a partir de la expresión:

$$F_j = X M u_j$$

donde u_j , es el vector propio obtenido de la diagonalización de la matriz $X M X' D$.

De aquí se procedió a la selección de las componentes que logran explicar la mayor variabilidad de la información, respecto a algunos criterio mencionados por Peña (2002).

Combinación del ACP global y el análisis de conglomerados

En base a los resultados obtenidos del análisis factorial, se toma como matriz de entrada para la clasificación de las unidades de manejo, las coordenadas factoriales ya seleccionadas, como propone (Lebart & Piron 1995), incorporando la clasificación jerárquica por el método de ward:

$$E_k = \sum_{i=1}^{n_k} \sum_{j=1}^n (x_{ij}^k - \bar{x}_j^k)^2$$

donde x_{ij}^k , es el valor en la componente j del ACP global, para la UM i del grupo k de las variables edafoclimáticas, $\bar{x}_j^k$, es la media de la componente j dentro del grupo k de las variables edafoclimáticas y n_k , es el número de unidades de manejo en el grupo k .

Este algoritmo indica que cada una de las unidades de manejo conforma un grupo, aglomerándose en cada paso.

A partir de la inspección del dendograma el cual resume el procedimiento de clasificación de las UM, y el histograma del índice de nivel, se logra tener una idea sobre el número de conglomerados a seleccionar; y para contar con otras particularidades en dicha decisión se tomó en cuenta dos criterios internos basados en el cálculo de índices de calidad (Davies Bouldin y Calinski-Harabasz).

Posteriormente se realiza la consolidación de la clasificación mediante **el método no jerárquico de las k-medias**, el cual requiere del número de clases que se decidió por el método de agrupamiento jerárquico anterior, como puntos iniciales para la obtención de una partición que minimice la inercia intraclases, de tal manera que:

$$SCDG = \sum_{k=1}^K \sum_{j=1}^p \sum_{i=1}^{n_k} (x_{ijk} - \bar{x}_{jk})^2$$

donde x_{ijk} es el valor de la componente j del ACP global para UM i del grupo k de las variables edafoclimáticas y $\bar{x}_{jk}$, es la media de la componente j dentro del grupo k .

Caracterización de los conglomerados y descripción de variables

Para distinguir las variables más características de cada grupo conformado se emplea los valores test que son índices descriptivos contruidos siguiendo la metodología de pruebas de hipotesis, pero sin el objetivo de hacer inferencias; con ello se tiene que para las variables continuas, los valores

test resultan de la comparación del promedio de la variable dentro de la clase con el promedio global, por tanto, la ordenación de los valores test dentro de cada clase permiten detectar las variables que caracterizan positivamente a cada una de las clases, dado que la media del grupo es suficientemente mayor que la media global, o negativamente cuando la media de la clase es inferior (Pardo & Del Campo 2007); por consiguiente se plantea la hipótesis nula (H_o) de que la media de la clase y la media global son iguales, frente a la hipótesis alternativa (H_a) de que la media de la clase y la media global son diferentes, aplicando el estadístico de prueba t-Student, con el fin de decidir sobre la importancia de las variables dentro de cada clase.

Finalmente se realiza su respectivo análisis descriptivo que sirve de soporte para comprender su comportamiento dentro de cada conglomerado.

7.8. Herramienta computacional

Para la consolidación de los resultados, se utilizó el programa estadístico **R Project for Statistical Computing** versión 3.1.0, que es un sistema para análisis estadísticos y gráficos, además es una alternativa de software libre.

Fue necesario la implementación de paquetes como el `sp`, `raster`, `rgdal` y `spatstat` para leer y procesar los datos que estaban en formato raster, el paquete `FactoMineR` se utilizó para la elaboración del AFM y la librería `FactoClass` para formar los grupos edafoclimáticos. También se recurre a la función `intCriteria` del paquete `clusterCrit` para el cálculo de los índices de validación: Davies Bouldin y Calinski-Harabasz.

8. Resultados

En este capítulo se presenta los resultados obtenidos a partir del análisis exploratorio de cada grupo de variables previamente seleccionadas para conocer propiedades y condiciones importantes del terreno y del cultivo del banano; además los resultados obtenidos del proceso de clasificación de las unidades de manejo según características similares mediante la conjugación del análisis factorial múltiple y el análisis de conglomerados, evaluando también los aspectos descriptivos de cada grupo respecto a las variables más significativas.

8.1. Análisis descriptivo

8.1.1. Análisis de las variables climáticas

En la tabla 8-1, se observa que las variables referentes a la temperatura, Bio 1 hasta Bio 7, son poco variables a lo largo del periodo analizado debido a que los coeficientes de variación están por debajo del 5 % a excepción de Bio 4 que representa la variación de la temperatura, obtuvo un valor medio de 25.47 % con una dispersión de 3.1413 % y coeficiente de variación de 12.33 %, indicando que en general las unidades de manejo parecen tener algunos meses calurosos y otros con temperaturas bajas, siendo una clara muestra del clima tropical en Colombia que se distingue por tener en el año dos periodos de invierno Eslava (1993).

	Min	Mean	Max	DesvEstand	CV
bio_1	26.3	26.8	27.12	0.124	0.004
bio_4	20.1	25.47	31.85	3.141	0.123
bio_5	30.9	31.81	32.54	0.361	0.011
bio_7	8.7	9.60	10.33	0.384	0.039
bio_12	2459	2796	3547	282.3	0.101
bio_14	65.9	75.54	91.15	8.384	0.111
bio_15	32.9	37.93	48.42	4.116	0.108

Tabla 8-1.: Estadísticas descriptivas variables de clima

Además se tiene que en promedio la Temperatura anual (Bio 1) de las unidades de manejo es de $26,54^{\circ}C$ valor que esta contenido dentro del rango de condiciones habituales para el cultivo de banano que varía entre $15^{\circ}C$ y $27^{\circ}C$. Por otra parte para el caso de las variables referentes a la precipitación Bio 12, Bio 14 y Bio 15, presentan mayor variabilidad, señalando que la cantidad

de lluvias difiere en algunos meses del año, además se evidencia que en promedio las unidades de manejo tienen una precipitación anual (Bio 12) de 2796 mm, condición que se encuentra dentro de los requerimientos para la producción del banano entre 2000 mm y 5000 mm. Las características encontradas hacen del Urabá Antioqueño una zona apta para la producción del banano.

8.1.2. Análisis de las variables del índice de vegetación normalizada (NDVI)

Teniendo en cuenta que este índice arroja valores de intensidad del verdor de la zona y da cuenta de la cantidad de vegetación presente en la superficie y su estado de salud o vigor vegetativo.

	Min.	Mean	Max.	DesvEsta	CV
nv_mean	0.423	0.805	0.893	0.05	0.06
nv_mrange	0.011	0.027	0.097	0.01	0.23
nv_range	0.024	0.062	0.181	0.02	0.24

Tabla 8-2.: Estadísticas descriptivas para el NDVI

En la tabla 8-2, se observa que la zona presenta abundante vegetación dado que en promedio el nivel vegetación es aproximadamente de 0.805, según Meneses (2011), los valores cercanos a 1 son indicadores de la actividad fotosintética de bosque templado, de selva y actividad agrícola, además existe un nivel de diferencia significativa en las variables que representan variabilidad temporal (dispersión) lo que sugiere que dicha condición no es uniforme entre áreas de la región bananera.

