

APLICACIÓN DE MÉTODOS MULTIVARIADOS PARA LA CARACTERIZACIÓN
BIOLÓGICA Y FÍSICO -QUÍMICA DE SISTEMAS DE PRODUCCIÓN DEL SUELO
EN CAQUETÁ

CYNTHIA FEJOO ROSERO

UNIVERSIDAD DEL VALLE – FACULTAD DE INGENIERÍA
ESCUELA DE ESTADÍSTICA
PROGRAMA ACADÉMICO DE ESTADÍSTICA
SANTIAGO DE CALI
AGOSTO 2013

APLICACIÓN DE MÉTODOS MULTIVARIADOS PARA LA CARACTERIZACIÓN
BIOLÓGICA Y FÍSICO -QUÍMICA DE SISTEMAS DE PRODUCCIÓN DEL SUELO
EN CAQUETÁ

CYNTHIA FEJOO ROSERO

Director:
Jaime Mosquera Restrepo
Estadístico

UNIVERSIDAD DEL VALLE – FACULTAD DE INGENIERÍA
ESCUELA DE ESTADÍSTICA
PROGRAMA ACADÉMICO DE ESTADÍSTICA
SANTIAGO DE CALI
AGOSTO 2013

CONTENIDO

	Pág.
INTRODUCCIÓN.....	8
1. FORMULACIÓN DEL PROBLEMA.....	10
1.1 DEFINICIÓN DEL PROBLEMA.....	10
1.2 OBJETIVOS.....	12
1.3 JUSTIFICACIÓN.....	13
2. ANTECEDENTES.....	14
2.1 ESTUDIOS BIOLÓGICOS.....	14
2.2 REFERENCIAS DE METODOLOGÍA ESTADÍSTICA.....	15
3. MARCO TEÓRICO.....	17
3.1 TÉRMINOS BIOLÓGICOS.....	17
3.1.1 Biodiversidad.....	17
3.1.2 Deforestación.....	17
3.1.3 Macroinvertebrados del suelo.....	18
3.1.4 El suelo.....	18
3.1.5 Sistemas de producción.....	18
3.1.6 Usos del suelo.....	19
3.2 MÉTODOS ESTADÍSTICOS.....	19
3.2.1 Análisis de experimentos.....	20
3.2.2 Análisis de varianza (ANOVA).....	21
3.2.3 Diseño de bloques completamente al azar con un solo factor.....	23
3.2.4 Alternativa no paramétrica para el ANOVA.....	25
3.2.5 Análisis Multivariante.....	28
3.2.5.1 Análisis de componentes principales (ACP).....	29
3.2.5.2 Análisis de correlación canónico (ACC).....	35

4. METODOLOGÍA.....	40
4.1 ÁREA DE ESTUDIO.....	40
4.2 RECOLECCIÓN DE LOS DATOS.....	41
4.3 ANÁLISIS ESTADÍSTICO DE LOS DATOS.....	42
4.3.1 Análisis exploratorio de datos.....	41
4.3.2 Análisis multivariado de datos.....	43
5. RESULTADOS.....	45
5.1 ANÁLISIS EXPLORATORIO DE DATOS.....	45
5.1.1 Análisis de variables físico – químicas.....	46
5.1.2 Análisis de varianza y validación de supuestos.....	49
5.2 ANÁLISIS MULTIVARIADO DE DATOS.....	50
5.2.1 Análisis de Componentes Principales.....	50
5.2.2 Análisis de Correlación Canónica.....	56
6. CONCLUSIONES.....	59
7. RECOMENDACIONES.....	60
BIBLIOGRAFÍA.....	61

LISTA DE TABLAS

	Pág.
Tabla 1. Análisis de varianza.....	21
Tabla 2. Anova de bloques.....	22
Tabla 3. Estructura de los datos.....	24
Tabla 4. Variables de estudio.....	36
Tabla 5. Frecuencia de usos del suelo en los sistemas de producción.....	39
Tabla 6. Valores p para la validación de supuestos y pruebas no paramétricas de variables con una profundidad.....	43
Tabla 7. Valores p para las pruebas post hoc.....	44
Tabla 8. Valores p para la validación de supuestos y pruebas no paramétricas de variables con 3 profundidades.....	45
Tabla 9. Valores p para prueba no paramétrica de la macrofauna.....	45
Tabla 10. Matriz de Coeficientes de las Componentes.....	48

LISTA DE FIGURAS

	Pág.
Figura 1. Modelo general de un proceso o sistema.....	18
Figura 2. Conjunto de datos.....	32
Figura 3. Mapa del Departamento de Caquetá.....	35
Figura 4. Estructura del conjunto de datos.....	38
Figura 5. Boxplot de las variables físico químicas medidas en tres profundidades.....	40
Figura 6. Boxplot variables físico – químicas al nivel de suelo.....	42
Figura 7. Valores propios del ACP.....	47
Figura 8. Círculo de correlaciones.....	47
Figura 9. Planos factoriales del ACP.....	50
Figura 10. Valores propios del ACC.....	50
Figura 11. Análisis de Correlación Canónica.....	53

Resumen

Este trabajo describe el efecto de los sistemas de producción y usos de suelo sobre la biodiversidad de la macrofauna y las propiedades físico – químicas del suelo en tres sistemas de producción de la tierra en municipios del Caquetá.

Se describe la relación entre factores ambientales y la composición físico – química del suelo. Los resultados muestran mayor riqueza de especies de macrofauna y estructura del suelo en el sistema agroforestal, un suelo mejor conservado. Mediante herramientas estadísticas de Análisis de Correlación Canónico, se evidenció que la abundancia de las poblaciones de macrofauna está influenciada por las condiciones del suelo.

INTRODUCCIÓN

En Colombia, el Amazonas hace parte de una de las regiones naturales, ocupando un área de 403.000 kilómetros cuadrados, equivalente al 35% del territorio nacional, ubicada entre los departamentos del Amazonas, Caquetá, Putumayo, Guainía y Vaupés. Dada su inigualable abundancia de especies vegetales y animales, el mundo se ha hecho consciente de su importancia ecológica para el futuro del planeta. Sin embargo, la perturbación de las selvas tropicales para establecer sistemas de producción intensivos, ha modificado muchos ciclos químicos y físicos del suelo, lo que ha deteriorado su cobertura vegetal y ha reducido la diversidad en muchas áreas (Hoffmann y Greef, 2003). La transformación de cobertura arbórea nativa a extensas áreas de pasturas, genera impactos ambientales negativos como erosión y compactación del suelo (Laurance, 2000). Estos impactos en la estructura y continuidad del paisaje afecta negativamente el movimiento de especies especialistas del bosque que tienen alta sensibilidad a cambios en las condiciones microambientales¹ (Lozano-Zambrano et al., 2009).

En la Amazonía la deforestación es la principal causa de pérdida de biodiversidad (Laurence et al. 2004). El piedemonte Amazónico colombiano ha sido afectado por la transformación de sus bosques para el establecimiento de sistemas productivos pecuarios (Maldonado y Velásquez 1994). En la actualidad, las condiciones económicas y sociológicas de la región determinan las prácticas de uso del suelo, los sistemas de producción y la biodiversidad, generando un complejo mosaico de fragmentos y parches de vegetación (Velásquez-Valencia 2009).

La mayoría de los organismos del suelo dependen para su funcionamiento de factores edáficos como la humedad, la temperatura, la aireación, la acidez y el estado nutricional, así como del suministro de energía. Los mismos factores influyen en la distribución espacial de los organismos. Entre los principales organismos del suelo, se mencionan: los transformadores de compuestos nitrogenados, los oxidantes minerales, los fijadores de nitrógeno, los descomponedores, los predadores de la fauna, los fitoparásitos y los mezcladores de detritos orgánicos. Lo anterior, motivó a que el grupo de investigación Gestión en Agroecosistemas Tropicales Andinos (GATA) de la Universidad Tecnológica de Pereira (UTP), planteara el proyecto Amaz BD: “Biodiversidad de los paisajes amazónicos, determinantes socio-económicos y producción de bienes y servicios”, realizado en el Departamento de Caquetá, ubicado al noroeste de la región Amazónica, con el objetivo de analizar y modelar las relaciones que unen el ambiente socioeconómico; la composición y estructura del paisaje; la biodiversidad de plantas, aves y macrofauna del suelo; y describir las relaciones entre producción de cultivos y otros bienes y servicios del ecosistema (definido por Daily et al., 1997) usados por poblaciones humanas. El grupo de Amaz BD estuvo a cargo de la planeación y ejecución de la recolección de los datos de acuerdo con el método de TSBF (Tropical Soil Biology and Fertility; Lavelle 1988), en el cual se evaluaron propiedades físicas y químicas del suelo y se tomó información de la cantidad de especies de macrofauna en los 135

¹ Ambiente pequeño, con las condiciones apropiadas para el desarrollo de un ser vivo.

puntos de muestreos. En el presente estudio, se tomó sólo el capítulo concerniente al análisis de la macrofauna del suelo, el cual contiene dos grupos de variables: biológicas, de tipo discreto, las cuales representan conteos de la macrofauna encontrada en cada punto de muestreo y físico – químicas que indican mediciones de las propiedades del suelo, son de tipo continuas, entre las cuales se encuentran el porcentaje de carbono, la humedad, la densidad aparente, entre otras. Se tiene como objetivo evaluar la composición biológica y físico – química del suelo y determinar si esta composición se ve afectada por el sistema de producción empleado y por el uso del suelo. Para ello, se cuentan con observaciones de tres sistemas de producción y siete diferentes usos del suelo, además, la aplicación de técnicas estadísticas multivariadas como el análisis de correspondencias canónico que permite evaluar los conjuntos de datos y determinar su relación.

1. FORMULACIÓN DEL PROBLEMA

1.1 DEFINICIÓN DEL PROBLEMA

La Selva Amazónica con su enorme superficie y complejidad ecosistémica es la más importante reserva biótica² existente en el mundo. Su base natural está conformada por numerosos ecosistemas disímiles que interactúan entre sí, que presentan una gran diversidad de flora y fauna y un alto grado de endemismos³. La Amazonía colombiana se encuentra relativamente bien conservada, por ello Colombia es reconocido como un país megadiverso al poseer el 10% de la biodiversidad terrestre mundial en tan sólo 0,7% del área. La gran región amazónica constituye el 45% del bosque tropical en el mundo (Malhi *et al.* 1999), y en Colombia, los departamentos de Amazonas, Caquetá y Putumayo, representan el 7% de la Amazonía total.

Otra dimensión no menos importante es la oferta hídrica de la Selva Amazónica, una de las mayores fuentes de agua dulce del planeta y reguladora del clima mundial. Los numerosos cuerpos de agua constituyen una de las mayores riquezas de la región, ya que allí se encuentran las cuencas del Amazonas y del Orinoco. La existencia de una estación seca y otra de lluvias, produce un gran descenso de las aguas y un gran ascenso que desborda sus riberas inundando grandes extensiones de bosque. Esta dinámica del sistema hídrico, genera complejos ecosistemas, hábitats estacionales y cadenas alimenticias que dan sustento a una biota increíblemente diversa.

El conjunto de estas características, ha permitido la existencia de una gran biodiversidad en las selvas amazónicas. Sin embargo, sus suelos son muy pobres en nutrientes, lo que sumado a las fuertes lluvias y a las altas temperaturas, le confieren también una gran fragilidad. El rompimiento de la dinámica natural de los ecosistemas para actividades agropecuarias, de extracción de maderas o de explotación minera de manera intensiva, causa la compactación, erosión y contaminación de los suelos, dificultando o evitando la regeneración de la vegetación y por lo tanto de los ecosistemas. Es por ello, es importante realizar estudios que permitan detectar el efecto del manejo de los suelos destinados a la ganadería sobre la biodiversidad, pues con la pérdida de macrofauna asociada al suelo, se afectan los procesos de degradación y mantenimiento estructural del mismo (Lavelle 1997; Blanchart *et al.* 1997; Chauvel *et al.* 1999), trayendo como consecuencia la degradación del suelo y afectando su calidad.

El proyecto Amaz BD: “Biodiversidad de los paisajes amazónicos, determinantes socio-económicos y producción de bienes y servicios” de la Universidad Tecnológica de Pereira, estudia varios factores relacionados con la biodiversidad y otros factores del piedemonte Amazónico del Departamento de Caquetá, el cual contiene un módulo de macrofauna y las

² Aquello que resulta característico de los organismos vivientes o que mantiene un vínculo con ellos. También es lo que pertenece o se asocia a la biota, un concepto que permite nombrar a la fauna y la flora de un cierto territorio.

³ Especie que sólo puede encontrarse naturalmente en un lugar.

propiedades del suelo, es decir, variables biológicas (conteos de especies) y físico – químicas (variables continuas), cuya relación se evaluará en el presente documento con el fin de analizar la composición biológica de la macrofauna y físico – química del suelo y su relación con los sistemas de producción, aportando nuevos elementos de análisis al estudio de la macrofauna. El trabajo de campo fue apoyado por la comunidad del municipio de Florencia y el equipo de Amaz BD, en el que se obtuvieron 135 puntos de muestreo, cada punto de muestreo se seleccionó de acuerdo al protocolo TSBF (Tropical Soil Biology and Fertility; Anderson e Ingram, 1993), el cual consiste en un monolito⁴ central (25 cm de lado por 20 cm de profundidad) de donde se extraen manualmente los macroinvertebrados e igualmente se miden las propiedades físicas y químicas del suelo. En la región del piedemonte se distinguen tres sistemas de producción denominados: Agroforestal, Silvopastoril y Tradicional. Los sistemas se caracterizan por la predominancia de algunos usos del suelo que a su vez se clasifican en siete tipos (Pastizal Mejorado, Pastizal Degradado, Pastizal con Árboles Dispersos, Rastrojo, Sistema Agroforestal, Cultivo de Caucho y Bosque Intervenido), por lo anterior, el interés de análisis es observar si hay relación entre el uso del suelo, la abundancia de especies de macrofauna y la estructura físico – química del suelo.

⁴ Columna o elemento arquitectónico tallados en un solo bloque de piedra.

1.2 OBJETIVOS

Objetivo General

Evaluar el efecto de los sistemas de producción y usos del suelo sobre la biodiversidad de la macrofauna y la estructura físico – química de suelos pertenecientes a Florencia, Caquetá.

Objetivos Específicos

- Evaluar el efecto del uso del suelo y los sistemas de producción sobre la composición físico – química del suelo.
- Evaluar el efecto del uso del suelo y los sistemas de producción sobre la relación existente entre la macrofauna y la composición físico – química del suelo.

1.3 JUSTIFICACIÓN

La Amazonía Colombiana es la principal fuente de recursos hídricos y forestales a nivel mundial, debido a su biodiversidad y riqueza del suelo, altamente influenciada por la abundancia y diversidad de macrofauna, quienes son importantes componentes del terreno debido a su alta capacidad en la formación de nutrientes para el agroecosistema. De este modo, los suelos suministran considerables servicios ecosistémicos⁵ de mucha importancia para el mantenimiento de los ecosistemas y de las sociedades humanas (Millennium Ecosystem Assessment, 1999). También, el valor cultural de los suelos como fuente de diversos materiales (tintas, arcilla) es otro componente poco valorado, pero de alta importancia.

En un escenario optimista, los sistemas de producción deben estar basados en métodos que ayuden a conservar la estructura natural del suelo, fortaleciendo el buen desarrollo de las plantas y así, su resistencia natural a plagas y enfermedades, potenciando al máximo el cuidado mediante adecuadas prácticas de cultivo que aseguren una buena salud del suelo y por tanto, que sea más resistente (Guía de la agricultura ecológica en la comunidad valenciana, CERA 1999).