8.1.3. Análisis de las variables del Suelo

Entre las variables de suelo que fueron caracterizadas por el RASTA se encuentran algunas propiedades físicas y químicas que se calcularon de forma cuantitativa y que describen las condiciones del suelo presentes en el Urabá.

	Min.	Mean	Max.	DesvEstand	CV
chid_mean	-10.69	4.69	7.050	1.09	0.23
diam_pond	-1.37	3.69	5.53	1.10	0.30
nive_frea	116.5	151.9	177.5	12.1	0.08
perc_arci	33.84	44.75	95.49	4.44	0.10
perc_aren	9.05	17.68	59.25	4.80	0.27
ph	5.23	5.62	6.21	0.16	0.03

Tabla 8-3.: Estadísticas descriptivas del Suelo

En la tabla 8-3, se logra evidenciar que hay un determinado rango de variabilidad entre las diferentes características, lo que sugiere que existen gradientes a través de la región y que probablemente hay un nivel de similitud al cual ciertas regiones comparten características, lo que hace posible una

clasificación agroecológica con el fin de optimizar los recursos y la producción. Posee suelos neutros debido a que su ph oscila entre 5.23 y 6.21, teniendo en cuenta la textura se puede decir que la zona es apta para la producción de banano ya que en promedio posee 44.75 % de arcilla y 17.65 % de arena lo que supone un equilibrio entre la permeabilidad del agua y retención de agua y de nutrientes.

8.2. Análisis multivariado

8.2.1. Análisis factorial múltiple (AFM)

El AFM es una técnica que permite reducir de la dimensionalidad inicial de la información existente, basada en el ACP, con el que se pretende dejar las bases para explicar el comportamiento de los individuos en función de los datos que cada uno de ellos posee sobre un gran conjunto de variables agrupadas en diferentes subtablas o grupos (Clima, NDVI y Suelo), en razón de su naturaleza, tipología, y características. Como todo análisis factorial se sustenta en el concepto de inercia, como una medida sintética de dispersión y en la utilización de una distancia que determinará la métrica a seguir. Se analiza individualmente cada grupo de variables o subtablas en diferentes fases a través de un ACP, ponderando cada subtabla por el inverso del primer valor propio del análisis precedente. Así, se equilibra el peso de los grupos en la obtención del primer factor de manera que el peso de cada grupo es el mismo y no dependerá del número y calidad de las variables empleadas. Posteriormente se calculan las coordenadas de las variables de cada uno de los grupos respecto a los factores para obtener la representación global, en este documento sólo se describen las relaciones entre grupos de variables y factores del análisis de componentes principales global.

	Valores propios	% Varianza	% Var. Acum.
comp. 1	1.489	34.68	34.68
comp. 2	1.020	23.74	58.43
comp. 3	0.867	20.19	78.62
comp. 4	0.383	8.920	87.54
comp. 5	0.303	7.063	94.61
comp. 6	0.114	2.665	97.27
⋮	⋮	⋮	⋮
comp. 15	0.00099	0.023	99.98
comp. 16	0.00051	0.011	100

Tabla 8-4.: Valores propios de las variables edafoclimáticas

Para la decisión de cuantos componentes seleccionar, se tuvo en cuenta algunos criterios que se encontraban en la literatura, entre ellos el de escoger la cantidad de ejes que recogen un porcentaje de variabilidad como mínimo del 75 %, o seleccionar aquellos con valores propios mayores a 1. En la tabla 8-4 y la figura 8-1, se observa que los tres primeros ejes factoriales logran una representación

apropiada constituida por el 78.62 % de la variabilidad total, demostrando que es posible reducir el número de variables con los nuevos factores los cuales tienen un elevado grado de información.

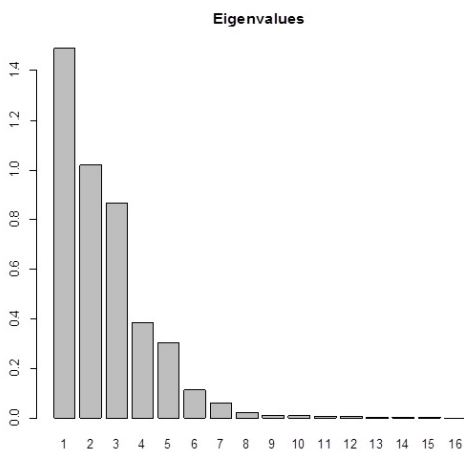


Figura 8-1.: Histograma de los valores propios

Con el objetivo de comparar globalmente los tres grupos de variables analizadas, cada uno es representado mediante un punto y proyectado sobre el primer plano factorial del AFM. La coordenada sobre un eje representa la inercia proyectada del grupo de variables sobre el eje e indica la importancia en la determinación de la dimensión (figura 8-2).

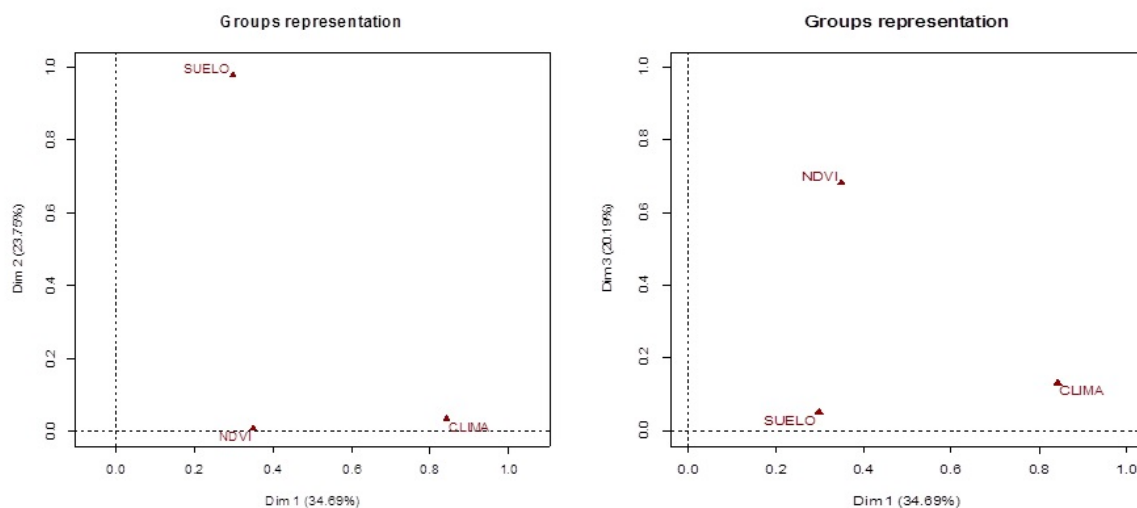


Figura 8-2.: Representación de los grupos en los planos factoriales 1-2 y 1-3

Se observa que el grupo de variables del clima presentan mayor proximidad al primer eje factorial, lo que sugiere que la primera dimensión puede estar explicada por el comportamiento de la temperatura y precipitación. Por otra parte las variables del suelo aportan significativamente a la formación del segundo eje factorial dada su cercanía por lo que esta dimensión está explicada por las propiedades físicas y químicas del terreno, consecuentemente se tiene que el NDVI ayuda a la determinación de la tercera dimensión.

Para comprender la relación entre los primeros factores y las variables, se recurre al círculo de correlaciones (ver Figura 8-3), en el que se muestra como quedan representadas las variables en las tres primeras dimensiones del ACP global y se puede apreciar que todas las variables Climáticas quedan representadas en el primer eje factorial, mientras que la mayoría de variables de Suelo están representadas en el segundo eje.

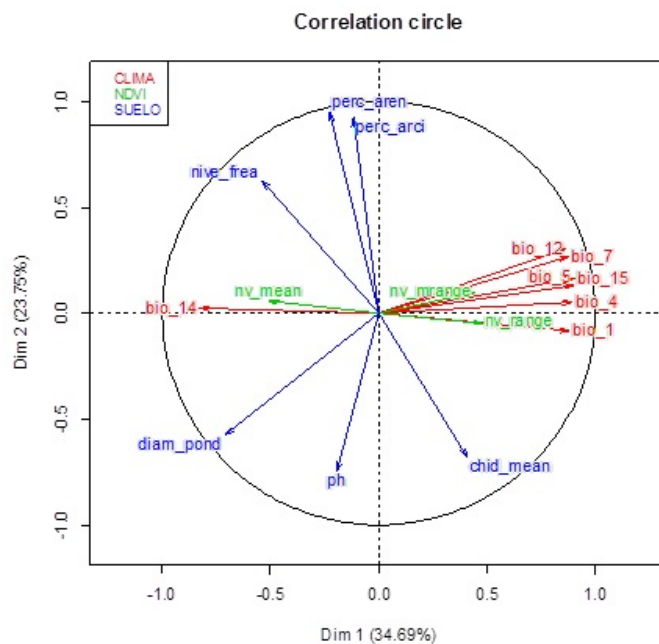


Figura 8-3.: Representación de las variables sobre el primer plano factorial 1 - 2

En el círculo de correlaciones (Figura 8-4), correspondiente a los ejes factoriales 1 y 3, se observa que aquellas variables que representan variabilidad temporal y tendencia central del NDVI aportan significativamente a la formación del tercer componente de modo que está asociado a la cantidad de vegetación y su variabilidad entre áreas de la región bananera.

Por lo tanto se concluye que el componente 1 es un indicador del comportamiento del clima, en el

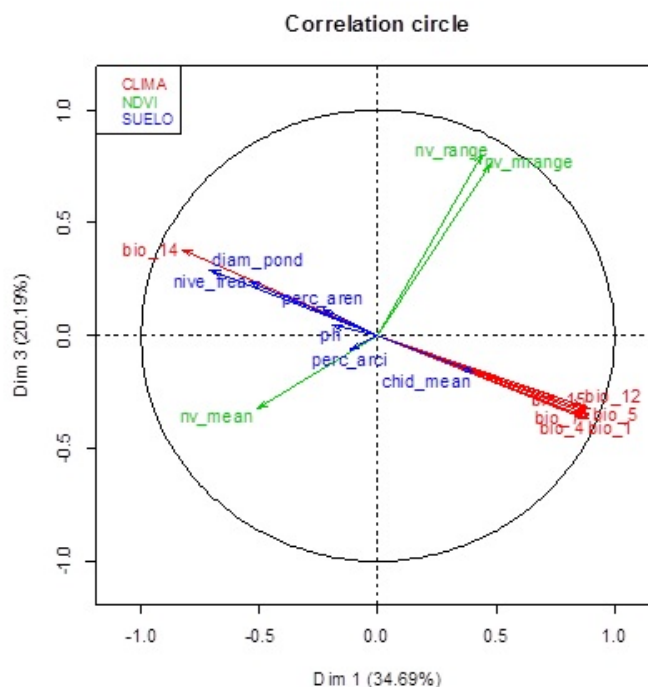


Figura 8-4.: Representación de las variables sobre el plano factorial 1 - 3

segundo eje se puede identificar con las propiedades físicas y químicas del suelo y el tercer eje indica la presencia de vegetación en la superficie.

8.2.2. Análisis de Conglomerados

Para la clasificación de las unidades de manejo de las fincas productoras de banano bajo condiciones ambientales similares, se usa como datos de entrada, las **primeras tres coordenadas factoriales** arrojadas por el ACP global que explican la mayor variabilidad (78.62%) de la información (ver Tabla 8-4), esto debido a que permiten balancear los grupos de variables y eliminar el ruido.

Se procedió entonces, a realizar el análisis de clasificación jerárquica aglomerativa por el método de Ward, que da como resultado el dendograma correspondiente a la primera gráfica de la figura 8-5 el cual resume el procedimiento de clasificación y en el se puede observar la estructura de las clases, e indica que hay dos grupos más representativos, sin embargo este es un valor subjetivo por tal motivo se incluyen otros criterios como el índice de nivel que se muestra en la segunda gráfica (Figura 8-5), según este método al conformar 5 clúster la ganancia en la inercia entre ellos es 0.58 relativamente alta comparada a cuando se conforman 6 clúster (0.4 aproximadamente), adicionalmente se utiliza

dos criterios internos basados en el cálculo de índices de calidad que permite conocer cuál es la mejor partición.

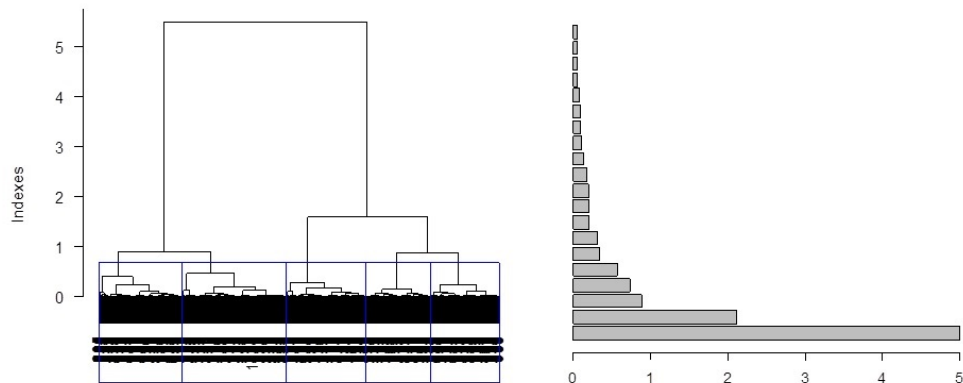


Figura 8-5.: Dendrograma e Histograma del índice de Nivel para variables edafoclimáticas

La primera gráfica de la figura 8-6, muestra el índice de Davies Bouldin calculado para diferentes particiones de los datos, teniendo en cuenta que el número de grupos que minimiza al DB se toma como el número óptimo de las agrupaciones se tiene entonces que la mejor partición se da cuando se tienen 5 grupos, en donde el índice toma el menor valor (0.537).

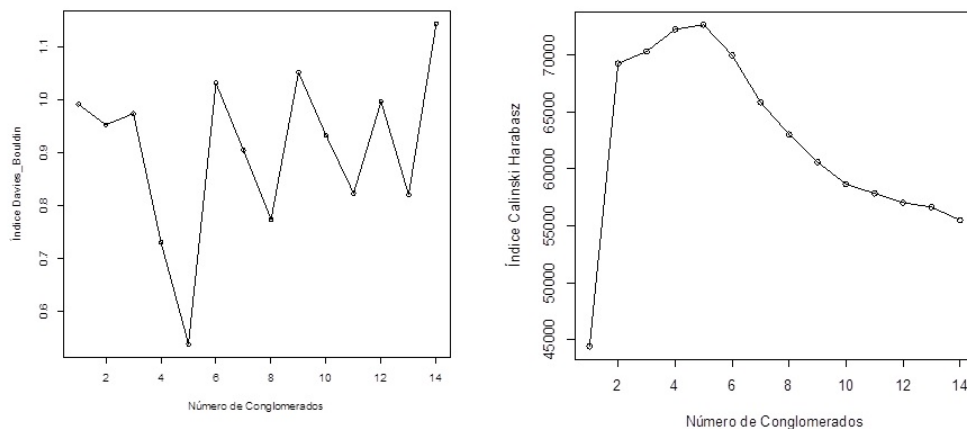


Figura 8-6.: Índice de Davies Bouldin y Calinski-Harabasz para variables edafoclimáticas

Respecto a lo obtenido por el índice de Calinski-Harabasz (Figura 8-6), la mejor partición es la que corresponde al mayor valor (75052.9) por lo que sugiere que se deberían elegir 5 grupos. Del comportamiento de los anteriores criterios se decide que el número adecuado para agrupar las

unidades de manejo e indicar al método de partición de las K-medias que inicie el algoritmo es con cinco conglomerados.