En Colombia, se han llevado a cabo algunos estudios como el realizado por Velásquez (2004), que explora las relaciones entre el uso de la tierra y la macrofauna con el fin de describir y predecir las condiciones de los agroecosistemas con sistemas de producción diversos. Es importante analizar el efecto de los usos del suelo y los sistemas de producción sobre variables cuantitativas agrupadas en biológicas (conteos de especies) y físico – químicas (variables continuas), las cuales fueron medidas en 135 puntos de muestreo, haciendo necesaria la aplicación de técnicas multivariadas, que examinen tal efecto, a través del análisis de relación. Además, varios estudios muestran que la estructura de la macrofauna, es más compleja en suelos mejor conservados o estructurados, propiciando, entre otros atributos, mayor diversidad y abundancia de macrofauna (Sevilla et al., 2002), indicando de esta manera, que tierras de buena calidad son aquellas que permiten maximizar la producción y minimizar la erosión (Doran y Parkin, 1994). Es por tanto, conveniente exponer el estudio sobre los sistemas actuales de producción y proporcionar a los sectores interesados, las características fundamentales del suelo bajo diferentes sistemas de producción, ya que con la pérdida de la macrofauna del suelo, se afectan los procesos de degradación y mantenimiento estructural del mismo (Lavelle 1997; Blanchart et al. 1997; Chauvel et al. 1999) que afectará al suelo para las diferentes actividades productivas.

⁵ Bienes y servicios de los que gozan las sociedades humanas que dependen totalmente de transformaciones de materiales y energía que sólo pueden ser obtenidas de la naturaleza.

2. ANTECEDENTES

Este capítulo muestra algunos trabajos referentes que presentan relación con los objetivos de estudio. Inicialmente se exponen antecedentes que corresponden al área de Biología, estudios de la macrofauna y estructura del suelo en diferentes tipos de paisajes. Posteriormente se presentan artículos relacionados con la caracterización estadística y técnicas multivariadas empleadas, estas publicaciones se presentan de forma ascendente en el tiempo.

2.1 ESTUDIOS BIOLÓGICOS

Con respecto al estudio de la relación de las especies de macroinvertebrados y el medio ambiente, se encuentran en la literatura varios trabajos de investigación realizados, entre los cuales se resalta el de Moya *et al.* (2004), en su artículo “Comunidad de macroinvertebrados en ríos de la cuenca Chipiriri con condiciones de referencia y alteradas, estudio preliminar para la bio – evaluación de arroyos”, el propósito fue determinar la composición y estructura de comunidades de macroinvertebrados bentónicos⁶ y su relación con factores ambientales en condiciones naturales y alteradas. Para tener una visión general acerca de la variabilidad ambiental de la zona estudiada, se realizó inicialmente un análisis de componentes principales (ACP) y para determinar asociaciones entre la estructura de las comunidades y las variables ambientales, se realizó un análisis de correspondencia canónica (ACC). El ACP muestra como resultado que la variación entre las variables ambientales puede ser explicada por los dos primeros componentes principales en un 51%. El primer eje separa los sitios con valores elevados de variables físicas como la velocidad de corriente, la longitud, el ancho, el tipo de sustrato dominante y la intermitencia. El segundo eje ordena los sitios en función de las variables químicas como el pH, la conductividad y el oxígeno disuelto. Concluyendo que tanto los sitios de referencia como perturbados no se separan marcadamente, lo que indica que ambos grupos de sitios comparten las mismas características ambientales. El ACC, indica que la varianza acumulada de la relación entre las variables ambientales y biológicas puede ser explicada por los dos primeros ejes en un 62%. La correlación especie – ambiente fue alta para ambos ejes (0.878 y 0.791, respectivamente). Los test de permutación de Monte Carlo (1000 permutaciones) muestra que ambos ejes fueron altamente significativas ($p=0.001$). Estos resultados sugieren que la distribución de las especies no fue debida al azar, lo que indica que la mayor parte de la variación explicable de los datos de especies fueron tomados en cuenta por las variables ambientales medidas.

Espinoza y Morales (2008), en su artículo “Macroinvertebrados bentónicos de la laguna Las Peonías, estado Zulia, Venezuela” con el objeto de evaluar la comunidad de macroinvertebrados bentónicos de la laguna Las Peonías en el estado de Zulia, realizó cinco muestreos en ocho estaciones, los cuales fueron muestreados en junio, julio, octubre, noviembre y diciembre del 2006. Se determinaron las variables fisicoquímicas del agua

⁶ Organismos que viven y realizan sus funciones vitales en dependencia estricta de un sustrato.

(salinidad, temperatura, turbidez y pH) y la granulometría⁷ y contenido de materia orgánica de los sedimentos. La comunidad estuvo compuesta por tres filos: Artropoda, Mollusca y Annelida. La densidad promedio fue de 1.680 ind/m², la menor densidad se observó durante noviembre (1.124 ind/m²) y diciembre (322 ind/m²) y la mayor en los meses junio (4.888 ind/m²) y julio (3.348 ind/m²). El taxón más abundante lo constituyeron las larvas de la familia Chironomidae (95%). Los valores del índice de diversidad ($H' = 0,12$), riqueza ($R1 = 0,96$) y equidad ($E = 0,01$) resultaron bajos, influenciados por la alta densidad de los individuos de esta familia indicando un alto nivel de contaminación en la laguna. La composición y distribución de los organismos presentaron relación con las variables fisicoquímicas del agua y del sedimento. Las variaciones y fluctuaciones en el contenido de materia orgánica y la salinidad se presentan como los reguladores de la comunidad de macroinvertebrados bentónicos de la laguna Las Peonías. Se aplicó análisis de componentes principales (ACP) a la matriz de correlaciones de las variables físico – químicas, con la finalidad de determinar cuáles parámetros explicaban la mayor variabilidad en la laguna. Se utilizó un análisis de correspondencia canónica (ACC) para identificar relaciones entre las variables fisicoquímicas y la estructura de la comunidad macrobentónica (Cuadras 1991). Los resultados del ACC evidenciaron que los parámetros que dominaron la distribución de los organismos en la laguna fueron la salinidad y el porcentaje de materia orgánica (% MO) en los sedimentos. La salinidad representó un 7% y la materia orgánica un 3% de la variabilidad entre los factores físico – químicos. Este valor para el resto de los parámetros fue menor del 1%. Aunque los porcentajes de variabilidad son bajos, se observó cómo los valores de densidad promedio de los organismos fluctuaban con respecto a los factores salinidad y contenido de materia orgánica. La salinidad se comportó de manera muy homogénea y el contenido de materia orgánica decreció, mientras la densidad aumentó y viceversa.

Ruiz (2004), en su trabajo doctoral “Desarrollo de un sistema bioindicador de calidad de la tierra basado en estudio de los macroinvertebrados” de la Universidad de Paris, propone un bioindicador de la calidad del suelo sobre la base de la composición y abundancia de las poblaciones de 15 órdenes de macroinvertebrados que habitan en el suelo. La primera parte de este trabajo demuestra la alta sensibilidad de los macroinvertebrados del suelo a diferentes parámetros del suelo y sus cambios que los hacen excelentes indicadores de la calidad. El segundo objetivo consiste en desarrollar un índice sintético de calidad del suelo, el Índice de Calidad del Suelo biótico (IBQS). Asimismo, se realizó análisis de varianza en los datos físico – químicos y los datos de la macrofauna. Por último se realizó el análisis multivariado sobre la abundancia de macroinvertebrados del suelo y los datos físico-químicos con el método del análisis de co – inercia para observar las posibles relaciones entre los dos conjuntos de datos.

2.2 REFERENCIAS DE METODOLOGÍA ESTADÍSTICA

Rosbaco, Bisaro, Romagnoli y Martignone (2004), en su artículo “Métodos estadísticos para la caracterización de ambientes climáticos y evaluación de respuestas de cultivos de soja”. Realizaron experimentos a campo en tres sistemas agrícolas, utilizando un diseño de bloques completos al azar con tres repeticiones. Los datos se procesaron por análisis de la variancia y

⁷ Método para la separación de un suelo en diferentes fracciones.

las interacciones se evaluaron según Análisis de Componentes Principales y Análisis de Conglomerados. Debido que las fechas de siembra y los suelos utilizados fueron similares, las diferencias entre años fueron causadas por las variables climáticas, temperatura, lluvias y radiación solar incidente en cada etapa fenológica. Los métodos estadísticos utilizados permitieron separar claramente los tres años climáticos y mostraron diferencias entre cultivares en cada año. Los rendimientos y sus componentes se explicaron en función de las variables climáticas y se identificaron ajustes ambientales en el número y tamaño de las semillas. Los métodos estadísticos aplicados en este trabajo permiten la identificación de los procesos ecofisiológicos del cultivar que determinan el rendimiento en relación con el ambiente. La difusión de esta técnica puede contribuir a incrementos sostenidos del rendimiento del cultivo.

Díaz – Quirós y Rivera – Rondón (2004), en su artículo “Diatomeas de pequeños ríos andinos y su utilización como indicadores de condiciones ambientales”, realizaron un estudio en veinte ríos de la cuenca alta y media del Río Bogotá (Cundinamarca, Colombia), donde se analizó la relación entre algunas variables físicas, químicas e hidrológicas con la comunidad de diatomeas perifíticas. A partir de un Análisis de Componentes Principales, se estableció que la conductividad, el pH, el amonio, la turbidez, el fósforo total, el dióxido de carbono y la velocidad de la corriente fueron las variables que determinaron los patrones físicos y químicos de los ríos. Mediante un Análisis de Correspondencia Canónica se estableció que el pH, el fósforo reactivo soluble, el nitrógeno total y la conductividad fueron las variables que determinaron la comunidad de diatomeas. Finalmente, se propone un modelo de regresión y calibración para inferir los valores de las variables significativas a partir de la estructura de la comunidad.

Gómez - Anaya, Palacios - Vargas y Castaño – Meneses (2010), en su artículo “Abundancia de colémbolos (Hexapoda: Collembola) y parámetros edáficos de una selva baja caducifolia”, describe y compara los patrones de abundancia de colémbolos de hojarasca y suelo de dos cuencas hidrológicas de la Estación de Biología Chamela en Jalisco – México, con base en colectas mensuales durante un año. Se analizó el efecto de los factores cuenca, biotopo y estacionalidad sobre la abundancia de las familias de Collembola. Mediante análisis de conglomerados (AC) se exploran las relaciones faunísticas entre los ensamblajes, y mediante análisis de correspondencia canónica (ACC), las relaciones entre los ensamblajes y las variables fisicoquímicas del suelo. Se obtuvieron 24.253 colémbolos pertenecientes a trece familias. El AC mostró mayor similitud entre biotopos que entre cuencas, mientras que en el ACC la variable más importante para explicar la distribución y abundancia de los Collembola fue la humedad.

3. MARCO TEÓRICO

En este capítulo se hace una breve descripción de la terminología utilizada en el trabajo, primero se definen los sistemas de producción y sus respectivos usos del suelo, la macrofauna y las propiedades físicas y químicas del suelo. Posteriormente, se presentan los métodos estadísticos que soportarán la metodología de análisis.

3.1 TÉRMINOS BIOLÓGICOS

Se describen algunos conceptos de biología utilizados en el presente trabajo, y así tener conocimiento de éstos para su interpretación y análisis.

3.1.1 Biodiversidad

También llamada diversidad biológica, es la variabilidad de organismos vivos de cualquier fuente, incluidos, entre otras cosas, los ecosistemas terrestres y marinos y otros ecosistemas acuáticos y los complejos ecológicos de los que forman parte; comprende la diversidad dentro de cada especie, entre las especies y de los ecosistemas.

3.1.2 Deforestación

Elimina la cobertura de los árboles en aras de la agricultura, actividades mineras, represas, creación y mantenimiento de la infraestructura, expansión de las ciudades y otras consecuencias debidas a un crecimiento rápido de la población. Entre las causas más importantes de la deforestación figuran la tala, la conversión del bosque a la agricultura y a la cría de ganado, la urbanización y la construcción de infraestructura, la minería y la explotación de petróleo, la lluvia ácida y los incendios. No obstante, ha habido una tendencia a hacer hincapié en los pequeños agricultores migratorios como causa principal de la pérdida de bosques. La tendencia general de estos agricultores es la de asentarse a lo largo de caminos que atraviesen el bosque, talar una parcela de tierra y utilizarla para la plantación de cultivos de subsistencia o de cultivos comerciales. En los bosques tropicales, ese tipo de prácticas terminan por provocar una rápida degradación del suelo ya que en gran medida es demasiado pobre como para resistir las prácticas agrícolas. Por consiguiente, a los pocos años el agricultor se ve forzado a talar otra parcela del bosque. La tierra agrícola degradada a menudo es utilizada algunos años más para la cría de ganado, lo cual equivale a firmar su sentencia de muerte ya que el ganado elimina los últimos rastros de fertilidad que podían quedar. El resultado es una parcela de tierra totalmente degradada que durante muchos años no podrá recuperar su biomasa original.

3.1.3 Macroinvertebrados del suelo

Son organismos cuyo ancho es > 2 mm, conocidos como macrofauna, intervienen en distintos procesos: en la agregación y estructura del suelo, en la textura y consistencia del suelo, en el movimiento y la retención del agua, en el intercambio gaseoso y en las propiedades químicas y nutricionales del mismo (Lal, 1989; Brussard, 1998; Neher, 1999; Lavelle y Spain, 2001) existiendo un fuerte efecto del tipo de vegetación sobre ellos (Brussard, 1998). Estos son importantes reguladores de muchos procesos del ecosistema: tienen efectos positivos en la conservación de la estructura del suelo; actúan sobre el microclima, la humedad y la aireación; pueden activar o inhibir la función de los microorganismos y están involucrados en la conservación y ciclado de nutrientes (Lavelle 1990, Lavelle et al. 1993, Salamanca y Chamorro 1994, Wolters y Ekschmitt 1997). Consecuentemente, la abundancia de estos organismos provisionan la buena estructura del suelo.

3.1.4 El suelo

Es considerado como uno de los recursos naturales más importantes, de ahí la necesidad de mantener su productividad, para que a través de él y las prácticas agrícolas adecuadas se establezca un equilibrio entre la producción de alimentos y el acelerado incremento del índice demográfico.

La composición química y la estructura física del suelo en un lugar dado, están determinadas por el tipo de material geológico del que se origina, por la cubierta vegetal, por la cantidad de tiempo en que ha actuado la meteorización, por la topografía y por los cambios artificiales resultantes de las actividades humanas. Las variaciones del suelo en la naturaleza son graduales, excepto las derivadas de desastres naturales. Sin embargo, el cultivo de la tierra priva al suelo de su cubierta vegetal y de mucha de su protección contra la erosión del agua y del viento, por lo que estos cambios pueden ser más rápidos. Los agricultores han tenido que desarrollar métodos para prevenir la alteración perjudicial del suelo debida al cultivo excesivo y para reconstruir suelos que ya han sido alterados con graves daños.

Las propiedades químicas y físicas de los suelos son controladas, en gran parte, por la arcilla y el humus, las que actúan como centros de actividad a cuyo alrededor ocurren reacciones químicas y cambios nutritivos. Las propiedades físicas más importantes son su textura, densidad aparente y estructura. Éstas definen la capacidad de infiltración y almacenamiento de agua y contribuyen a definir su capacidad de uso; y las propiedades químicas más importantes son la capacidad de intercambio catiónico, el grado de acidez o alcalinidad y la concentración de sales en el perfil. Las propiedades químicas de los suelos están, sobre todo, asociadas con su capacidad de uso con fines agrícolas.

3.1.5 Sistemas de producción

Las prácticas agroforestales son la mejor alternativa para el uso sostenible de las tierras de laderas tropicales. Casi todos los sistemas agrícolas tradicionales, los cuales incluyen los

sistemas ganaderos, tienen árboles intercalados con cultivos o manejados en una forma zonal alternando árboles y cultivos y/o pastos; es decir, son sistemas agroforestales, aún con la modernización de la agricultura de la región, los paisajes agrícolas todavía contienen un alto número de árboles, estos árboles cumplen con muchos propósitos como producción (madera, leña, forraje, frutas, medicinas, etc.) además de servicios (sombra para cultivos y/o animales, protección como en el caso de cortinas rompevientos, etc.). Los árboles aumentan la diversidad biológica del agroecosistema creando en sus ramas, en sus raíces y en la hojarasca, hogares para otros organismos (Beer et al., 2004). Estos sistemas a su vez posee varias clases, entre ellas las que se estudian en el presente trabajo, el Sistema Agroforestal, la cual es la forma de usar la tierra que involucra la combinación de especies leñosas (árboles, arbustos y palmas) con especies herbáceas en sistemas boscosos, en procura de la sostenibilidad del sistema (árboles y bosque); el Sistema silvopastoril (SSP) que es un manejo de suelos sostenible, el cual incrementa el potencial del mismo combinando la producción de cultivos (incluyendo frutales, leñosos) o animales y árboles forestales en forma simultánea o secuenciada en el tiempo, sobre la misma unidad de tierra, y aplica prácticas de manejo que son compatibles con las prácticas culturales de la población local (árboles y pastizal); y el Sistema tradicional que son siembras de pastizales constituidas fundamentalmente en plantas nativas, pudiendo además tener otros usos alternativos (recreación, leña, etc.).