Posteriormente se realiza la consolidación de la clasificación mediante el proceso de agregación alrededor de centros móviles (K-medias) el cual requiere del número de las clases obtenidas en el procedimiento anterior como puntos iniciales para la obtención de una partición que minimice la inercia intraclases. En la primera gráfica de la figura 8-7 se muestra la conformación de los cinco grupos sobre el primer plano factorial del AFM y en la segunda gráfica (Figura 8-7) se presenta las unidades de manejo georeferenciadas sobre el Urabá Antioqueño identificando las cinco zonas con características edafoclimáticas similares, las cuales se puede diferenciar por colores.

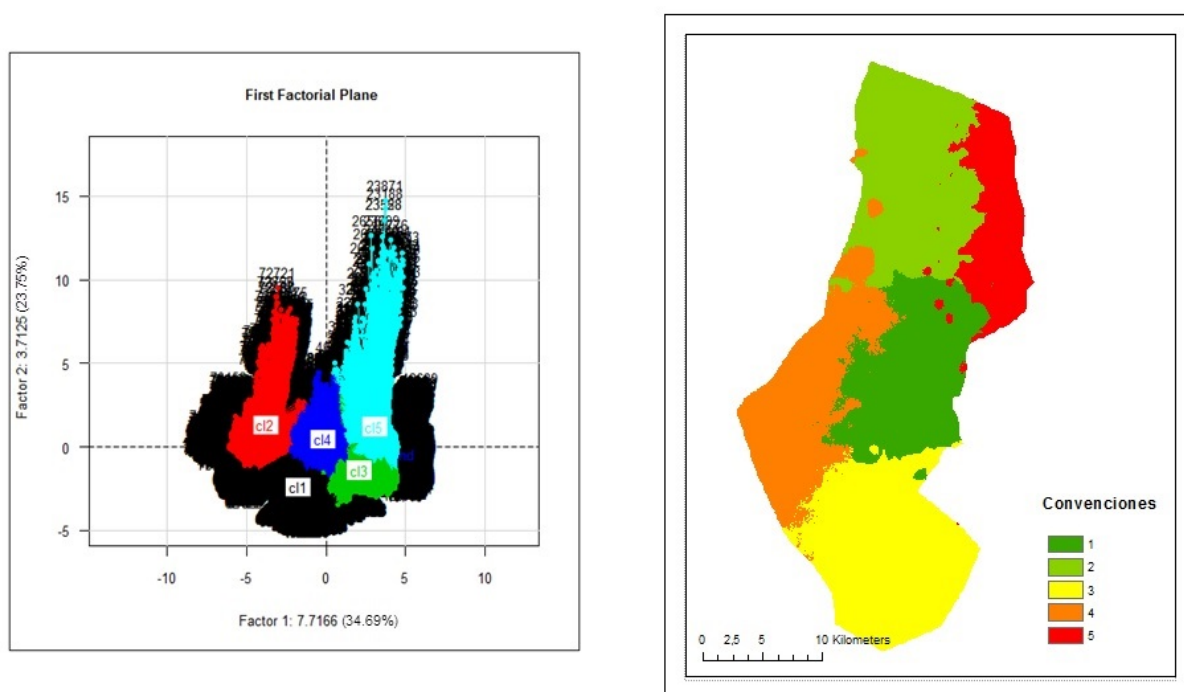


Figura 8-7.: Representación de los individuos sobre el primer plano factorial y la distribución de los lotes agrupados sobre el Urabá Antioqueño

8.2.3. Caracterización de los conglomerados y descripción de variables

Para distinguir las variables más significativas de cada grupo conformado, se emplea los valores test que son cuantiles de la distribución normal estándar, en el sentido en que su media dentro de los grupos es razonablemente superior a la global; además se realiza su respectivo análisis exploratorio el cual sirve de soporte para comprender su comportamiento dentro de cada clase. A continuación

se presenta los resultados obtenidos y en el **ANEXO C** se encuentran las tablas referidas a las estadísticas descriptivas de cada grupo.

Grupo 1: En la tabla **8-5** se evidencia que las propiedades físicas del suelo (perc arci, perc aren, diam pond, nive frea), el promedio de vegetación (nv mean), la temperatura y su variabilidad (bio 1, bio 5 y bio 7) son las características más significativas de la clase dado que su media dentro del grupo es superior a la media global, es necesario tener en cuenta que como el tamaño de la clase y el poblacional (15402 y 78638 respectivamente) son muy grandes y el error estandar asociado al valor t es muy pequeño, se tiende a rechazar la hipótesis nula, asumiendo entonces, que la media del grupo es mayor a la media global, aunque dicha diferencia entre medias no es muy apreciable, el tamaño de muestra tan grande genera la significancia de estas variables.

NOTA: se aplica el mismo procedimiento de caracterización descrito anteriormente a los grupos restantes, teniendo en cuenta los grandes tamaños muestrales que presentan.

	Test.Value	Class.Mean	Frequency	Global.Mean
nv_mean	62.72	0.83	15402	0.81
perc_arci	39.74	46.03	15402	44.75
nive_frea	30.62	154.60	15402	151.94
perc_aren	29.34	18.69	15402	17.67
diam_pond	26.54	3.90	15402	3.69
bio_7	20.92	9.67	15402	9.61
bio_5	20.61	31.87	15402	31.81
bio_1	15.32	26.81	15402	26.80
bio_4	-9.27	25.26	15402	25.47
chid_mean	-19.43	4.53	15402	4.69
bio_12	-37.65	2719.30	15402	2796.1
bio_15	-39.59	36.76	15402	37.93
nv_mrange	-43.20	0.03	15402	0.03
ph	-45.30	5.57	15402	5.62
nv_range	-59.99	0.06	15402	0.06
bio_14	-69.41	71.33	15402	75.54

Tabla **8-5.**: Caracterización a partir de valores test: Cluster 1

Las unidades de manejo de esta zona se distinguen por poseer abundante vegetación, además tiene suelos con texturas arcillosas dado que están formados principalmente por arcilla (46.03 %) y bajos contenidos de arena (18,69 % en promedio) lo que los hace impermeables y difíciles de trabajar, el diámetro medio ponderado oscila entre 1.809 mm y 4.747 mm sugiriendo que la estructura del suelo es poco estable, en términos de clima la zona presenta en promedio una temperatura anual de 26.81°C aunque dicha condición es poco variable a lo largo del año según el rango de temperatura y la temperatura máxima del mes más caliente presenta un promedio de 31.87°C.

Grupo 2: en esta área la precipitación del mes más seco fluctúa entre 75.11 y 91.15 mm, su temperatura promedio es de 26.66° C al año, en cuanto a las propiedades del suelo se distingue por ser moderadamente ácidos ya que su pH promedio es de 5.69, el diámetro medio de los agregados resulto ser una variable relevante y especialmente significativa el cual repercute en la capacidad de infiltración de los suelos, en este caso el promedio es equivalente a 4.629 mm, según la literatura por ser un valor alto, el tamaño de los agregados es grande lo que produce una obstrucción en la porosidad del suelo disminuyendo su conductividad hidráulica; además posee vegetación densa ya que el promedio del índice de vegetación (nv mean) oscila entre 54.7 % y 87.54 % que es poco constante a lo largo de la región.

	Test.Value	Class.Mean	Frequency	Global.Mean
bio_14	201.12	86.40	18431	75.54
diam_pond	132.18	4.63	18431	3.69
nv_mean	81.28	0.83	18431	0.81
nive_frea	74.98	157.76	18431	151.94
ph	70.15	5.69	18431	5.62
chid_mean	-9.97	4.62	18431	4.69
perc_arci	-23.97	44.07	18431	44.75
perc_aren	-30.49	16.73	18431	17.67
nv_mrange	-60.30	0.03	18431	0.03
nv_range	-62.51	0.06	18431	0.06
bio_12	-142.42	2536.95	18431	2796.09
bio_15	-155.94	33.80	18431	37.93
bio_4	-160.62	22.21	18431	25.47
bio_1	-167.78	26.66	18431	26.80
bio_5	-193.29	31.36	18431	31.81
bio_7	-201.25	9.11	18431	9.61

Tabla 8-6.: Cluster 2

Grupo 3: dadas las condiciones más significativas, esta zona presenta precipitaciones promedio de 3213 mm que están repartidas regularmente a lo largo de todo el año, con temperaturas entre 26.69 y 27.12°C anuales, mientras que el mes más caliente alcanza una temperatura promedio de 32.24°C, son características uniformes durante el periodo evaluado debido a que los valores de sus coeficientes de variación son muy pequeños 0.236 % y 0.396 % respectivamente. De acuerdo al índice de vegetación promedio se observa que el área tiene extensa vegetación en la superficie y por ende mayor actividad agrícola además existe un nivel de disimilitud espacial en la zona, que se observa en las variables que representan variabilidad temporal (dispersión). Su suelo es de tipo arcilloso ya que en promedio la proporción de partículas minerales en el área de estudio para la arena es de 18.48 % y para la arcilla de 46.6 % por lo que carecen de buenas posibilidades de aireación, sin embargo el diámetro medio ponderado que varía entre 1.147 y 3.526 mm determina que son suelos poco estables pero con buenos movimientos de agua en la tierra debido a que el tamaño de los agregados es pequeño.

	Test.Value	Class.Mean	Frequency	Global.Mean
bio_12	242.51	3213.25	20062	2796.1
bio_15	228.98	43.68	20062	37.93
bio_4	210.03	29.49	20062	25.47
bio_7	202.07	10.08	20062	9.61
bio_5	193.95	32.24	20062	31.81
bio_1	180.36	26.93	20062	26.80
perc_arci	70.37	46.66	20062	44.75
chid_mean	48.92	5.02	20062	4.69
nv_mrange	41.25	0.03	20062	0.03
nv_range	35.61	0.07	20062	0.06
perc_aren	27.39	18.48	20062	17.67
nive_frea	-45.86	148.57	20062	151.9
nv_mean	-70.35	0.79	20062	0.81
ph	-126.14	5.50	20062	5.62
bio_14	-152.61	67.74	20062	75.54
diam_pond	-235.22	2.12	20062	3.69

Tabla 8-7.: Cluster 3

Grupo 4: se caracteriza por suelos con moderados flujos de agua y cantidades de pH entre 5.56 y 6.10 lo que implica que son ligeramente ácidos, esta es una característica de alta importancia en la calidad de los suelos ya que tiene influencia sobre otras condiciones químicas y sobre aspectos microbiológicos; el índice de vegetación promedio presenta un valor mínimo del 57.1 % y un valor máximo del 85.84 % lo que determina que la vegetación es densa, húmeda y bien desarrollada, no obstante su respectivo coeficiente de variación (4.59 %) sugiere que hay poca variabilidad, es decir que la zona presenta una topografía bastante uniforme. Respecto al clima se tiene que el comportamiento de la temperatura media anual no es muy diferente a la de los otros conglomerados varía entre 26.6°C y 27.03°C durante el periodo evaluado, con lluvias abundantes repartidas regularmente a lo largo de todo el año.

	Test.Value	Class.Mean	Frequency	Global.Mean
ph	166.38	5.81	15593	5.62
chid_mean	80.54	5.32	15593	4.69
nv_range	71.05	0.07	15593	0.06
diam_pond	64.51	4.20	15593	3.69
bio_1	62.37	26.85	15593	26.80
bio_5	40.42	31.92	15593	31.81
bio_4	37.92	26.32	15593	25.47
nv_mrange	26.33	0.03	15593	0.03
bio_7	22.77	9.67	15593	9.61
bio_15	19.58	38.51	15593	37.93
bio_12	-19.37	2756.88	15593	2796.1
nv_mean	-56.03	0.79	15593	0.81
bio_14	-68.36	71.43	15593	75.54
perc_aren	-154.85	12.34	15593	17.67
perc_arci	-163.12	39.56	15593	44.75
nive_frea	-178.66	136.51	15593	151.9

Tabla 8-8.: Cluster 4

Grupo 5: las unidades de manejo de esta zona se distinguen por suelos con texturas arcillosas lo que los hace impermeables y difíciles de trabajar, en promedio el diámetro medio ponderado es 4.039 mm aunque son suelos estables el tamaño de los agregados es grande lo que produce escasos flujos de agua en la tierra, por otra parte el nivel freático varía entre 137.1 y 177.5 cm característica que es adecuada para el cultivo que se produce en la zona.

	Test.Value	Class.Mean	Frequency	Global.Mean
perc_aren	159.29	25.19	9150	17.67
nive_frea	147.55	169.40	9150	151.9
bio_14	112.70	84.82	9150	75.54
perc_arci	89.64	48.67	9150	44.75
nv_mrange	44.32	0.03	9150	0.03
diam_pond	32.12	4.04	9150	3.69
nv_range	20.08	0.07	9150	0.06
nv_mean	-19.70	0.80	9150	0.81
bio_7	-63.04	9.37	9150	9.61
bio_12	-70.86	2599.50	9150	2796.1
ph	-72.01	5.50	9150	5.62
bio_15	-80.63	34.67	9150	37.93
bio_5	-84.09	31.51	9150	31.81
bio_4	-109.01	22.10	9150	25.47
bio_1	-120.05	26.65	9150	26.80
chid_mean	-129.43	3.29	9150	4.69

Tabla 8-9.: Cluster 5

En esta área la precipitación del mes más seco fluctúa entre 71.64 y 91.1 mm y la temperatura promedio se mantiene bajo condiciones estimadas entre los 26,65°C, la cual no difiere a la de los otros grupos.