3.1.6 Usos del suelo

Es cualquier tipo de utilización humana de un terreno, en ambiente construido como campos de cultivo, pasturas y asentamientos urbanos. Entre los cuales se enfatizan el Pastizal mejorado (PM) que corresponde a cultivos de pastos nativos e introducidos principalmente *Brachiaria* sp.⁸ y *Clidemia rubra*⁹ sin la presencia de árboles forrajeros, con actividad ganadera intensiva; el Pastizal degradado (PD) que son pastos nativos sin presencia de árboles forrajeros; el Pastizal con árboles dispersos (PDAD) donde la cobertura arbórea representa menos del 10% del total de la superficie de la superficie muestreada; el Rastrojo (RS) compuesta por la vegetación de más de diez años de hasta 8 m de altura y con DAP (diámetro a la altura del pecho) >25 mm, incluye especies arbustivas y herbáceas; el Agroforestal (SAF) el cual implica la combinación de árboles forestales con cultivos, con ganadería, o una combinación de ambos, incluye pastos combinados con gramíneas para pastoreo, bancos de proteína y producción agroforestal con frutales amazónicos, así como árboles maderables nativos; el Cultivo de caucho (CH) Conformado por plantaciones de *Hevea brasiliensis*¹⁰; y Bosque intervenido (BI) el cual son bosques con extracción de madera con > 25% de cobertura arbórea.

3.2 MÉTODOS ESTADÍSTICOS

Se citan y describen los métodos utilizados para el desarrollo del proyecto, análisis exploratorio de datos y análisis multivariado.

⁸ Cultivo de pasto.

⁹ Especies de planta de flores.

¹⁰ Árbol de caucho de la familia de las euforbiáceas (*Hevea brasiliensis*), de 20 a 30 m de altura.

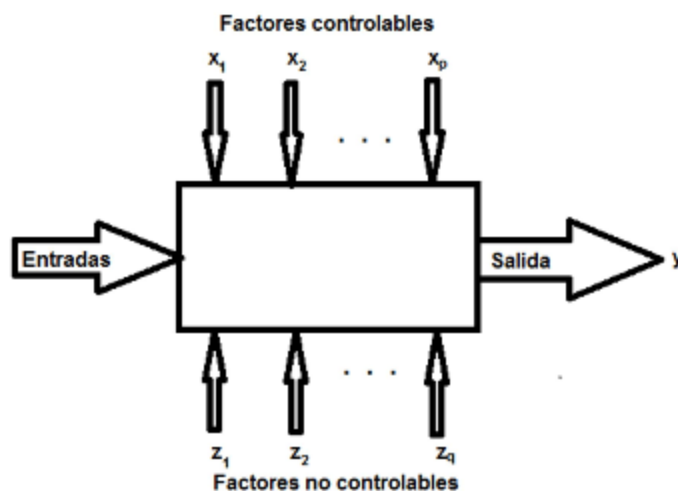
3.2.1 Análisis de experimentos

Un experimento es una prueba o serie de pruebas en el que se hacen cambios deliberados en las variables de entrada x 's de un proceso o sistema con el objetivo de observar e identificar las razones de los cambios que pudieran observarse en las respuestas de salidas y 's, (ver figura 1). El objetivo del experimento es el de determinar cuáles son las variables que tiene mayor influencia sobre las respuestas y 's; el ajuste de las x 's que tiene mayor influencia para que las y 's esté casi siempre cerca de algún valor nominal deseado; el ajuste de las x 's que tiene mayor influencia para que la variabilidad de y 's sea reducida; y el ajuste de las x 's que tiene mayor influencia para que los efectos de las variables no controlables sean mínimos. En resumen, es un procedimiento mediante el cual se trata de comprobar, confirmar o verificar una o varias hipótesis relacionadas con un determinado fenómeno, mediante el análisis de variables. La experimentación constituye uno de los elementos claves del método científico y es fundamental para ofrecer explicaciones causales.

Un experimento aleatorio o estadístico presenta las siguientes características¹¹:

- Todos los posibles resultados del experimento son conocidos antes de hacer una realización del experimento.
- El resultado exacto en cualquier ejecución del experimento no es predecible (aleatoriedad).
- El experimento puede ser repetido bajo similares condiciones.
- Existe un patrón predecible a lo largo de muchas ejecuciones (regularidad estadística).

En cualquier experimento, los resultados y conclusiones que puedan obtenerse dependerán en gran medida de la manera en que se recolectan los datos.



¹¹ http://www.virtual.unal.edu.co/cursos/ciencias/2001065/html/un2/cont_205_47.html

Figura 1. Modelo general de un proceso o sistema.

El proceso puede visualizarse como una combinación de métodos, personas o recursos que transforman cierta entrada en una salida que tiene una o más respuestas observables $y's$. Algunas variables del proceso $x_1, x_2, \dots, x_p$ son controlables, mientras que otras $z_1, z_2, \dots, z_q$ son no controlables.

3.2.2 Análisis de varianza de un factor (ANOVA)

El análisis de varianza es una técnica estadística para el análisis de datos experimentales u observacionales que se deriva de la partición de la variabilidad total en las partes que la componen.

Para el caso de un diseño experimental que tiene en cuenta un factor con los niveles de experimentación, el modelo asumido para la relación es la Ecuación (1). La hipótesis que se prueba es que las medias poblacionales son iguales, implicando que los grupos no difieren entre si y en consecuencia que el factor es independiente de la variable respuesta.

Así, se tienen a tratamientos o niveles diferentes de un solo factor que quieren compararse. La respuesta observada de cada uno de los a tratamientos es una variable aleatoria que puede representarse mediante el siguiente modelo

$$y_{ij} = \mu_i + \varepsilon_{ij} \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, n \end{cases} \quad (1)$$

μ_i es la media del nivel del factor o tratamiento i – *ésimo*, ε_{ij} es un componente del error aleatorio que incorpora todas las demás fuentes de variabilidad del experimento, incluyendo las mediciones, la variabilidad que surge de factores no controlados, las diferencias entre las unidades experimentales a las que se aplican los tratamientos, y el ruido de fondo general en el proceso y y_{ij} s la observación ij – *ésima*. Este modelo supone que los errores tienen media cero y varianza constante, de tal modo que

$$E(y_{ij}) = \mu_i, \quad Var(y_{ij}) = \sigma_\varepsilon$$

La ecuación (1) recibe el nombre de modelo de medias. Otra forma de escribir el modelo en términos de los efectos es

$$\mu_i = \mu + \tau_i \quad i = 1, 2, \dots, a.$$

De tal modo que la ecuación (1) se convierte en

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, n \end{cases} \quad (2)$$

μ es un parámetro común a todos los tratamientos al que se llama la media global, y τ_i es un parámetro único del tratamiento i – *ésimo* al que se le llama el efecto del tratamiento i – *ésimo*.

La ecuación (2) es llamada el modelo de los efectos.

Tanto el modelo de las medias como el modelo de los efectos son modelos estadísticos lineales; es decir, la variable de respuesta y_{ij} es una función lineal de los parámetros del modelo. En el modelo de efectos, μ es una constante y los efectos de los tratamientos τ_i representan desviaciones de esta constante cuando se aplican los tratamientos específicos con la restricción de que $\sum_{i=1}^a \tau_i = 0$.

Se requiere que el experimento se lleve a cabo en orden aleatorio para que el ambiente en el que se apliquen los tratamientos sea lo más uniforme posible. Por lo tanto, el diseño experimental es un diseño completamente aleatorizado. Los objetivos serán probar las hipótesis apropiadas acerca de las medias de los tratamientos y estimarlas.

Para probar las hipótesis, se supone que los errores del modelo son variables aleatorias que siguen una distribución normal e independiente con media cero $E(\varepsilon) = 0$ y varianza constante $Var(\varepsilon) = \sigma^2$. Esto implica que las observaciones

$$y_{ij} \sim N(\mu + \tau_i, \sigma^2)$$

La variabilidad total de las observaciones se divide en aquella explicable por los tratamientos y la debida al error experimental, como en la ecuación (3).

$$\sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^a \sum_{j=1}^n [(\bar{y}_{i.} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.})]^2 \quad (3)$$

$$\begin{aligned} \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = \\ n \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2 + 2 \sum_{i=1}^a \sum_{j=1}^n (\bar{y}_{i.} - \bar{y}_{..})(y_{ij} - \bar{y}_{i.}) \end{aligned} \quad (4)$$

Sin embargo, el término del producto cruzado de la ecuación (4) es cero, puesto que

$$\sum_{j=1}^n (y_{ij} - \bar{y}_{i.}) = y_{i.} - n\bar{y}_{i.} = y_{i.} - n(y_{i.}/n) = 0$$

Se tiene, por tanto

$$\sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = n \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2 \quad (5)$$

La ecuación (5) establece que puede hacerse la partición de la variabilidad total de los datos, medida por la suma de cuadrados total corregida, en una suma de cuadrados de las diferencias entre los promedios de los tratamientos y el gran promedio, más una suma de cuadrados de las diferencias de las observaciones dentro de los tratamientos y el promedio de los tratamientos. Entonces, la diferencia entre los promedios de los tratamientos observados y el gran promedio es una medida de las diferencias entre las medias de los tratamientos, mientras que las diferencias de las observaciones dentro de un tratamiento y el promedio del tratamiento, pueden deberse únicamente al error aleatorio. Por tanto, la ecuación (5) puede escribirse simbólicamente como

$$SS_{TOTAL} = SS_{TRATAMIENTOS} + SS_{ERROR}$$

Donde a $SS_{TRATAMIENTOS}$ se le llama la suma de cuadrados debida a los tratamientos (es decir, entre los tratamientos), y a SS_{ERROR} se le llama la suma de cuadrados debida al error (es decir, dentro de los tratamientos). Hay $N = an$ observaciones en total; por lo tanto, SS_{TOTAL} tiene $N - 1$ grados de libertad. Hay a niveles del factor (y medias de a tratamientos), de donde $SS_{TRATAMIENTOS}$ tiene $a - 1$ grados de libertad. Dentro de cualquier tratamiento hay n réplicas que proporcionan $n - 1$ grados de libertad con los cuales estimar el error experimental. Puesto que hay a tratamientos, se tienen $a(n - 1) = an - a = N - a$ grados de libertad para el error.

Después de realizar los cálculos de las sumas de cuadrados, los grados de libertad, el cuadrado medio y la F, se elabora la tabla del ANOVA, que reúne toda la información como se muestra en la tabla 1.

Tabla 1. Análisis de varianza

Fuente de variación	Suma de cuadrados	Grados de libertad	Promedio de los cuadrados (MS)	Estadístico de prueba F
Tratamientos	$SS_{TRATAMIENTOS}$	$a - 1$	$\frac{SS_{TRATAMIENTOS}}{a - 1}$	$\frac{MS_{TRATAMIENTOS}}{MS_{ERROR}}$
Error	SS_{ERROR}	$N - a$	$\frac{SS_{ERROR}}{N - a}$	
Total	SS_{TOTAL}	$N - 1$		

3.2.3 Anova de un factor con bloque

El anova de un factor con bloque es el más sencillo de este tipo de diseños utilizados para controlar y reducir el error experimental, en él las unidades experimentales quedan estratificadas en bloques de unidades homogéneas, cada tratamiento se asigna al azar a un número igual (por lo general uno) de unidades experimentales en cada bloque y es posible hacer comparaciones más precisas entre los tratamientos dentro del conjunto homogéneo de unidades experimentales en un bloque. En la prueba de significación, sólo interesa investigar la diferencia entre tratamientos sin que la condición bloque interfiera. El modelo lineal para un

experimento en un diseño de bloques completos aleatorizado requiere un término que represente la variación identificable en las observaciones como consecuencia de los bloques. La respuesta de la unidad con el i – *ésimo* tratamiento en el j – *ésimo* bloque se escribe como

$$y_{ij} = \mu + \tau_i + \rho_j + \varepsilon_{ij} \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, n \end{cases}$$

Donde μ es la media general, τ_i es el efecto del tratamiento y ε_{ij} es el error experimental. El efecto del bloque ρ_j , representa la desviación promedio de las unidades en el bloque j a partir de la media general. Se supone que los efectos del tratamiento y del bloque son aditivos, Aditividad significa que no existe interacción entre tratamientos y bloques; también se supone que los errores experimentales son independientes, con medias cero y varianza común σ^2 . La suposición de independencia se justifica a través de la asignación aleatoria de los tratamientos a las unidades experimentales.

La suma de cuadrados totales en un diseño de bloques completamente aleatorizados con un solo factor se descompone de la siguiente manera

$$\sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = n \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2 + a \sum_{j=1}^n (\bar{y}_{.j} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$$

$$SS_{TOTALS} = SS_{TRATAMIENTOS} + SS_{BLOQUES} + SS_{ERROR}$$

Luego de descomponer la suma de cuadrados totales. Se ilustrara los grados de libertad, estadísticos de prueba F en la tabla para la Anova de un factor con bloques, ver tabla 2.

Tabla 2. Anova de un factor con bloque

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio (MS)	Estadístico de prueba F
Tratamientos	$SS_{TRATAMIENTOS}$	$a - 1$	$\frac{SS_{TRATAMIENTOS}}{a - 1}$	$\frac{MS_{TRATAMIENTOS}}{MS_{ERROR}}$
Bloques	$SS_{BLOQUES}$	$n - 1$	$\frac{SS_{BLOQUES}}{n - 1}$	
Error	SS_{ERROR}	$(a - 1)(n - 1)$	$\frac{SS_{ERROR}}{(a - 1)(n - 1)}$	
Total	SS_{TOTAL}	$an - 1$		

3.2.3 Alternativas no paramétricas para anova

Cuando se analizan datos medidos por una variable cuantitativa, las pruebas estadísticas de estimación y contraste frecuentemente empleadas se basan en suponer que se ha obtenido una muestra aleatoria de una distribución de probabilidad normal. En muchas ocasiones esta suposición no resulta válida, en cuyo caso es posible utilizar las pruebas estadísticas que no se basan en ninguna suposición en cuanto a la distribución de probabilidad a partir de la que fueron obtenidos los datos, y por ello se denominan pruebas no paramétricas.

En la mayor parte de las pruebas no paramétricas, los resultados estadísticos se derivan únicamente a partir de procedimientos de ordenación y recuento, por lo que su base lógica es de fácil comprensión.

La prueba de *Kruskal Wallis* es un método no paramétrico análogo al ANOVA, en el cual los datos son reemplazados por rangos. Es una extensión de la prueba de la U de Mann-Whitney para 3 o más grupos. La prueba de Kruskal-Wallis no asume normalidad en los errores, ya que no asume un modelo de análisis en oposición al tradicional ANOVA. Bajo la hipótesis nula, asume que los datos vienen de la misma distribución. Una forma común en que se viola este supuesto es con datos heterocedásticos.

Planteamiento de Hipótesis

$$\begin{aligned} H_0: \theta_1 &= \theta_2 = \dots = \theta_a \\ H_1: \theta_i &\neq \theta_j \quad \text{para al menos un par } (i, j) \end{aligned}$$

Se utiliza para probar la hipótesis nula de que las muestras independientes provienen de poblaciones con medianas iguales y la hipótesis alternativa es la aseveración de que las poblaciones tienen medianas que no son iguales.

El procedimiento parte de una transformación de los datos originales a un sistema de ranqueo r_i que representa la posición asumida por el dato i –ésimo en un ordenamiento ascendente dentro del conjunto global de datos.

El estadístico está dado por

$$K = (N - 1) \frac{\sum_{i=1}^a n_i (\bar{r}_i - \bar{r})^2}{\sum_{i=1}^a \sum_{j=1}^{n_i} (r_{ij} - \bar{r})^2}$$

n_i es el número de observaciones en el grupo i ($i = 1, 2, \dots, a$), r_{ij} es el rango (entre todas las observaciones) de la observación j ($j = 1, 2, \dots, n_i$), en el grupo i , N ($N = \sum_{i=1}^a n_i$) es el número total de observaciones entre todos los grupos y a es el número de tratamientos o grupos.

$\bar{r}_i = \frac{\sum_{j=1}^{n_i} r_{ij}}{n_i}$, $\bar{r} = \frac{N+1}{2}$ es el promedio global de r_{ij} . El denominador para K es exactamente

$$\frac{(N-1)N(N+1)}{12}$$

Entonces puede expresarse como

$$K = \frac{12}{N(N+1)} \sum_{i=1}^a n_i (\bar{r}_i - \bar{r})^2.$$

Este estadístico sigue asintóticamente una distribución χ^2_{a-1} , por tanto, el p-valor es aproximado por $P(\chi^2_{a-1} \geq K)$.

Al emplear la prueba de Kruskal – Wallis, existe un factor de corrección que debe aplicarse siempre que existan muchos empates

$$K^* = \frac{K}{1 - \frac{\sum t^3 - t}{N^3 - N}}$$

Si el resultado de la prueba de Kruskal-Wallis por rangos es significativo, indica que hay una diferencia entre al menos dos de las medianas de la muestra en el conjunto de medianas. Como resultado, el investigador puede concluir que existe una alta probabilidad de que al menos dos de las muestras representan poblaciones con diferentes valores de la mediana. Cuando el valor de K es significativo, es necesario llevar a cabo comparaciones contrastantes entre grupos específicos. Varias fuentes como: Siegel y Castellan (1988), describen el procedimiento para la comparación de Kruskal-Wallis, que es esencialmente la aplicación del método de Bonferroni de un solo factor entre sujetos de análisis varicentro el modelo de Kruskal-Wallis. A través del uso de la inecuación

$$|\bar{R}_i - \bar{R}_j| \geq Z_{adj} \sqrt{\frac{n(N+1)}{12} \left[\frac{1}{n_a} + \frac{1}{n_b} \right]}$$

Donde, n_a y n_b representan el número de sujetos en cada uno de los grupos que participan en la comparación. El valor de la Z_{adj} se obtiene de la tabla de la distribución normal, ver anexos. En el caso de una hipótesis alternativa no direccional Z_{adj} es el valor Z por encima del cual una proporción de casos correspondiente al valor de $\alpha_{FW}/2c$ cae (c es el número total de comparaciones que son llevadas a cabo). En el caso de una hipótesis alternativa direccional Z_{adj} es el valor Z por encima del cual una proporción de casos correspondiente al valor de $\alpha_{FW}/2c$ cae.

Este procedimiento permite identificar la diferencia mínima requerida entre las medias de las filas de los dos grupos (designados como CD_{kw}) con el fin de que difieran uno de otro en el nivel de significación especificado previamente.

La prueba de *Friedman* es una prueba no paramétrica desarrollada por el economista Milton Friedman. Homóloga a la prueba ANOVA diseño de un factor con bloques, el método consiste en ranquear los datos por bloque, remplazándolos.

Plantea la hipótesis nula

$$H_0: \theta_1 = \theta_2 = \dots = \theta_a$$

$$H_1: \theta_i \neq \theta_j \quad \text{para al menos un par } (i, j)$$

Es decir, las respuestas asociadas a cada uno de los tratamientos tienen la misma distribución de probabilidad o distribuciones con la misma mediana, frente a la hipótesis alternativa de que por lo menos la distribución de una de las respuestas difiere de las demás.

Los datos se disponen en una tabla en la que en cada fila se recogen las respuestas de los n elementos de cada grupo a los a tratamientos, ver tabla 3.

Tabla 3. Estructura de los datos

Bloques	Tratamientos			
	1	2		a
1	X_{11}	X_{12}		X_{1a}
2	X_{21}	X_{22}		X_{2a}
3	X_{31}	X_{32}		X_{3a}
....				
n	X_{n1}	X_{n2}		X_{na}

A las observaciones de cada fila se les asignan rangos de menor a mayor desde 1 hasta a ; a continuación se suman los rangos correspondientes a cada columna, siendo R_j la suma correspondiente a la columna j —ésima—. Si la hipótesis nula es cierta, la distribución de los rangos en cada fila se debe al azar, y es de esperar que la suma de los rangos correspondientes a cada columna sea aproximadamente igual a $n(a + 1)/2$. La prueba de Friedman determina si las R_j observadas difieren significativamente del valor esperado bajo la hipótesis nula.

El estadístico de prueba es

$$x_r^2 = \frac{12}{na(a+1)} \left[\sum_{i=1}^a R_j^2 \right] - 3n(a+1)$$

Si la hipótesis nula es cierta y el número de columnas y/o de filas es moderadamente grande la distribución de x_r^2 se aproxima a una chi-cuadrado con $a - 1$ grados de libertad; de forma que se rechaza la hipótesis nula para valores de x^2 superiores al valor crítico para el nivel de significación fijado.

En el caso de existir diferencias significativas, diversas fuentes (Daniel en 1990 y Siegel y Castellan en 1988) describen un procedimiento de comparación para Friedman de dos vías. A través del uso de la ecuación

$$CD_F = Z_{adj} \sqrt{\frac{na(a+1)}{6}}$$

El valor de la Z_{adj} se obtiene de la tabla de la distribución normal, ver anexos. En el caso de una hipótesis alternativa no direccional Z_{adj} es el valor Z por encima del cual una proporción de casos correspondiente al valor de $\alpha_{FW}/2c$ cae (c es el número total de comparaciones que son llevadas a cabo). En el caso de una hipótesis alternativa direccional Z_{adj} es el valor Z por encima del cual una proporción de casos correspondiente al valor de $\alpha_{FW}/2c$ cae. Esta ecuación, permite identificar la diferencia mínima requerida entre las sumas de las filas de las dos condiciones (designado como CD_F) en orden para que se diferencian entre sí en el nivel de significación especificado previamente.

3.2.4 Análisis multivariante

El desarrollo teórico del siglo XX y las aplicaciones crecientes de la estadística han hecho de las técnicas del Análisis Multivariante junto con el Análisis de Regresión, uno de los instrumentos más empleados para el estudio del entorno ambiental, económico y social. El análisis multivariado consta de métodos multidimensionales los cuales tratan a más de dos variables simultáneamente, es diferente a las técnicas usadas en estadística descriptiva que no tratan más que una o dos variables a la vez. Permiten la comparación entre numerosas variables, lo que lo hace más rico en información que el examen por separado de cada una, debido a que sintetizan la información considerablemente, extrayendo las tendencias más sobresalientes de datos demasiado numerosos.

El principio de los métodos multivariados consiste en construir y representar sobre gráficos dos nubes de puntos correspondientes respectivamente a las filas y a las columnas de la tabla estudiada.

3.2.4.1 Análisis de Componentes Principales (ACP)

Técnica proveniente del análisis exploratorio de datos, que permite sintetizar la información, reduciendo la dimensión de análisis. Es decir, ante una tabla de datos con muchas variables, el objetivo será reducirlas a un menor número perdiendo la menor cantidad de información posible. Se aplica a tablas bidimensionales que cruzan individuos y variables cuantitativas. Los términos individuos y variables encubren nociones diferentes. En relación con los individuos aquello que se trata de evaluar es su semejanza, dos individuos se asemejan más cuanto más próximos sean sus valores en el conjunto de las variables; con respecto a las variables, lo que se trata de evaluar es su relación, la relación entre dos variables se mide a través del coeficiente de correlación lineal, o su covarianza.

El ACP es uno de los métodos más utilizados en minería de datos. El primero que la introdujo fue Pearson en 1901 y se desarrolló independientemente en 1933 por Hotelling y la primera implementación computacional se dio en el año 1960. Fue aplicado para analizar encuestas de opinión pública por Jean Pagès.

Estas nuevas componentes principales o factores son calculados como una combinación lineal de las variables originales, y serán linealmente independientes. La interpretación del ACP no se da a priori, sino que será deducida tras observar la relación de los componentes principales con las variables originales, tomando en cuenta las correlaciones y la información que se tenga de la investigación.

Los n individuos de una tabla de datos se pueden ver como una nube de puntos en $\mathbb{R}^p$, con su centro de gravedad localizado en el origen, y lo que se busca es un subespacio q -dimensional ($\mathbb{R}^q$) de $\mathbb{R}^p$ ($q < p$), usualmente un plano, tal que la proyección ortogonal de los n puntos sobre $\mathbb{R}^q$ tienen varianza máxima, lo cual permitirá el estudio de relaciones entre los individuos (filas) de la tabla de datos. Las observaciones pueden ser representadas tanto en el espacio de las variables como en el espacio de los individuos.

Los valores de las nuevas variables se llaman puntajes de las componentes principales. Las variables y las nuevas variables se conservan centradas con media cero. La variabilidad total de las variables nuevas como la de las variables originales es la misma, es decir, la variabilidad total de los datos no se altera por transformaciones ortogonales.

Es importante resaltar el hecho de que el concepto de mayor información se relaciona con el de mayor variabilidad o varianza. Cuanto mayor sea la variabilidad de los datos (varianza) se considera que existe mayor información, lo cual está relacionado con el concepto de entropía.

Cálculo de las componentes principales

Se considera una serie de variables $X_1, X_2, \dots, X_p$, sobre un grupo de individuos y se trata de calcular, a partir de ellas, un nuevo conjunto de variables $Y_1, Y_2, \dots, Y_p$, incorrelacionadas entre sí, cuyas varianzas vayan decreciendo progresivamente.

Cada Y_j , donde $j = 1, 2, \dots, p$ es una combinación lineal de las $X_1, X_2, \dots, X_p$, originales, es decir,

$$Y_j = u_{j1}X_1 + u_{j2}X_2 + \dots + u_{jp}X_p = Xu_j,$$

Siendo

$$u'_j = (u_{1j}, u_{2j}, \dots, u_{pj})a'_j = a_{1j}, a_{2j}, \dots, a_{pj}$$

un vector de constantes, y

$$X = [X_1, X_2, \dots, X_p]$$

donde X es una matriz de n filas y p variables. Para mantener la ortogonalidad de la transformación se impone que el módulo del vector

$$u'_j = (u_{1j}, u_{2j}, \dots, u_{pj})$$

sea

$$u'_j u_j = \sum_{k=1}^p u_{kj}^2 = 1.$$

El primer componente se calcula eligiendo u_1 de modo que tenga la mayor varianza posible, sujeta a la restricción de que $u'_1 u_1 = 1$. El segundo componente principal se calcula obteniendo u_2 de modo que la variable obtenida, Y_2 esté incorrelacionada con Y_1 .

Del mismo modo se eligen $Y_1, Y_2, \dots, Y_p$, incorrelacionadas entre sí, de manera que las variables aleatorias obtenidas vayan teniendo cada vez menor varianza.

Para el proceso de extracción de factores se quiere elegir u_1 de modo que se maximice la varianza de Y_1 sujeta a la restricción de que $u'_1 u_1 = 1$, es decir

$$Var(Y_1) = Var(Xu_1) = u'_1 \Sigma u_1.$$

El método habitual para maximizar una función de varias variables sujeta a restricciones, es el método de los multiplicadores de Lagrange.

Se maximiza la función $u'_1 \Sigma u_1$ sujeta a la restricción $u'_1 u_1 = 1$.

Así, se construye la función L:

$$L(u_1) = u_1' \Sigma u_1 - \lambda(u_1' u_1 - 1)$$

Se busca el máximo, derivando e igualando a 0:

$$\frac{\partial L}{\partial u_1} = 2\Sigma u_1 - 2\lambda u_1 = 0$$

Siendo lo anterior, un sistema lineal de ecuaciones. Por el teorema de Roché-Frobenius, para que el sistema tenga una solución distinta de 0 la matriz $(\Sigma - \lambda I)$ tiene que ser singular. Esto implica que el determinante debe ser igual a cero:

$$|\Sigma - \lambda I| = 0$$

y de este modo, λ es un valor propio de Σ . La matriz de covarianzas Σ es de orden p y si además es definida positiva, tendrá valores propios distintos $\lambda_1, \lambda_2, \dots, \lambda_p$, tales que $\lambda_1 > \lambda_2 > \dots > \lambda_p$. Desarrollando la expresión anterior, se tiene:

$$\begin{aligned} (\Sigma - \lambda I)u_1 &= 0 \\ \Sigma u_1 - \lambda u_1 &= 0 \\ \Sigma u_1 &= \lambda u_1 \end{aligned}$$

Entonces:

$$Var(Y_1) = Var(Xu_1) = u_1' \Sigma u_1 = u_1' \lambda u_1 = \lambda u_1' u_1 = \lambda * 1 = \lambda$$

Luego, para maximizar la varianza de Y_1 se tiene que tomar el mayor valor propio λ_1 , y el correspondiente vector propio u_1 .

Donde, u_1 es un vector que da la combinación de las variables originales que tiene mayor varianza, es decir, si

$$u_1' = (u_{11}, u_{12}, \dots, u_{1p}),$$

Entonces

$$Y_1 = u_{11}X_1 + u_{12}X_2 + \dots + u_{1p}X_p = Xu_1.$$

El segundo componente principal, $Y_2 = Xu_2$ se obtiene mediante un argumento similar. Además, se quiere que Y_2 esté incorrelado con el anterior componente Y_1 , es decir, $Cov(Y_2, Y_1) = 0$. Por lo tanto:

$$Cov(Y_2, Y_1) = Cov(Xu_2, Xu_1) = u'_2 E[(X - \mu)(X - \mu)'] u_1 = u'_2 \Sigma u_1$$

Es decir, se requiere que $u'_2 \Sigma u_1 = 0$.

Como se tenía que $\Sigma u_1 = \lambda u_1$, lo anterior es equivalente a

$$u'_2 \Sigma u_1 = u'_2 \lambda u_1 = \lambda u'_2 u_1 = 0$$

esto equivale a que $u'_2 u_1 = 0$, es decir, que los vectores sean ortogonales.

De este modo, se maximiza la varianza de Y_2 , es decir, $u'_2 \Sigma u_2$, sujeta a las siguientes restricciones:

$$\begin{aligned} u'_2 u_2 &= 1 \\ u'_2 u_1 &= 0 \end{aligned}$$

Se toma la función:

$$L(u_2) = u'_2 \Sigma u_2 - \lambda(u'_2 u_2 - 1) - \delta u'_2 u_1$$

y se deriva

$$\frac{\partial L(u_2)}{\partial u_2} = 2\Sigma u_2 - 2\lambda u_2 - \delta u_1 = 0$$

si se multiplica por u'_1 , entonces:

$$2u'_1 \Sigma u_2 - \delta = 0$$

porque

$$\begin{aligned} u'_1 u_2 &= u'_2 u_1 = 0 \\ u'_1 u_1 &= 1 \end{aligned}$$

luego

$$\begin{aligned} \delta &= 2u'_1 \Sigma u_2 = 2u'_2 \Sigma u_1 = 0 \\ \delta &= 2a'_1 \Sigma a_2 = 2a'_2 \Sigma a_1 = 0 \end{aligned}$$

Ya que $Cov(Y_2, Y_1) = 0$.

De este modo, $\frac{\partial L(u_2)}{\partial u_2}$ queda finalmente como:

$$\frac{\partial L(u_2)}{\partial u_2} = 2\Sigma u_2 - 2\lambda u_2 - \delta u_1 = 2\Sigma u_2 - 2\lambda u_2 = (\Sigma - \lambda_2 I)u_2 = 0$$

Entonces, se elige a λ_2 como el segundo mayor valor propio de la matriz Σ con su vector propio asociado $u_2 a_2$. Por lo anterior, se concluye que el j -ésimo componente le correspondería el j -ésimo valor propio. Así, todos los componentes Y se pueden expresar como el producto de una matriz A formada por los vectores propios, multiplicada por el vector X que contiene las variables originales $X_1, X_2, \dots, X_p$:

$$Y = XA$$

donde

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix}, X = [X_1, X_2, \dots, X_p], A = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1p} \\ u_{21} & u_{22} & \cdots & u_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ u_{p1} & u_{p2} & \cdots & u_{pp} \end{bmatrix}$$

Como

$$\begin{aligned} \text{Var}(Y_1) &= \lambda_1 \\ \text{Var}(Y_2) &= \lambda_2 \\ &\vdots \\ \text{Var}(Y_p) &= \lambda_p \end{aligned}$$

La matriz de covarianzas de Y será

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix}$$

Porque $Y_1, Y_2, \dots, Y_p$, se han construido como variables incorrelacionadas.