9. Conclusiones y Recomendaciones

- La metodología de la zonificación agroecológica junto con métodos multivariantes son de gran utilidad en el desarrollo de las actividades agropuecuarias ya que permiten caracterizar e identificar los sistemas productivos en base a variables climáticas, edáficas y topográficas, el cual pueden ser implementados a cualquier tipo de cultivo.
- Los resultados obtenidos con las técnicas multivariantes, permitieron conocer aspectos generales y de poco control presentes en el Urabá Antioqueño y que influyen en la productividad y desarrollo del cultivo característico de la región el Banano.
- La implementación del análisis factorial múltiple como metodología previa a una clasificación de zonas agroecológicas de las fincas bananeras del Urabá antioqueño permitió resumir la información en tres factores que son: comportamiento del clima, propiedades físicas y químicas del suelo y la presencia de vegetación en la región, el cual explican la mayor variabilidad (78.62 %) presente en los datos. Del análisis de clúster para estas nuevas variables se logra conformar un total de 5 grupos homogéneos, en los que se encontró que la variabilidad de las características climáticas es poco apreciable a lo largo de la región ya que predominó una temperatura media aproximada a 26°C con variaciones inferiores a 1°C y precipitaciones alrededor de los 2459 y 3547 mm al año.
- En cada grupo se observó abundante vegetación condición que es poco variable debido a que esa área está dedicada principalmente a la producción del banano por ende que la cubierta vegetal sea densa; sin embargo las variables referentes al suelo muestran tener mayor variabilidad dentro de los grupos, además sus características difieren entre cada clase, por lo que son consideradas muy influyentes en la clasificación de las unidades de manejo, por otra parte se encontró que la condición más distintiva en toda la zona es la acidez de los suelos debido a que su pH varía aproximadamente entre 5 y 6, característica que hace del Urabá una región ideal para la producción del banano, además son suelos con buena capacidad de almacenamiento de agua y nutrientes por los altos contenidos de arcilla. De forma general se puede evidenciar que las clases son conformadas a partir de su distancia espacial debido a que las características pueden ser muy similares por su cercanía.
- Debido a la cantidad de datos con los que se llevó a cabo el procedimiento para la caracterización de las zonas, se notó que la clasificación de los individuos es muy sensible respecto a la selección del número de conglomerados, por tanto se hizo necesario recurrir

a diferentes métodos con el fin de elegir el número adecuado de grupos que establezca la tipificación de los individuos.

- Aunque la variabilidad dentro de la región no es muy alta para las variables estudiadas, en el agrupamiento de los individuos se encontró sectores con similitudes agroecológicas notables.
- Las variables empleadas fueron ideales para la caracterización debido a su influencia sobre el cultivo predominante en la región, sin embargo considerar otras variables de productividad y manejo, tal como el rendimiento por árbol anual, grupos varietales, cultivo asociado, entre otras, sería de gran interés para encontrar posibles diferencias dentro de los grupos edafoclimáticos, aunque hasta la fecha han sido difíciles de involucrar en el análisis.

A. Anexo: Descripción de las variables

Variables	Descripción
Bio 1	Temp. media anual (°C)
Bio 2	Rango de temperatura diurno medio (°C)
Bio 3	Isotermalidad (Bio 2/Bio 7)
Bio 4	Estacionalidad de temperatura (desviación estandar*100) (°C)
Bio 5	Temperatura máxima del mes mas caliente (°C)
Bio 6	Temperatura mínima del mes mas frío (°C)
Bio 7	Rango de temperatura anual (°C)
Bio 8	Temperatura media del trimestre mas humedo (°C)
Bio 9	Temperatura media del trimestre mas seco (°C)
Bio 10	Temperatura media del trimestre mas caliente (°C)
Bio 11	Temperatura media del trimestre mas frio (°C)
Bio 12	Precipitación total anual (mm)
Bio 13	Precipitación del mes mas húmedo (mm)
Bio 14	Precipitación del mes mas seco (mm)
Bio 15	Estacionalidad de precipitación (CV) (%)
Bio 16	Precipitación del trimestre mas húmedo (mm)
Bio 17	Precipitación del mes mas seco (mm)
Bio 18	Precipitación del trimestre mas caliente (mm)
Bio 19	Precipitación del trimestre mas frío (mm)

Tabla A-1.: Descripción de variables de clima 1950-2000

Variables	Descripción
nv_max	máximo histórico período
nv_mean	promedio histórico período
nv_min	mínimo histórico período
nv_mrange	rango medio anual período
nv_range	rango total histórico período
nv_stdv	desviación estándar período

Tabla **A-2.**: Descripción de variables del NDVI 2000-2009

Variables	Descripción
chid_barr	Cond. hidr. pozo barrenado (cm/h)
chid_inve	Cond. Hidr. Pozo invertido (cm/h)
chid_labo	Cond. Hidr. Método laboratorio (cm/h)
dens_real	Densidad real (g/cm ³)
diam_pond	Diametro medio ponderado (mm)
esta_agre	Estado de agregación
hgra_ccam	Humedad gravimetrica saturación (%)
nive_frea	Nivel freático (cm)
perc_arci	Arcilla (%)
perc_aren	Arena (%)
perc_limo	Limo (%)
ph	Potencial de hidrógeno

Tabla **A-3.**: Descripción de variables del Suelo 2000-2009

B. Anexo: Análisis de correlación

	nv_max	nv_mean	nv_min	nv_mrange	nv_range	nv_stdv
nv_max	1	0.984	0.954	-0.205	-0.158	-0.172
nv_mean	0.984	1	0.985	-0.327	-0.307	-0.312
nv_min	0.954	0.985	1	-0.416	-0.447	-0.439
nv_mrange	-0.205	-0.327	-0.416	1	0.760	0.794
nv_range	-0.158	-0.307	-0.447	0.760	1	0.932
nv_stdv	-0.172	-0.312	-0.439	0.794	0.932	1

Tabla **B-1.**: Matriz de correlación para el NDVI

	chid_barr	chid_inve	chid_labo	dens_real	diam_pond	esta_agre	hgra_ccam	nive_frea	perc_arci	perc_aren	perc_limo	ph
chid_barr	1	-0.71	0.51	-0.39	-0.07	0.73	0.77	-0.77	-0.69	-0.83	-0.69	0.37
chid_inve	-0.71	1	-0.93	0.31	0.04	-0.72	-0.95	0.30	0.68	0.69	0.68	0.14
chid_labo	0.51	-0.93	1	-0.42	0.23	0.73	0.91	-0.03	-0.71	-0.61	-0.71	-0.18
dens_real	-0.39	0.31	-0.42	1	-0.77	-0.77	-0.55	0.36	0.86	0.73	0.86	-0.65
diam_pond	-0.07	0.04	0.23	-0.77	1	0.49	0.20	0.16	-0.51	-0.33	-0.51	0.49
esta_agre	0.73	-0.72	0.73	-0.77	0.49	1	0.91	-0.51	-0.93	-0.97	-0.93	0.54
hgra_ccam	0.77	-0.95	0.91	-0.55	0.20	0.91	1	-0.42	-0.85	-0.87	-0.85	0.17
nive_frea	-0.77	0.30	-0.03	0.36	0.16	-0.51	-0.42	1	0.48	0.72	0.48	-0.67
perc_arci	-0.69	0.68	-0.71	0.86	-0.51	-0.93	-0.85	0.48	1	0.90	1.00	-0.49
perc_aren	-0.83	0.69	-0.61	0.73	-0.33	-0.97	-0.87	0.72	0.90	1	0.90	-0.62
perc_limo	-0.69	0.68	-0.71	0.86	-0.51	-0.93	-0.85	0.48	1	0.90	1	-0.49
ph	0.37	0.14	-0.18	-0.65	0.49	0.54	0.17	-0.67	-0.49	-0.62	-0.49	1