Se tiene que

$$\Lambda = \text{Var}(Y) = A' \text{Var}(X) A = A' \Sigma A$$

O bien

$$\Sigma = A' \Lambda A$$

Ya que A es una matriz ortogonal por lo que $AA' = I$.

Interpretación

No hay criterios estrictamente formales para la determinación del número de componentes principales a escoger. Los criterios sugeridos son de tipo empírico, y se basan en la variabilidad que en una situación particular se quiere mantener. Existen algunas ayudas gráficas con las cuales se decide el número adecuado de componentes, el gráfico de valores propios. Otro criterio consiste en tomar un número de componentes igual al número de valores propios que están por encima de la media; de esta manera si las componentes han sido generadas desde la matriz de correlaciones, se seleccionan las componentes cuyos valores propios asociados sean mayores que 1.

En el gráfico biplot del primer plano factorial se superponen vectores que representan a cada una de las variables originales, éstos representan la dirección de mejor ajuste para cada una de las variables, en el sentido de que, si se proyecta los puntos que representan a los individuos sobre uno de los vectores, las puntuaciones obtenidas estarán más correlacionadas con la variable original que las proyecciones en cualquier otra dirección. El coseno del ángulo que forma cada vector con el eje, mide aproximadamente la correlación con el mismo y los cosenos de los ángulos entre dos vectores aproximan la correlación entre las variables a las que representan, es decir, a menor ángulo menor correlación.

A las correlaciones al cuadrado entre la variable y el eje se le denomina contribución relativa del factor al elemento, y mide la parte de la variabilidad de la variable que explica cada uno de los ejes. Las variables con contribuciones altas en uno de los ejes y bajas en los demás son las que han de interpretarse para cada eje, ya que son características exclusivas del mismo. Como las componentes son incorreladas, tienen información independiente por lo que la suma de las correlaciones al cuadrado es 1. La parte explicada por un plano, se calcula simplemente sumando la parte explicada (contribución) por los ejes que lo componen, a esta cantidad se la denomina también “calidad de la representación”, y puede interpretarse como la correlación al cuadrado entre los valores de la variable original y las proyecciones de los puntos sobre la dirección que representa a la variable.

De la misma manera que hemos hecho para las variables es posible definir calidades de representación para los individuos. La calidad de la representación con respecto a un plano se mide de la misma manera, es decir, como el coseno al cuadrado del ángulo que forman el vector y el plano. Este coseno al cuadrado es la suma de los cosenos al cuadrado de los ángulos con los ejes que forman el plano.

Para la definición de cada uno de los componentes, es importante conocer qué modalidades de las variables en estudio ha contribuido en la elaboración de cada uno de los ejes, es decir, el peso que tiene cada modalidad en la definición de cada uno de los ejes. Este peso se denomina contribución absoluta y representa porcentualmente la importancia que tiene cada modalidad en la formación del eje. La suma de todas las contribuciones absolutas tanto para las frecuencias activas como para los individuos en cada eje será igual a 100.00 y a 1 respectivamente. Definidas las contribuciones absolutas, se calculan las contribuciones relativas, estas proveen información de cuanto de la inercia de una modalidad esta explicada por el eje. Es importante entender las diferencias entre las dos clases de contribuciones: La

contribución de las modalidades al eje son una guía para la interpretación del eje, mientras que las contribuciones relativas indican que tan bien una modalidad es descrita por el eje. Usualmente, una alta contribución de los puntos a las dimensiones implica también una alta contribución relativa.

Después de la descripción de los ejes del estudio, se debe analizar los planos factoriales que se conforman con la unión de los ejes que se deciden tener en cuenta, por tanto al realizar la lectura de los planos factoriales es importante tomar a aquellos elementos que presentan una buena calidad de representación en el plano. De esta manera se evita interpretar elementos que debido a efectos propiamente de la gráfica distorsione las asociaciones.

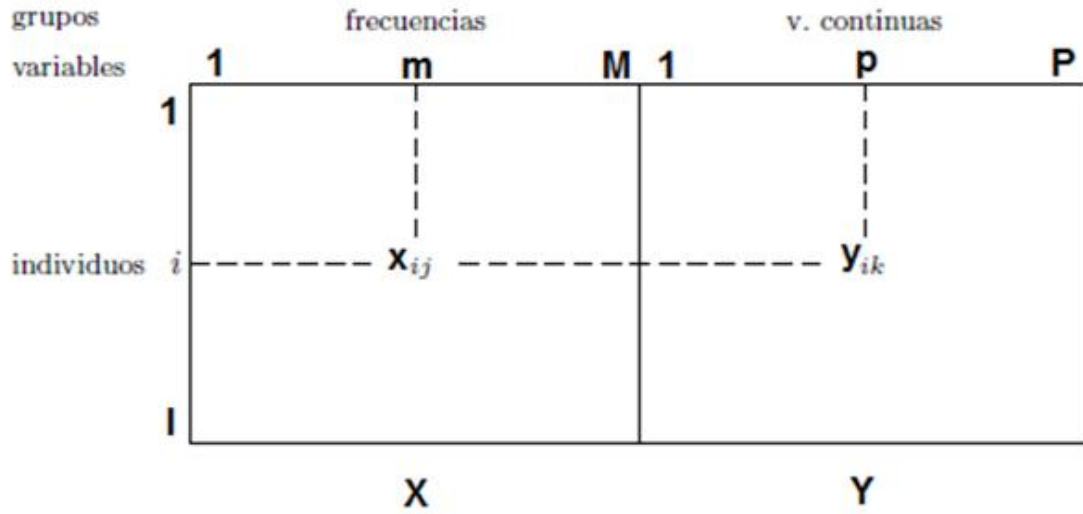
3.2.4.2 Análisis de Correlación Canónico (ACC)

Herramienta creada por Hotelling (1936), utilizada en situaciones donde un conjunto de variables se debe dividir en dos grupos para estudiar la relación existente entre las variables de éstos, así el ACC analiza tablas de frecuencias-variables continuas, es decir, tablas en las cuales las unidades estadísticas están descritas por dos grupos de variables, uno de frecuencias y otro de variables continuas, los datos del grupo de frecuencias pueden ser conteos, porcentajes o respuestas binarias (presencia/ausencia). Además, fue propuesto por Ter-Braak (1986) para estudiar la influencia de las condiciones del medio ambiente en la distribución de las especies de flora y fauna. En Ecología, se estudia la influencia de las condiciones del medio ambiente (grupo de variables continuas) en la distribución de las especies de flora y fauna (grupo de frecuencias) (Chessel et al. 1987, Lebreton et al. 1988, Lebreton et al. 1991, Doledec & Chessel 1991, Birks & Austin 1994, Villalobos et al. 2000, Pavoine et al. 2003, Urbina & Londoño 2003, Bertiet al. 2004).

Se puede considerar al ACC como una generalización del modelo de regresión múltiple; el cual busca establecer la relación entre un conjunto de variables predictoras y un conjunto de variables respuesta. El ACC se propone determinar la correlación entre una combinación lineal de las variables de un conjunto y una combinación lineal de las variables del otro conjunto. Es decir, una combinación U y otra V, tal que la correlación entre éstas sea tan grande como se pueda. El procedimiento de optimización es similar al de componentes principales, excepto que aquí se maximiza la correlación en lugar de la varianza. Las variables U y V reciben el nombre de variables canónicas.

Este análisis está profundamente relacionado con el análisis canónico discriminante y tiene ciertas propiedades análogas al análisis de componentes principales y al análisis factorial, en el que en lugar de tratar de estudiar las dependencias internas entre las variables de un mismo grupo, en el caso de la correlación canónica lo que se estudia es la relación o dependencia entre dos grupos de variables. Sean $X_1, X_2, \dots, X_m$ y $Y_1, Y_2, \dots, Y_p$, dos conjuntos de variables, uno de variables explicativas (independientes) y el otro de variables dependientes (respuesta), respectivamente. Ver figura 2.

Figura 2. Conjunto de datos



Si Z es el vector de variables de tamaño $(1 \times (m + p))$, se puede considerar que éste ha sido particionado verticalmente en la forma $Z = (X \parallel Y)$, donde X contiene m variables y Y las p restantes. Sin pérdida de generalidad, se asume que $E(Z) = 0$.

La matriz de covarianzas del vector Z se particiona en forma análoga a la hecha anteriormente; es decir,

$$\Sigma = \begin{pmatrix} \Sigma_{XX} & : & \Sigma_{XY} \\ \Sigma_{YX} & : & \Sigma_{YY} \end{pmatrix}.$$

Las matrices Σ_{XX} y Σ_{YY} son las matrices de las covarianzas “dentro” de cada conjunto de variables y las matrices Σ_{XY} y Σ_{YX} son las matrices de las covarianzas “entre” los conjuntos. El siguiente esquema muestra la partición

$$\begin{matrix} X_1, X_2, \dots, X_m & Y_1, Y_2, \dots, Y_p \\ \begin{matrix} X_1 \\ X_1 \\ \vdots \\ X_m \\ - \\ Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{matrix} & \begin{pmatrix} (\Sigma_{XX})_{m \times m} & : & (\Sigma_{XY})_{m \times p} \\ & \vdots & \\ & \vdots & \\ & \vdots & \\ & \vdots & \\ \dots & \dots & \dots \\ & \vdots & \\ (\Sigma_{YX})_{p \times m} & : & (\Sigma_{YY})_{p \times p} \end{pmatrix} \end{matrix}$$

El objetivo del análisis canónico es encontrar una combinación lineal de las m variables predictoras X (independientes) que maximice la correlación con una combinación lineal de las p variables respuesta Y (dependientes). Explícitamente, se trata de encontrar entre las siguientes combinaciones lineales:

$$\begin{aligned} U_1 &= X\alpha_1 = \alpha_{11}X_1 + \alpha_{21}X_2 + \cdots + \alpha_{m1}X_m \\ U_2 &= X\alpha_2 = \alpha_{12}X_1 + \alpha_{22}X_2 + \cdots + \alpha_{m2}X_m \\ &\vdots \\ U_r &= X\alpha_r = \alpha_{1r}X_1 + \alpha_{2r}X_2 + \cdots + \alpha_{mr}X_m \end{aligned}$$

Aquella que tenga la correlación más alta con alguna de las siguientes combinaciones lineales:

$$\begin{aligned} V_1 &= Y\gamma_1 = \gamma_{11}Y_1 + \gamma_{21}Y_2 + \cdots + \gamma_{p1}Y_p \\ V_2 &= Y\gamma_2 = \gamma_{12}Y_1 + \gamma_{22}Y_2 + \cdots + \gamma_{p2}Y_p \\ &\vdots \\ V_r &= Y\gamma_r = \gamma_{1r}Y_1 + \gamma_{2r}Y_2 + \cdots + \gamma_{pr}Y_p \end{aligned}$$

Con $r = \min\{m, p\}$. Las combinaciones lineales se escogen de tal forma que: la correlación entre U_1 y V_1 sea máxima; la correlación entre U_2 y V_2 sea máxima con la restricción que estas variables estén no correlacionadas con U_1 y V_1 ; la correlación entre U_3 y V_3 sea máxima sujeta a la no correlación con U_1 y V_1 , U_2 y V_2 , y así sucesivamente. Cada par de variables $(U_1, V_1), (U_2, V_2), \dots, (U_r, V_r)$ representa, independientemente, la relación entre los conjuntos de variables X y Y . El primer par (U_1, V_1) tiene la correlación más alta y es más importante; el segundo par (U_2, V_2) tiene la segunda correlación más alta, y así sucesivamente, el r - *ésimo* par (U_r, V_r) tiene la r - *ésima* correlación más alta (en orden descendente).

En forma resumida se muestra el procedimiento para maximizar la correlación, se describen las correlaciones señaladas como: $U = X\alpha$ y $V = Y\gamma$ respectivamente. La correlación entre U y V está dada por

$$\rho(\alpha, \gamma) = \frac{\alpha' \Sigma_{XY} \gamma}{\{(\alpha' \Sigma_{XX} \alpha)(\gamma' \Sigma_{YY} \gamma)\}^{\frac{1}{2}}}$$

Como $\rho(\alpha, \gamma)$ es invariante por transformaciones de escala sobre α y γ normalizados, entonces, se requiere que α y γ sean tales que U y V tengan varianza uno; es decir, que

$$\begin{aligned} \text{Var}(U) &= \text{Var}(X\alpha) = \alpha' \Sigma_{XX} \alpha = 1 \\ \text{Var}(V) &= \text{Var}(Y\gamma) = \gamma' \Sigma_{YY} \gamma = 1 \end{aligned}$$

Con $E(U) = E(V) = 0$. Entonces, maximizar $\alpha' \Sigma_{XY} \gamma$, con las restricciones $\alpha' \Sigma_{XX} \alpha = 1$ y $\gamma' \Sigma_{YY} \gamma = 1$ por multiplicadores de Lagrange el problema se trata de maximizar

$$\varphi = \alpha' \Sigma_{XX} \gamma - \frac{1}{2} \phi(\alpha' \Sigma_{XX} \alpha - 1) - \frac{1}{2} \mu(\gamma' \Sigma_{YY} \gamma - 1)$$

con φ y μ los respectivos multiplicadores de Lagrange.

Interpretación

El Coeficiente Canónico: muestra la importancia de la variable original para derivar las funciones canónicas. En la interpretación de los coeficientes, la magnitud y la dirección son importantes. Entre más grande sea el coeficiente, más importante será la variable de interés. Las variables con un coeficiente negativo indican una relación inversa; sin embargo, la interpretación de la importancia relativa o su contribución, tomando en cuenta el coeficiente canónico está sujeto a los mismos criterios de los coeficientes de regresión. Es decir, un coeficiente pequeño puede significar que la variable correspondiente es irrelevante o puede ser que está compartiendo su importancia en otra variable por estar correlacionada con ella.

Correlación canónica: muestra la relación entre dos variables. El cuadrado de esta correlación muestra la cantidad de varianza compartida entre las variables dependientes e independientes. Se conoce también como raíz canónica.

La prueba de significancia: se puede usar la de la prueba de F, pero la más comúnmente utilizada es la Chi- cuadrado con un nivel de significancia de 0.05.

Para determinar el número de funciones canónicas, es similar al análisis de factores, la primera función será aquella con mayor varianza en los grupos de las variables, la segunda será aquella con tanta posible varianza, que no tiene en cuenta en la primera función y así sucesivamente.

El círculo de correlaciones en el ACC para las variables continuas se construye buscando en cada eje, las correlaciones entre las variables continuas estandarizadas y las componentes principales estandarizadas de las filas del ACC(U,V). La contribución a la formación de los ejes es nula. La calidad de la representación en el plano se observa visualmente al dibujar el círculo de radio uno en el plano factorial. Los coeficientes canónicos de las variables continuas (componentes principales del ACC en términos de las variables continuas) son los coeficientes de la regresión múltiple ponderada de las coordenadas factoriales estándar de las filas del ACC sobre las variables continuas estandarizadas.

El gráfico donde aparecen los tres objetos de estudio relacionados: individuos, frecuencias y variables continuas. Los elementos (individuos, frecuencias) que participan en el ACC (U,V) se denominan activos, son representados en el gráfico por puntos, al igual que en el biplot con escalamiento tipo 2; las variables continuas se proyectan como elementos suplementarios. La coordenada de la proyección de una variable continua suplementaria en el ACC (U,V) equivale a su correlación con el eje y se representa por flechas desde el centro del gráfico por los coeficientes canónicos.

La prueba de permutación Monte Carlo es una prueba de hipótesis para determinar relación lineal entre frecuencias y variables continuas (Ter-Braak & Smilauer 2002). La hipótesis a contrastar es

Ho: Las columnas-frecuencias no están relacionadas linealmente con las columnas-variables continuas.