Tabla **B-2.**: Matriz de correlación para variables de Suelo

	bio_1	bio_2	bio_3	bio_4	bio_5	bio_6	bio_7	bio_8	bio_9	bio_10	bio_11	bio_12	bio_13	bio_14	bio_15	bio_16	bio_17	bio_18	bio_19
bio_1	1	0.76	-0.87	0.89	0.93	0.004	0.87	0.88	0.95	0.95	0.91	0.83	0.85	-0.89	0.87	0.82	-0.89	-0.30	0.84
bio_2	0.76	1	-0.69	0.79	0.89	-0.47	0.93	0.72	0.82	0.83	0.72	0.87	0.82	-0.77	0.83	0.85	-0.73	-0.08	0.81
bio_3	-0.87	-0.69	1	-0.9	-0.89	0.24	-0.89	-0.78	-0.86	-0.89	-0.79	-0.79	-0.79	0.91	-0.85	-0.76	0.89	0.37	-0.81
bio_4	0.89	0.79	-0.90	1	0.92	-0.23	0.91	0.8	0.91	0.94	0.77	0.90	0.89	-0.875	0.94	0.89	-0.95	-0.44	0.89
bio_5	0.93	0.89	-0.89	0.92	1	-0.21	0.97	0.86	0.957	0.97	0.88	0.89	0.87	-0.92	0.91	0.87	-0.9	-0.25	0.87
bio_6	0.004	-0.46	0.24	-0.23	-0.21	1	-0.4	-0.02	-0.05	-0.13	0.08	-0.35	-0.28	0.16	-0.29	-0.34	0.14	-0.03	-0.31
bio_7	0.87	0.93	-0.89	0.91	0.97	-0.4	1	0.81	0.90	0.93	0.81	0.90	0.87	-0.89	0.91	0.88	-0.87	-0.23	0.88
bio_8	0.88	0.72	-0.77	0.8	0.86	-0.02	0.81	1	0.86	0.87	0.88	0.76	0.79	-0.81	0.77	0.74	-0.77	-0.15	0.80
bio_9	0.95	0.81	-0.86	0.91	0.95	-0.05	0.90	0.86	1	0.97	0.90	0.87	0.86	-0.88	0.9	0.85	-0.9	-0.30	0.86
bio_10	0.95	0.83	-0.89	0.94	0.97	-0.13	0.93	0.87	0.97	1	0.90	0.87	0.87	-0.92	0.91	0.85	-0.91	-0.28	0.88
bio_11	0.91	0.72	-0.79	0.77	0.88	0.08	0.81	0.88	0.90	0.90	1	0.71	0.72	-0.87	0.74	0.68	-0.77	-0.11	0.74
bio_12	0.83	0.86	-0.78	0.90	0.89	-0.35	0.90	0.76	0.87	0.87	0.71	1	0.98	-0.74	0.97	0.99	-0.89	-0.39	0.96
bio_13	0.85	0.81	-0.79	0.89	0.87	-0.29	0.87	0.79	0.86	0.87	0.72	0.98	1	-0.75	0.96	0.98	-0.89	-0.40	0.97
bio_14	-0.89	-0.77	0.91	-0.88	-0.92	0.16	-0.9	-0.81	-0.88	-0.92	-0.87	-0.75	-0.75	1	-0.81	-0.71	0.86	0.19	-0.77
bio_15	0.87	0.83	-0.85	0.94	0.91	-0.29	0.91	0.77	0.9	0.91	0.74	0.97	0.96	-0.81	1	0.97	-0.95	-0.50	0.94
bio_16	0.82	0.85	-0.76	0.89	0.87	-0.35	0.88	0.74	0.85	0.85	0.68	0.99	0.98	-0.71	0.97	1	-0.89	-0.45	0.95
bio_17	-0.89	-0.72	0.88	-0.95	-0.89	0.14	-0.87	-0.77	-0.90	-0.91	-0.77	-0.89	-0.88	0.86	-0.95	-0.88	1	0.61	-0.87
bio_18	-0.3	-0.08	0.37	-0.44	-0.25	-0.04	-0.24	-0.15	-0.30	-0.28	-0.11	-0.39	-0.40	0.19	-0.5	-0.45	0.61	1	-0.32
bio_19	0.84	0.81	-0.81	0.89	0.87	-0.31	0.88	0.80	0.86	0.88	0.74	0.96	0.97	-0.77	0.94	0.95	-0.87	-0.32	1

Tabla B-3.: Matriz de correlación variables climáticas

C. Anexo: Análisis descriptivos para los grupos según variables edafoclimáticas

	Min.	Mean	Max.	Desv.Est	CV
bio_1	26.69	26.81	26.91	0.039	0.0015
bio_5	31.63	31.87	32.12	0.11	0.0035
bio_7	9.375	9.666	9.922	0.117	0.0121
nv_mean	0.521	0.827	0.894	0.025	0.0307
diam_pond	1.809	3.903	4.747	0.498	0.1275
nive_frea	134.7	154.6	172.6	7.182	0.0465
perc_arci	39.71	46.03	56.98	2.001	0.0435
perc_aren	13.37	18.69	29.88	2.552	0.1365

Tabla C-1.: Análisis descriptivo grupo edafoclimático 1

	Min.	Mean	Max.	Desv.Est	CV
bio_14	75.11	86.4	91.15	3.3964	0.0393
nv_mean	0.547	0.8303	0.8754	0.033	0.0401
diam_pond	3.9	4.629	5.526	0.2662	0.0575
nive_frea	137.4	157.8	171.9	6.1498	0.0390
ph	5.445	5.69	5.94	0.1054	0.0185

Tabla C-2.: Análisis descriptivo grupo edafoclimático 2

	Min.	Mean	Max.	Desv.Est	CV
bio_1	26.69	26.93	27.12	0.0635	0.0024
bio_4	25.7	29.49	31.85	1.1103	0.0376
bio_5	31.8	32.24	32.54	0.1275	0.0040
bio_7	9.553	10.082	10.336	0.1326	0.0131
bio_12	2877	3213	3547	170.582	0.0531
bio_15	38.95	43.68	48.42	2.0411	0.0467
nv_mrange	0.014	0.0292	0.067	0.0059	0.2033
nv_range	0.034	0.0655	0.1574	0.014	0.2211
chid_mean	-3.455	5.016	7.05	1.0384	0.2070
perc_arci	38.29	46.66	74.28	3.3090	0.0709
perc_aren	10.97	18.48	42.79	3.9734	0.2150

Tabla C-3.: Análisis descriptivo grupo edafoclimático 3

	Min.	Mean	Max.	Desv.Est	CV
bio_1	26.6	26.85	27.03	0.0557	0.0021
bio_4	21.58	26.32	31.3	2.0398	0.0775
bio_5	31.28	31.92	32.35	0.1770	0.0055
bio_7	8.905	9.671	10.156	0.1961	0.0203
bio_15	33.73	38.51	44.79	2.5168	0.0654
nv_range	0.024	0.07	0.1722	0.0190	0.2716
chid_mean	3.247	5.322	6.509	0.4357	0.0819
diam_pond	2.788	4.201	5.282	0.4089	0.0973
ph	5.564	5.808	6.107	0.0986	0.0170

Tabla C-4.: Análisis descriptivo grupo edafoclimático 4

	Min.	Mean	Max.	Desv.Est	CV
bio_14	71.64	84.82	91.1	4.1457	0.0489
nv_range	0.0299	0.065	0.1809	0.0152	0.2335
diam_pond	-1.369	4.039	4.92	0.8503	0.2105
nive_frea	137.1	169.4	177.5	4.8480	0.0286
perc_arci	44.31	48.67	95.49	6.5234	0.1340
perc_aren	17.9	25.19	59.25	5.1134	0.2030