Para esta prueba la estadística que se usa es la pseudo-F:

$$pseudo - F = \frac{Inercia(ACC)/S}{\phi^2 - Inercia(ACC)}$$

Donde

$$S = \min\{I - 1, J - 1, K\}$$

Si el *p - valor* es significativo (*p - valor* < α), las columnas-frecuencias están relacionadas linealmente a las columnas-variables continuas.

4. METODOLOGÍA

Este capítulo comprende los métodos propuestos para lograr los objetivos del proyecto. Inicialmente, se incluye una breve descripción del procedimiento de recolección de datos en el estudio macro, el área de estudio y finalmente la metodología estadística propuesta para el desarrollo de los objetivos planteados.

4.1 ÁREA DE ESTUDIO

El estudio se realiza en los municipios de Belén de los Andaquíes, Morelia y Florencia del departamento de Caquetá, ubicado al noroeste de la región Amazónica colombiana, entre los $0^{\circ}42'17''$ de latitud sur y $2^{\circ}04'13''$ de latitud norte y entre los $74^{\circ}18'39''$ y $79^{\circ}19'35''$ de longitud oeste. La precipitación media es de 3.800 mm/año; humedad superior al 80% y temperatura promedio de 25°C al año. La zona de vida es bosque húmedo tropical. Prevalecen los suelos planos con altitudes entre los 200 a 400 m. (Gobernación del Caquetá 2008, Gutiérrez et al. 2004). La mayoría del área se encuentra dedicada a ganadería de tipo extensivo. En esta región de piedemonte¹² se distinguen tres sistemas productivos denominados: Agroforestal, Silvopastoril y Tradicional.



Figura 3. Mapa del Departamento de Caquetá

¹² Unión de la Cordillera de los Andes con la Amazonía.

4.2 RECOLECCIÓN DE LOS DATOS

El trabajo de campo fue apoyado por la comunidad del municipio de Florencia y el equipo de trabajo de Amaz BD Colombia de la Universidad Tecnológica de Pereira. Se realizó durante el mes de junio de 2008. En cada uno de los tres sistemas de producción se seleccionaron nueve fincas y sobre cada finca se escogieron al azar cinco puntos, separados entre sí por una distancia aproximada de 230 m, para un total de 135 puntos de muestreo en toda el área de estudio.

En cada punto de muestreo se empleó el método de TSBF (Tropical Soil Biology and Fertility; (Lavelle 1988, Anderson y Ingram 1993), que consistió en marcar un monolito central de 25 cm por cada lado y examinar en tres profundidades (a nivel del suelo, 10cm y 20 cm de profundidad); a cinco metros al norte (MAD N) y cinco metros al sur (MAD S) se hizo escrutinio de hojarasca y suelo en un área de 1 m² y hasta una profundidad de 10 cm.; a cinco metros al este (FORM E) y cinco metros al oeste (FORM O), se marcó un cuadro de 50 cm de lado al cual se agregó aproximadamente dos litros de formol al 5%; se realizó una búsqueda directa de macrofauna del suelo durante 10 minutos/hombre alrededor del área muestreada.

En total el muestreo comprendió 135 monolitos, 270 muestras de hojarasca, 270 muestras tratadas con formol y 1350 horas/hombre de búsqueda directa. Para efecto de este estudio se analizó únicamente la información de los macroinvertebrados del suelo y la estructura física y química del mismo. Las especies se identificaron con las claves de Watkins, 1976: Palacio, 1999. Los especímenes se preservaron en alcohol al 80%. En total, se analizaron 135 monolitos, que corresponden a los puntos de muestreo en los cuales se tomaron conteos de la abundancia de especies de macrofauna del suelo (variables discretas) y mediciones de las propiedades físico – químicas (variables continuas). Algunas de las variables físico – químicas son medidas al nivel del suelo (0 cm) y las otras en las tres profundidades (0, 20 y 30 cm) como se presenta en la siguiente tabla:

Tabla 4. Variables de estudio

Variables medidas al nivel del suelo	Variables medidas en tres profundidades (0, 10 y 20 cm)
Potasio (K)	Porcentaje de carbono (%C)
Magnesio (Mg)	Stock de carbono
Calcio (Ca)	Porcentaje de nitrógeno (%N)
Aluminio (Al)	Densidad aparente (Da)
pH	
Humedad (Hum)	
NH_4	
Permeabilidad (Penet)	

4.3 ANÁLISIS ESTADÍSTICO DE LOS DATOS

Se presentan los métodos utilizados en el proyecto, iniciando con el análisis exploratorio de la información el cual dará una idea general la distribución de los datos y finalizando con el análisis multivariado de datos que permitirá observar la relación entre las variables biológicas y las físico – químicas.

4.3.1 Análisis exploratorio de datos

Para tener un conocimiento general del estudio se realizó una descripción de los datos, examinando las variables mediante el análisis exploratorio y observar el comportamiento individual y la existencia de diferencias significativas entre los sistemas de producción y las tres profundidades, previo a la aplicación de herramientas estadísticas multivariadas.

Los datos se agrupan en variables continuas las cuales son las mediciones de las propiedades físicas y químicas del suelo, unas medidas en tres profundidades (0, 20 y 30 cm) y otras solo al nivel del suelo (0 cm); y variables discretas, pertenecientes a conteos de especies de la macrofauna encontrada en cada uno de los 135 puntos de muestreo.

La distribución de variables cuantitativas se graficaron mediante box – plot en función del sistema de producción asociado y la profundidad de medición. Además, para el caso de variables físico – químicas se valora su homogeneidad entre los diferentes sistemas de producción, a través de pruebas de hipótesis de igualdad de promedio e igualdad de varianzas. El modelo planteado para la comparación de promedios para variables medidas al nivel del suelo es:

$$y = \mu + \tau_i + \varepsilon_{ij}$$

Donde μ es la media global, y τ_i es un parámetro único que corresponde al efecto del i – *ésimo* sistema de producción y ε_{ij} es un componente del error aleatorio que incorpora todas las demás fuentes e variabilidad del estudio. Para las variables medidas a tres niveles de profundidad se realizó una anova por bloques, ubicando la profundidad como factor bloque, puesto que es algo implícito de estudio que no se modifica. De esta manera se sigue el modelo estadístico:

$$y = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

Donde μ es la media global, y τ_i es el efecto del sistema de producción i – *ésimo*, τ_i es el efecto de la j – *ésima* profundidad del suelo y ε_{ij} es el término del error que sigue una distribución normal e independiente con media cero y varianza σ^2 .

Se evalúan los supuestos de homogeneidad de varianza y normalidad de los errores sobre los residuales a través de las pruebas de Bartlett y Shapiro Wilks.

En caso de no aprobarse el supuesto, se implementan las pruebas no paramétricas de Kruskal Wallis y Friedmann, para los modelos con y sin factor bloque respectivamente.

4.3.2 Análisis multivariado de datos

Se evaluó el efecto del uso del suelo y los sistemas de producción sobre la composición físico – química del suelo y observar relaciones entre los mismos, a través del Análisis de Componentes Principales estableciendo posibles componentes que permitieran ordenar los usos en escala del paisaje y establecer similitudes entre los puntos de muestreo. Con este método se tiene como objetivo, estructurar el conjunto multivariado de variables, mediante la reducción del número de variables, siendo esta reducción los nuevos componentes que son combinaciones lineales de las originales, que contiene la mayor parte de la variabilidad presente en el conjunto inicial. Además, se ingresaran dos variables cualitativas correspondientes a los sistemas de producción y usos del suelo como variables suplementarias. La finalidad del ACP en el presente trabajo, es la de visualizar la relación entre las variables y valorar gráficamente un posible efecto de los sistemas de producción.

Posteriormente, de acuerdo a los antecedentes, se identificó al Análisis de Correlación Canónica, como el método adecuado para relacionar las características de biodiversidad de la macrofauna (variables biológicas) y la composición físico – química del suelo, puesto que es utilizado en situaciones donde un conjunto de variables, que vienen divididas en dos grupos, para estudiar la relación existente entre sus variables de éstos, así el ACC analiza tablas de frecuencias-variables continuas, es decir, tablas en las cuales las unidades estadísticas están descritas por dos grupos de variables, uno de frecuencias (conteos de especies de macrofauna) y otro de variables continuas (mediciones de las propiedades físico – químicas del suelo), como se aprecia en la figura 2.

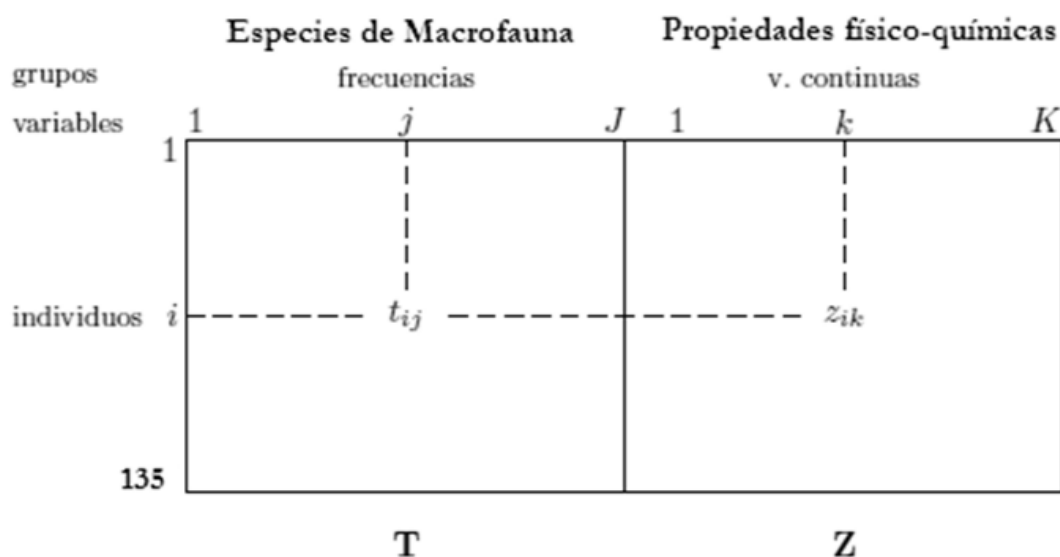


Figura 4. Estructura del conjunto de datos

Por consiguiente, se tiene una estructura de datos adecuada para realizar un ACC y procede a estudiar la relación existente entre estos conjuntos de datos, observando la correlación entre una combinación lineal de las variables biológicas y una combinación lineal de las variables físico – químicas y así, determinar si existe un grado de asociación entre los dos conjuntos de variables, sus sistemas de producción y sus respectivos usos del suelo que se proyectan en el mismo gráfico como variables suplementarias que no intervienen en el proceso pero que enriquecen la presentación, lo cual permite caracterizar los sistemas de producción evaluados en el departamento de Caquetá. En conclusión, se obtiene un diagrama de ordenamiento formado por un sistema de ejes donde se muestran los puntos de muestreo, las especies y variables físico – químicas (Ter Braak & Verdonshot, 1995).

El ACC propuesto por Ter-Braak (1986) para estudios medioambientales, es uno de los métodos que permite estudiar la relación entre un grupo de frecuencias y un grupo de variables continuas sobre un mismo conjunto de individuos. El grupo de frecuencias juega el papel de variables de respuesta y el grupo de variables continuas juega el papel de variables explicativas que son de tipo cuantitativo.

5. RESULTADOS

Se presentarán los resultados obtenidos en la fase de análisis para responder a los objetivos planteados; primero se realiza el análisis exploratorio de datos y finalmente el análisis multivariado.

5.1 ANÁLISIS EXPLORATORIO DE DATOS

En el siguiente capítulo se examinan los datos previamente a la aplicación de las técnicas multivariadas, y así se obtiene un entendimiento básico de las relaciones existentes entre las variables analizadas. Además, se organizan los datos y se evalúan datos ausentes, se identifican casos atípicos, se observa la existencia de diferencias significativas entre los sistemas de producción y entre las tres profundidades de valoración, de igual manera se proporciona una perspectiva razonable para la interpretación de los resultados.

En esta región se distinguen tres sistemas de producción, es decir, que se utilizan tres estrategias para el cuidado y producción del suelo, el sistema Tradicional es el más común, por lo general se usan químicos para controlar plagas, enfermedades y para incrementar la producción, se manejan las hierbas, los cultivos son monocultivos es decir grandes extensiones de una sola planta que se cultiva en este caso pastizal; el sistema Agroforestal, es el más amigable con el medio ambiente, no hay uso de químicos y se busca que sus cultivos estén acompañados por más plantas, árboles que les proporcione sombra, es decir, es un intento por parecerse más al bosque y ser más natural; y el sistema Silvopastoril, es el que tiene ganado, presenta algunos árboles y arbustos ricos en proteína para alimento del mismo ganado y se reemplazan las cercas de alambres por árboles (cercas vivas). Estos sistemas de producción se caracterizan por la predominancia de algunos usos del suelo que a su vez se clasifican en siete tipos: *Pastizal mejorado* (PM), *Pastizal degradado* (PD), *Pastizal con árboles dispersos* (PDSD), *Rastrojo* (RS), *Sistema agroforestal* (SAF), *Cultivo de caucho* (CH) y *Bosque intervenido* (BI). A continuación se presentan las frecuencias de los usos del suelo para cada sistema de producción, éstos se encuentran distribuidos de la siguiente forma:

Tabla 5. Usos del suelo en los sistemas de producción

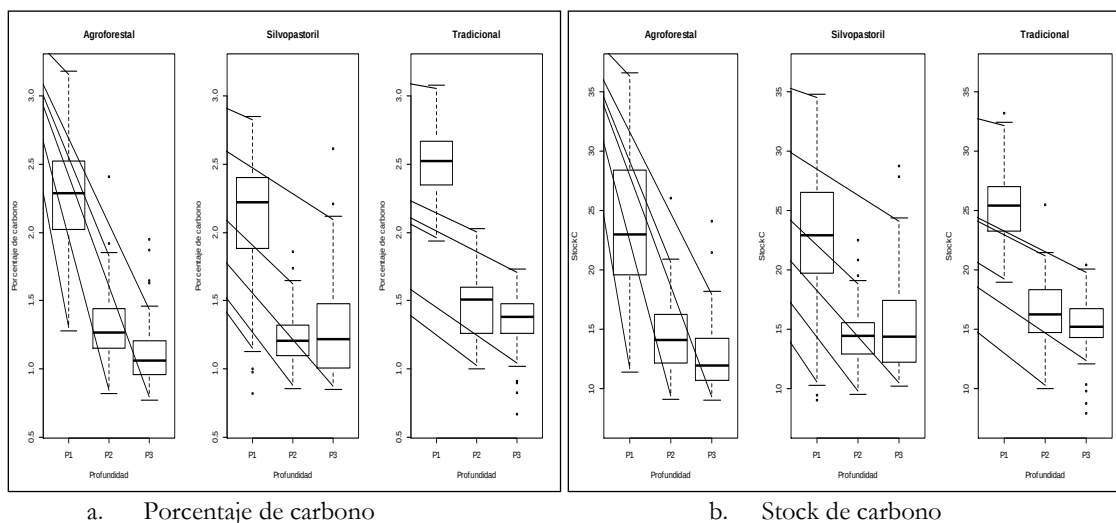
Sistema / Uso	BI	CH	PD	PDAD	PM	RS	SAF	TOTAL
Agroforestal	1	9	9	4	3	4	15	45
Silvopastoril	4	3	16	5	6	6	5	45
Tradicional		2	24	9	7	3		45
TOTAL	5	14	49	18	16	13	20	135

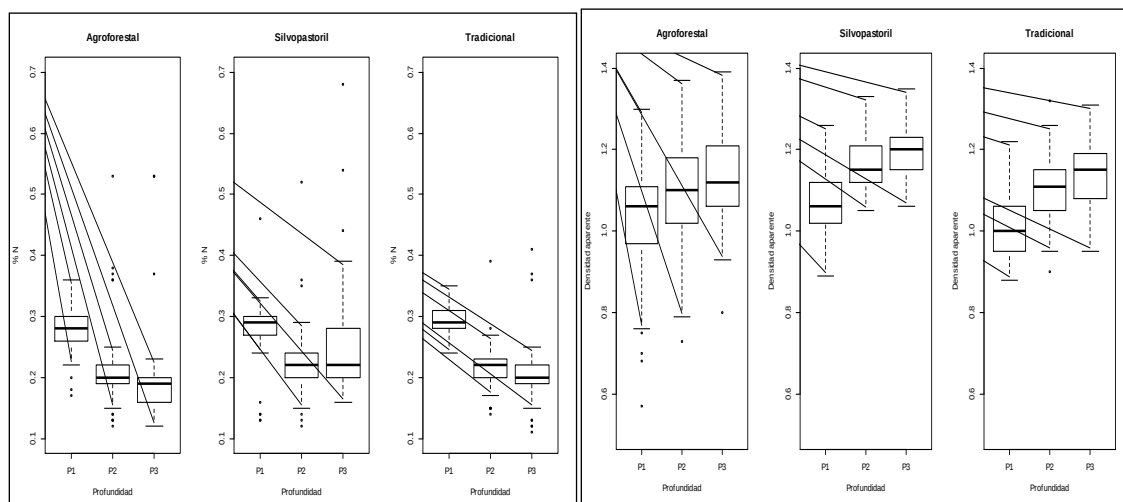
El Sistema Agroforestal lo compone en su mayoría las plantaciones agroforestales, es decir, especies que proveen forraje, cultivos de caucho y un porcentaje menor de pastos nativos, frutales amazónicos y árboles maderables nativos, este sistema está orientado a mejorar la

productividad de las tierras y a su vez ser ecológicamente sustentable, favoreciendo a la conservación de la diversidad biológica dentro de los paisajes, suministrando hábitats y recursos para las especies de animales y plantas; el Sistema Silvopastoril es una combinación natural de componentes leñosos, se caracteriza por tener árboles y pastos nativos, ausencia de árboles forrajeros con actividad ganadera extensiva que controla la vegetación y contribuye al abono de los suelos, vegetación de más de 10 años, existen pequeños cultivos y bosques con extracción de madera, teniendo los bosques un rol protector sobre el ganado, puesto que la sombra disminuye el gasto metabólico de regulación de la temperatura corporal. En el Sistema Tradicional, más del 50% de la extensión de suelo está conformado por el uso Pastizal Degradado, se caracteriza por tener pastos nativos (*Brachiaria sp.*, gramínea que provee una buena protección contra la erosión) sin presencia de árboles forrajeros, un poco de actividad ganadera y la cobertura arbórea representa menos del 10% del total de la superficie muestreada.

5.1.1 Análisis de variables físico - químicas

Los siguientes gráficos proporcionan una visión general del comportamiento de las variables físico – químicas para los diferentes sistemas en cada una de las profundidades del suelo.





c. Porcentaje de nitrógeno d. Densidad aparente
 Figura 5. Boxplot de las variables físico químicas medidas en tres profundidades

La figura anterior presenta la distribución de las variables físico – químicas evaluadas en los tres sistemas de producción a tres profundidades P1, P2 y P3, es decir, 0, 10 y 20 cm respectivamente. Se aprecia que el comportamiento de carbono y nitrógeno es similar en los sistemas tradicional y agroforestal, notándose en la primera capa del suelo, mayor su acumulación, mientras que en los niveles más profundos hay menor presencia de éstos, esto es debido a que estos elementos se encuentra casi en la superficie del suelo. Además, el carbono se vincula con la cantidad y disponibilidad de nutrientes del suelo, y existen prácticas de manejo que generan un detrimento de éste en el tiempo, a la vez hay prácticas que favorecen su acumulación. La densidad aparente es similar entre sistemas, aunque en el sistema agroforestal es más variable, sin embargo, a medida que se hace más profundo, la densidad aparente se incrementa, es decir, a nivel del suelo se encuentra una densidad baja, generalmente, equivale a más porosidad y mayores agregados del suelo, lo que corresponde a mayor estabilidad, menos compactación y, probablemente, mayor contenido de humedad que un suelo con una densidad mayor.

En conclusión, no se observan diferencias entre los sistemas en cuanto a las variables anteriores, pero si estructuras distintas del suelo en su profundidad.

Ahora se presentan las variables físico – químicas que solo se miden al nivel del suelo (0 cm):

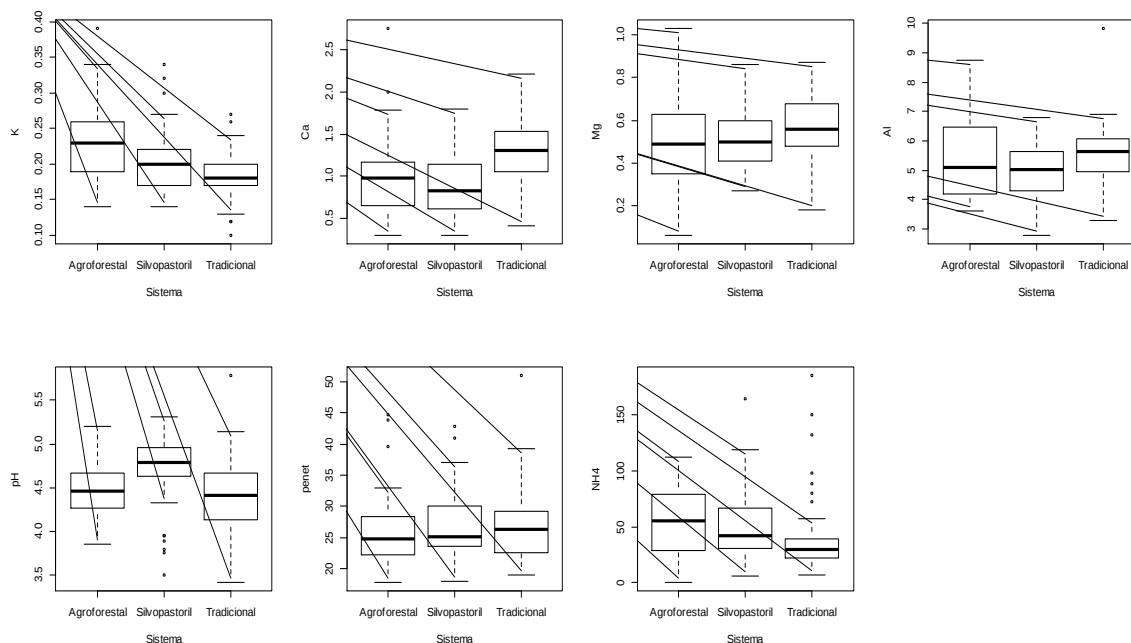


Figura 6. Box – plot variables físico – químicas al nivel de suelo

El sistema tradicional parece tener una acumulación distinta de Aluminio (Al), Magnesio (Mg), Calcio (Ca), Potasio (K) y Amonio (NH_4) comparado con los sistemas Agroforestal y Silvopastoril, es decir, el potasio y el amonio su acumulación es menor y es debido a que son unos de los nutrientes minerales más importantes para las plantas y este sistema de producción es algo deficiente como manejo del suelo; y el calcio, magnesio y aluminio hay mayor concentración de estos elementos, estos son menos demandantes para las plantas. Por otra parte, la capacidad de permeabilidad (Penet) es relativamente similar entre sistemas, además, ésta suele aumentar por la existencia de fallas, grietas, juntas u otros defectos estructurales del suelo. También los factores químicos tienen una influencia directa en la permeabilidad. El pH del suelo es más alto en el sistema silvopastoril y la estructura de los otros dos sistemas es similar, el pH aporta una información de suma importancia en diversos ámbitos de la edafología. Uno de los más importantes deriva del hecho de que las plantas tan solo pueden absorber los minerales disueltos en el agua, mientras que la variación del pH modifica el grado de solubilidad de los minerales. Así, el aluminio es más soluble a un pH bajo, y cuando tal hecho ocurre, pueden ser absorbidos por las raíces, siendo tóxicos a ciertas concentraciones. Por el contrario, determinadas sales minerales que son esenciales para el desarrollo de las plantas, son menos solubles a un pH alto, lo que tiene como resultado que bajo tales condiciones son menos disponibles con vistas a ser absorbidos y nutrir las plantas. De igual forma, en la naturaleza, existen especies vegetales adaptadas a ambientes extremadamente ácidos y básicos. La estructura del suelo se ve influenciada por la naturaleza y la cantidad de iones presentes, es decir, de los elementos que participan directa o indirectamente en todas las actividades hidrodinámicas, químicas y biológicas del suelo.

5.1.2 Análisis de varianza y validación de supuestos

Se presenta una tabla que contiene la validación de los supuestos de normalidad y homogeneidad de varianza para los residuales del modelo de diseño de un factor y diseño de bloques completamente al azar con un solo factor. En la cual se encontró que solo que cumple el supuesto de normalidad para la variable Mg (Magnesio) y el supuesto de homogeneidad de varianza no se cumple para ninguna de las variables, por ello se procedió a realizar pruebas no paramétricas, las cuales no tienen como condición normalidad ni homogeneidad de varianza, a las variables con tres profundidades se les aplicó la prueba de Friedmann asumiendo la profundidad como un factor bloque y a las variables medida a nivel del suelo se les aplicó la prueba de Kruskal Wallis, igualmente para las variables biológicas sólo se utilizaron pruebas no paramétricas, puesto que al ser variables de conteo no es adecuado pensar en su normalidad.

Tabla 6. Valores p para la validación de supuestos y pruebas no paramétricas de variables con 1 profundidad

Variable	Normalidad	Homogeneidad	Anova de un factor	Prueba de Kruskal Wallis
K	0.000	0.002	2.615e-06	0.000
Ca	0.001	0.097	3.705e-06	0.000
Mg	0.510	0.008	0.029	0.029
Al	0.012	0.009	0.013	0.017
pH	0.004	0.077	0.001	0.000
Humedad	0.084	0.048	0.334	0.542

Según la tabla anterior, el resultado de la prueba de Kruskal Wallis indica que la variable humedad es la única que no presenta diferencias significativas entre los sistemas de producción, es decir, la humedad no varía con el uso de suelo. Por tanto, se realizan la comparación por pares utilizando la prueba de rangos de Wilcoxon para las demás variables y establecer entre cuales sistemas presentan las diferencias significativas el valor p es ajustado con el método de bonferroni.

Tabla 7. Valores p para las pruebas post hoc

	Potasio		Calcio		Magnesio		Aluminio		pH	
	Agrof.	Silvop.	Agrof.	Silvop.	Agrof.	Silvop.	Agrof.	Silvop.	Agrof.	Silvop.
Silvopastoril	0.027	-	0.081	-	0.079	-	0.5815	-	0.00034	-
Tradicional	2.1e-05	0.072	0.00016	3.9e-06	0.082	0.052	0.9510	0.0056	0.084	0.00043

Los resultados de las pruebas anteriores se contrastaron con un nivel de significancia del 0.05, se puede observar que la variable K (Potasio) presenta diferencias significativas entre los sistemas agroforestal y silvopastoril, debido a que los valores encontrados en las mediciones de este elemento son más elevados en el sistema agroforestal. Por otra parte, los sistemas

tradicional y agroforestal también presentan diferencias en las medianas, debido a que el sistema tradicional presenta valores más bajos y tiene una desviación estándar menor.

La variable Ca (Calcio) presenta diferencias entre los sistemas tradicional y agroforestal, puesto que los valores de calcio en el sistema tradicional poseen una alta amplitud y son mayores. De igual forma, el sistema tradicional presenta diferencias con el sistema silvopastoril, ya que la concentración de calcio en el sistema tradicional es más alta.

La variable Al (Aluminio) solo presenta diferencias significativas entre los sistemas tradicional y silvopastoril, debido a que la concentración de aluminio es mayor en el sistema tradicional.

La variable pH presenta diferencias significativas entre los sistemas silvopastoril y agroforestal, y entre silvopastoril y tradicional, esto puede ser debido a que el pH del suelo en el sistema silvopastoril es mayor que en el de los otros dos sistemas.

Tabla 8. Valores p para la validación de supuestos y pruebas no paramétricas de variables con 3 profundidades

Variable	Normalidad	Homogeneidad	Anova de bloques	Prueba de Friedmann
%C	9.2E-10	2.6E-07	6.667E-09	0.10
Stock C	7.904E-09	1.211E-09	0.0001073	0.26
%N	< 2.2e-16	2.7E-14	0.01273	0.03
Da	1.4E-06	1.6E-01	1.98E-08	1.98E-05

Se observa en la tabla 8 que las variables porcentaje de Nitrógeno y Densidad Aparente, presentan diferencias significativas, por tanto se evaluará cuáles son los niveles que muestran las diferencias.

Tabla 9. Valores p para prueba no paramétrica de la macrofauna

Variable	Prueba de Kruskal Wallis	Variable	Prueba de Kruskal Wallis	Variable	Prueba de Kruskal Wallis
Lombriz	0.9646	Aran	0.3931	Lhym	0.387
Coleóptera	0.0004842	Opil	8.63E-06	Lcol	0.01113
Hemíptera	0.8131	Term	1.35E-08	Lbla	0.5234
Formidae	0.2567	Ldipt	0.5642	Ldiplo	0.2115
Isóptera	0.177	Llepid	0.5297	Diplop	0.08519
Blattodea	0.3484	Lhemi	0.4018	Quilop	0.09052
Orthop	0.07911	Lorth	0.6544	Hirud	0.1931

La tabla 9, expone los resultados de la prueba no paramétrica de Kruskal Wallis realizada para las variables de la macrofauna (biológicas), en la cual se observa que hay diferencias entre los tres sistemas de las especies de macrofauna: coleóptera, opilión y termita, esto es debido a que

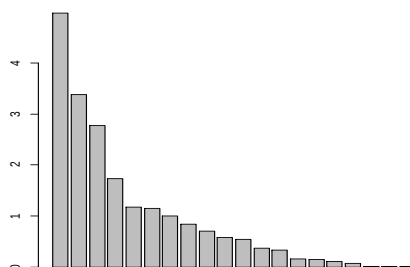
las cantidades encontradas de éstas difieren significativamente entre los sistemas. La especie coleóptera y las larvas de coleóptera se encuentra más abundante en los sistemas silvopastoril y tradicional; los opiliones y termitas abundan más en el sistema agroforestal, es decir, de acuerdo a las condiciones del suelo, se pueden presentar mayor o menor número de especies en el terreno.

5.2 ANÁLISIS MULTIVARIADO DE DATOS

El siguiente capítulo muestra el análisis multivariado con la finalidad de observar las relaciones entre los conjuntos de variables que se estudiaron en el presente trabajo.

5.2.1 Análisis de Componentes Principales

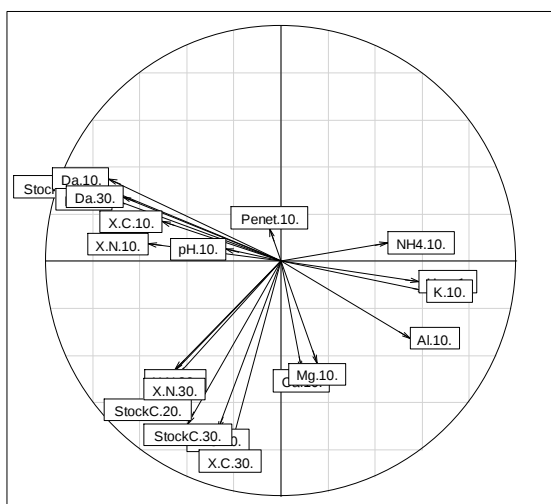
Se analizan los datos de las variables físico – químicas a través de la técnica de Análisis de Componentes Principales (ACP), y así observar las relaciones entre todas las variables y los tres sistemas.