Tabla C-5.: Análisis descriptivo grupo edafoclimático 5

Bibliografía

- Anselin, L., B. R. & Lowengberg-DeBoer, J. (2004), 'A spatial econometric approach to the economics of site-specific nitrogen management in corn production', *American Journal of Agricultural Economics* pp. 671–683.
- AUGURA (2012), Coyuntura bananera colombiana, Technical report.
- Bécue, M. & Pagés, J. (2008), 'Multiple factor analysis and clustering of a mixture of quantitative, categorical and frequency data', *Computational Statistics and Data Analysis* 52(6) p. 3255–3268.
- Bonilla, J., P. J. U. M. & E., C. (2000), 'Informe nacional sobre el uso y manejo de plaguicidas en Colombia, tendiente a identificar y proponer alternativas para reducir el escurrimiento de plaguicidas al mar Caribe', *Global Environment Facility* . MINISTERIO DEL MEDIO AMBIENTE. PROYECTO PNUMA//UCR/CAR.
- Carbonell, J., A. A. O. B. T. J. Q. R. & Isaacs, E. (2001), 'Zonificación agroecológica para el cultivo de caña de azúcar en el valle del río Cauca. tercera aproximación', *Canicaña, Cali*. **29 edn.**
- Casa, R. & Castrignanó, A. (2008), 'Analysis of spatial relationships between soil and crop variables in a durum wheat field using a multivariate geostatistical approach', *European Journal of Agronomy* pp. 331–342.
- Cock, J. (2007), 'Sharing commercial information in: Innovation workshop for the agricultural sector: Site specific agriculture based on sharing farmers experiences', *Innovation Workshop for the Agricultural Sector: Site Specific Agriculture based on Sharing Farmers Experiences* .
- Cock, J., I. D. & Estrada, M. (2006), 'Guía práctica para la caracterización del suelo y del terreno rasta', *Biometrika* **2**.
- Cortes, A. (2010), Análisis de la variabilidad espacial y temporal de la precipitación en una ciudad de media montaña andina, caso de estudio: Manizales., PhD thesis, Universidad Nacional.
- Díaz, L. G. (2007), *Estadística Multivariada: Inferencia y Métodos*, Vol. 2 edición.
- Dorado, H. (2013), 'Análisis multivariado para la caracterización de factores edafoclimáticos y agrupamiento de sitios con presencia de plátano', *Universidad del Valle* .

- Escofier, B. & Page, J. (1992), 'Método factoriales simples y múltiples. objetivos, métodos e interpretación', *Balboa edn. Universidad del País Vasco* pp. 507–538.
- Eslava, J. (1993), 'Climatología y diversidad climática de colombia', *Rev. Acad. Colomb. Cienc.* **18**, 507–538.
- Espinal, C., M. H. & Peña, Y. (2005), 'La cadena del banano en colombia una mirada global de su estructura y dinamica 1991-2005,', *Ministerio de Agricultura y Desarrollo Rural Observatorio Agro cadenas Colombia*, .
- FAO (2005), Food and agriculture organization of united nations, Technical report. Estadísticas.
- Gabriels, D., L. D. & Pinto, J. (2011), 'Métodos para determinar la conductividad hidráulica saturada y no saturada de los suelos'.
- Gallardo, D. (2011a), 'Ampliación de análisis de datos multivariantes', *Departamento de Estadística e Investigación Operativa*. .
- Gallardo, J. (2011b), 'Métodos jerárquicos de análisis clúster: Capítulo 3'. Curso de diplomatura estadística teórico práctico, Universidad de Granada.
- Gildardo, E., P. C. G. R. & Martín, J. (2006), 'Manejo sostenible del cultivo del plátano', *Corporación Colombiana de Investigación Agropecuaria* .
- Guarín, G. (2011), 'Impacto de la variabilidad climática en la producción de banano en el urabá antioqueño'.
- Hijmans, R., C. S. P. J. J. P. & Jarvis, A. (2005), 'Very high resolution interpolated climate surfaces for global land areas', *International Journal of Climatology* .
- Hutchinson, M. (1984), 'A summary of some surface fitting and contouring programs for noisy data', *CSIRO Division of Mathematics and Statistics* . Consulting Report ACT 84/6. Canberra, Australia.
- IGAC (2007), Antioquia: Características geográficas, Technical report.
- IGAC (2012), Hacia un dialogo entre regiones, Technical report. Encuentro Internacional sobre Asociatividad Regional y Ordenamiento Territorial Bogota D.C.
- Jiménez, D., C. J. J. A. S. H. V. D. P. U. & Barreto, M. (2011), 'Interpretation of commercial production information: A case study of lulo, an underresearched andean fruit', *Elsevier 104(1)* pp. 258–270.
- Lebart, L., M. A. & Piron, M. (1995), *Statistique exploratoire multidimensionnelle: visualisation et inférence en fouilles de données.*, Vol. 2 edn, Dunod.

- Leiva, F. (2003), 'Aproximación al manejo sitio-específico del suelo para una agricultura sostenible y competitiva en colombia', *Journal of the American Statistical Association*. .
- Meneses, C. (2011), 'El índice normalizado diferencial de la vegetación como indicador de la degradación del bosque'.
- Moreira, A. (1996), 'Los sistemas de información geográfica y sus aplicaciones en la conservación de la diversidad biológica', *Ambiente y desarrollo* pp. 80–86.
- Mundial, B. (2008), Informe sobre el desarrollo mundial 2008. agricultura para el desarrollo. coeditado con Mundi-Prensa y Mayol Ediciones, S.A. Colombia.
- OMM (1996), Vocabulario meteorológico internacional. Publicación No. 182. Ginebra, Suiza.
- Pabón J., Zea J., L. G. H. G. G. O. & J., M. (2002), 'La atmósfera, el tiempo y el clima data', *Instituto de Hidrología, Meteorología y Estudios Ambientales* . IDEAM.
- Pardo, C. & Del Campo, P. (2007), 'Combinación de métodos factoriales y de análisis de conglomerados en r: el paquete factoclass', *Revista Colombiana de Estadística* **30**, 231–245.
- Peña, D. (2002), *Análisis de Datos Multivariantes*, Mc Graw Hill, Interamericana de España.
- Pérez, C. (2007), 'Minería de datos. técnicas y herramientas', *Thomson* .
- Rodríguez, J. & Rodríguez, A. (2002), 'Aspectos socioeconómicos del cultivo del plátano en colombia', *Infomusa* 10(1) .
- Ríos, G., R. M. B. M. F. G. P. J. M. J. G. J. & Echeverry, D. (2004), 'Zonificación, caracterización y tipificación de los sistemas de producción de lulo (*solanum quitoense* lam) en el eje cafetero', *REVISTA CORPOICA* .
- Ruiz, N., M. G. T. C. & Suaza, S. (2011), 'El urabá antioqueño: un mar de oportunidades y potencialidades', *Gobernación de Antioquia* . Departamento Administrativo de Planeación.
- SAC (2011), Sector agroindustrial colombiano, Technical report.
- Teran, C., J. C. G. C. & Villaneda, E. (1998), 'Metodología para la zonificación agroclimática de la región de la mojana mediante el sistema de información geográfica arc/info', *Revista Corpoica* pp. 1–8.
- Valencia, A. (2013), 'Análisis multivariado para la caracterización de unidades de manejo o lotes de aguacate según variables edafoclimáticas relacionadas con la productividad, en ejercicio de la agricultura específica por sitio en colombia.'. Universidad del Valle.