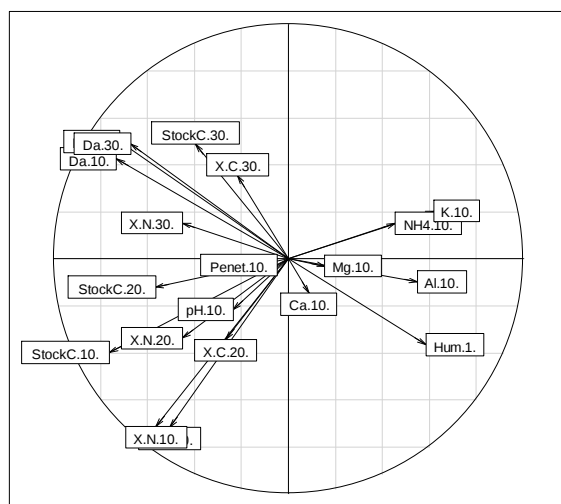
Valor propio	Porcentaje	Porcentaje Acum.
5,0	24,9	24,9
3,4	16,9	41,8
2,8	13,8	55,6
1,7	8,7	64,3
1,2	5,9	70,2
1,1	5,7	75,9

Figura 7. Valores propios ACP

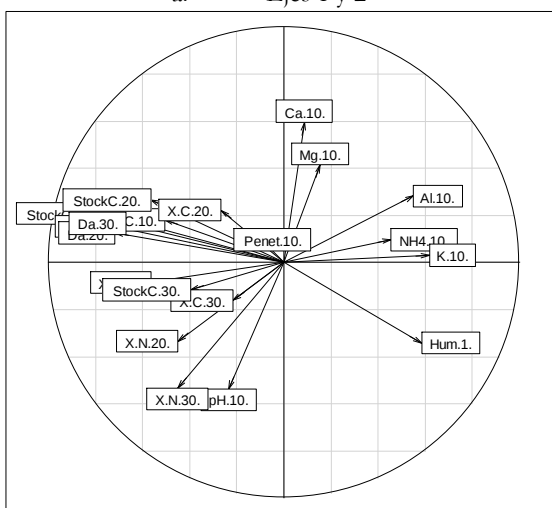
La figura 7, muestra los valores propios de la matriz de correlación que se obtienen a partir de la matriz de datos. Un criterio para escoger el número de componentes es que tengan un valor propio superior a 1 (Díaz, 2007). En este caso se toman las primeras cuatro componentes que explican un 64,3% de la varianza, además, se observa que a partir del 5 componente los valores propios son similares. Sin embargo, la representación gráfica de los ejes se presenta dos dimensiones.



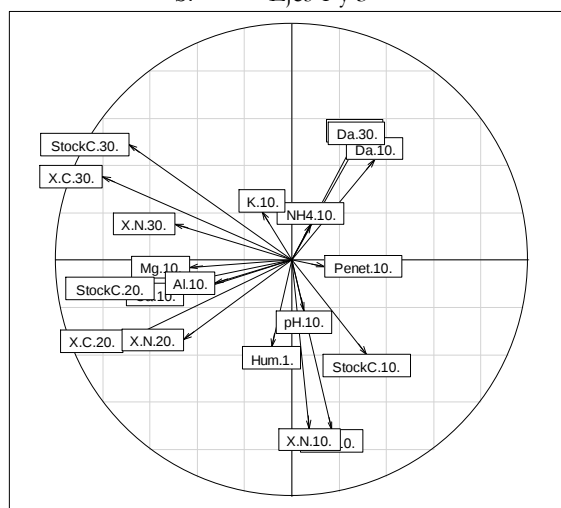
a. Ejes 1 y 2



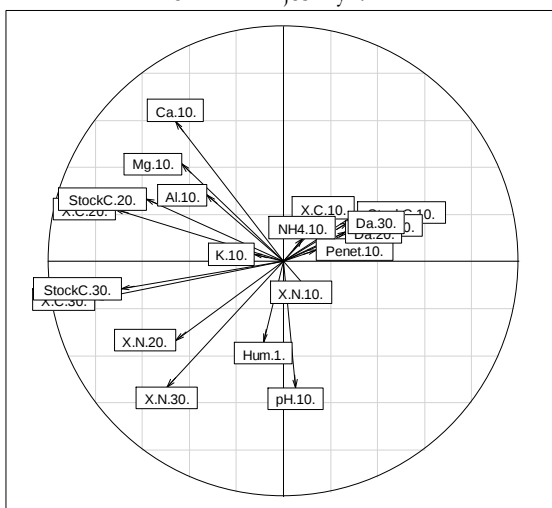
b. Ejes 1 y 3



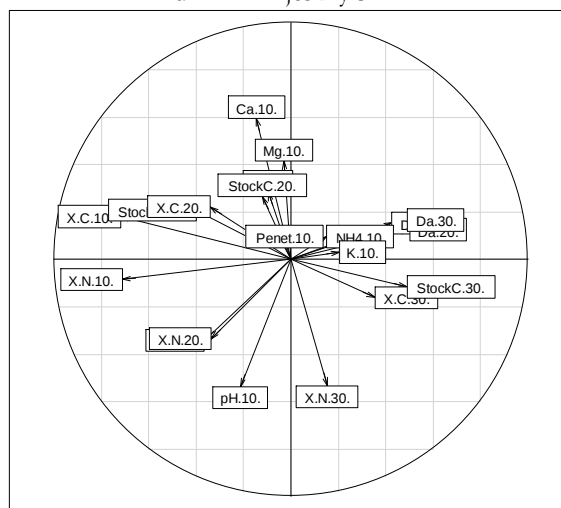
c. Ejes 1 y 4



d. Ejes 2 y 3



e. Ejes 2 y 4



f. Ejes 3 y 4

Figura 8. Círculo de correlaciones

Como se puede apreciar en la figura 8 en el eje 1 y 2, las variables que más contribuyen a la generación de la primera componente son la Densidad Aparente al nivel del suelo, a 10 y 20 cm de profundidad, contribuyendo el -0.73, -0.71 y -0.67 respectivamente y el Stock de Carbono a nivel del suelo, estas variables contrastan con las variables el potasio (0.61), el aluminio (0.54) y la humedad (0.58); es decir, al aumentar este último grupo de variables la densidad aparente disminuye y se encontraría un suelo más poroso lo que indicaría mayor humedad y estabilidad del suelo. Asimismo, esta componente la conforman variables que son medidas a nivel del suelo, en cuanto a la densidad aparente, se puede decir que ésta no tiene una fuerte variación al cambiar el nivel de profundidad.

La segunda componente principal la conforman las variables que medidas en niveles profundos del suelo como porcentaje de carbono (20 y 30 cm de profundidad) y stock de carbono (20 y 30 cm de profundidad), elemento químico de importancia relevante para el nutrimento del suelo. El carbono del suelo se relaciona con la sustentabilidad de los sistemas agrícolas afectando las propiedades del suelo relacionadas con el rendimiento sostenido de los cultivos. El carbono se vincula con la cantidad y disponibilidad de nutrientes del suelo, al aportar elementos como el nitrógeno cuyo aporte mineral es normalmente deficitario.

La tercera y cuarta componente se encuentran compuestas por el porcentaje de carbono y nitrógeno al nivel del suelo; y calcio y pH al nivel de suelo respectivamente, consecuentemente se puede decir que estas dos componentes reúnen las variables que se encuentran en la capa superficial del suelo.

Asimismo, se observa que el pH y la permeabilidad, no son variables que contribuyan de manera significativa al estudio, ver tabla 10.

Por tanto, se definen tres grupos de variables el primero lo componen variables medidas a nivel del suelo (Stock de Carbono, %Carbono, %N, pH y la Densidad aparente); el segundo grupo lo conforman variables medidas a mayor profundidad (Stock de Carbono a 10 y 20 cm, %Carbono a 10 y 20 cm, %Nitrógeno a 10 y 20 cm); y el tercer grupo con mayor presencia de Amonio, Potasio, Aluminio y Humedad.

Tabla 10. Matriz de Coeficientes de las Componentes

	Coordenadas de las componentes				Contribuciones			
	Comp1	Comp2	Comp3	Comp4	Comp1	Comp2	Comp3	Comp4
%C nivel del suelo	-0,50	0,17	-0,72	0,18	5,10	0,80	18,60	1,90
%C 20cm	-0,27	-0,71	-0,34	0,22	1,40	15,00	4,20	2,90
%C 30cm	-0,22	-0,80	0,35	-0,16	0,90	19,00	4,50	1,50
Stock C nivel del suelo	-0,76	0,32	-0,40	0,21	11,70	3,00	5,80	2,50
Stock C 20cm	-0,57	-0,58	-0,12	0,27	6,40	10,00	0,50	4,10
Stock C 30cm	-0,40	-0,69	0,49	-0,11	3,20	14,10	8,70	0,80
%N nivel del suelo	-0,56	0,07	-0,71	-0,08	6,40	0,20	18,30	0,40
%N 20cm	-0,45	-0,46	-0,34	-0,33	4,10	6,20	4,10	6,40
%N 30cm	-0,45	-0,50	0,15	-0,53	4,10	7,20	0,80	16,40
Densidad aparente nivel del suelo	-0,73	0,35	0,43	0,15	10,80	3,60	6,60	1,40
Densidad aparente 20cm	-0,72	0,26	0,50	0,13	10,30	2,10	9,10	0,90
Densidad aparente 30cm	-0,67	0,27	0,49	0,17	9,10	2,20	8,70	1,60
Permeabilidad	-0,05	0,14	-0,03	0,05	0,00	0,60	0,00	0,20
NH4 al nivel del suelo	0,45	0,08	0,15	0,10	4,10	0,20	0,80	0,60
Potasio al nivel del suelo	0,62	-0,13	0,20	0,03	7,70	0,50	1,50	0,10
Calcio al nivel del suelo	0,09	-0,46	-0,15	0,60	0,20	6,20	0,80	20,50
Magnesio al nivel del suelo	0,16	-0,43	-0,03	0,41	0,50	5,60	0,00	9,90
Aluminio al nivel del suelo	0,55	-0,33	-0,10	0,28	6,10	3,10	0,30	4,60
pH	-0,23	0,05	-0,21	-0,54	1,10	0,10	1,70	16,60
Humedad	0,59	-0,09	-0,37	-0,34	6,90	0,20	4,80	6,70

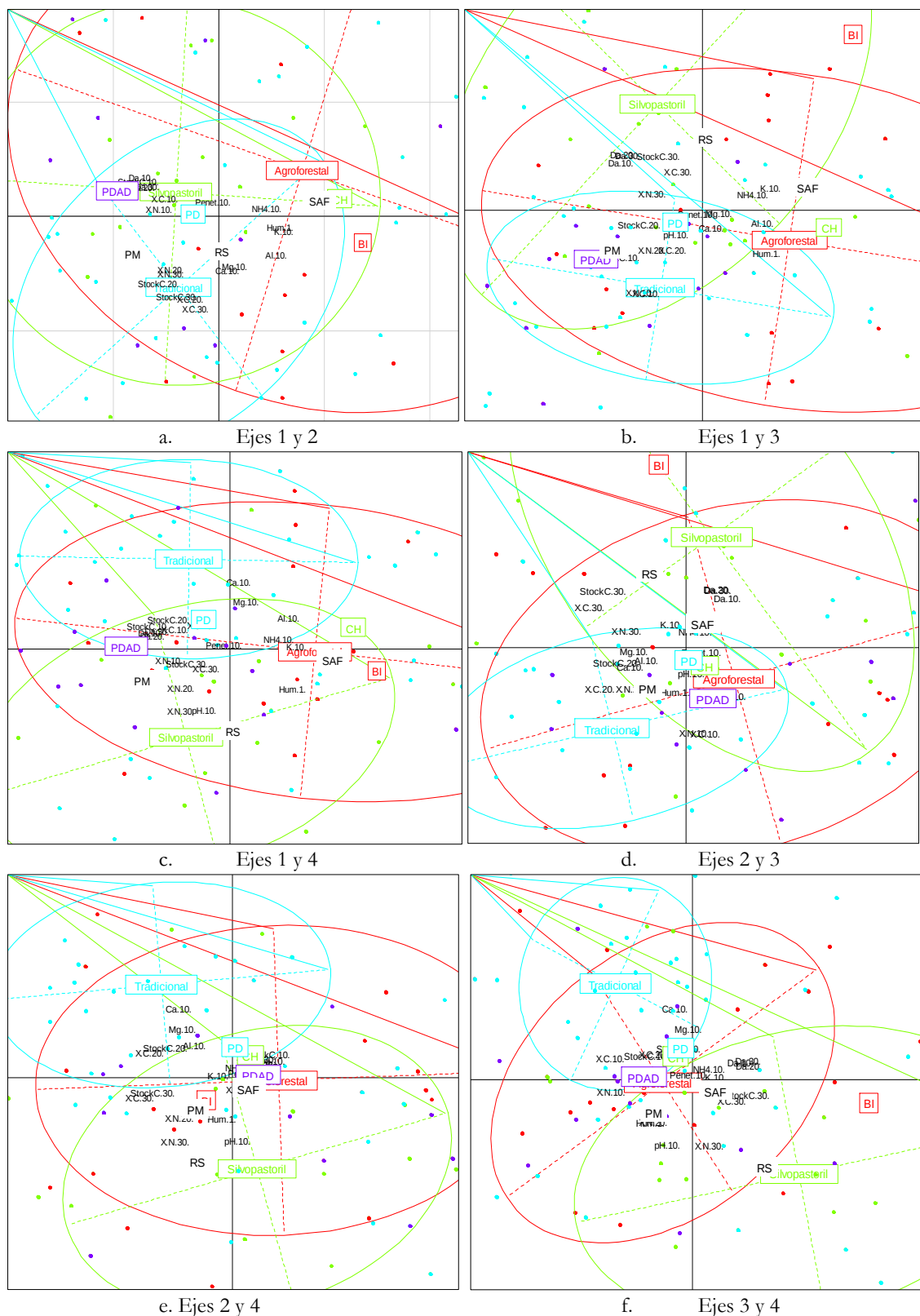


Figura 9. Planos factoriales del ACP

En la figura 9 se han ubicado las propiedades físico – químicas del suelo de acuerdo con sus coordenadas respecto a los primeros cuatro ejes factoriales. Además, se anexan las

variables suplementarias sistemas de producción y usos del suelo, dando lugar a la apreciación de las relaciones existentes entre éstas.

En la variación de los ejes factoriales, se observa que el sistema Silvopastoril lo caracteriza los usos pastizal con árboles dispersos (PDAD) y un poco de pastizal degradado (PD); el sistema Tradicional está representado por los usos pastizal mejorado (PM), rastrojo (RS) y un poco de pastizal degradado (PD), como se puede ver en la figura 9, la asociación entre usos y sistemas no varía significativamente en el cambio de ejes. Sin embargo, cada sistema lo caracterizan ciertas propiedades, el sistema Agroforestal está compuesto por el amonio, la humedad, el potasio y el aluminio, variables que solo se encuentran a nivel del suelo, no obstante hay presencia de carbono, nitrógeno y las demás variables en menor cantidad; el sistema Tradicional está caracterizado por mayor acumulación de carbono, esto no es muy adecuado para el suelo puesto que debilita la presencia de las demás propiedades del suelo; y el sistema Silvopastoril está caracterizado por el carbono, el nitrógeno y la densidad aparente en niveles profundos del suelo, también posee en proporciones adecuadas elementos como el amonio, potasio, lo que conlleva a decir que el sistema tiene una estructura adecuada del suelo en sus niveles profundos; mientras que el sistema agroforestal tiene una estructura completa saludable, puesto que la presencia de varios componentes es lo que hace rico al suelo.

Además, se concluye que el carbono es uno de los elementos principales para el suelo en proporciones adecuadas, puesto que sirve como regulador de los demás procesos y elementos del suelo y se encuentra en los tres sistemas de producción, asimismo, se observa la relación inversa del carbono con el potasio, aluminio y humedad con la densidad aparente.

5.2.2 Análisis de Correlación Canónica

Se utilizó este análisis para examinar el efecto del uso de la tierra con la abundancia de especies de macrofauna y sus propiedades físicas y químicas.

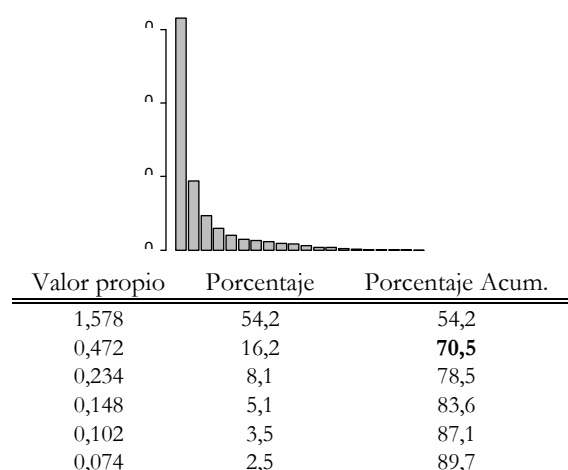
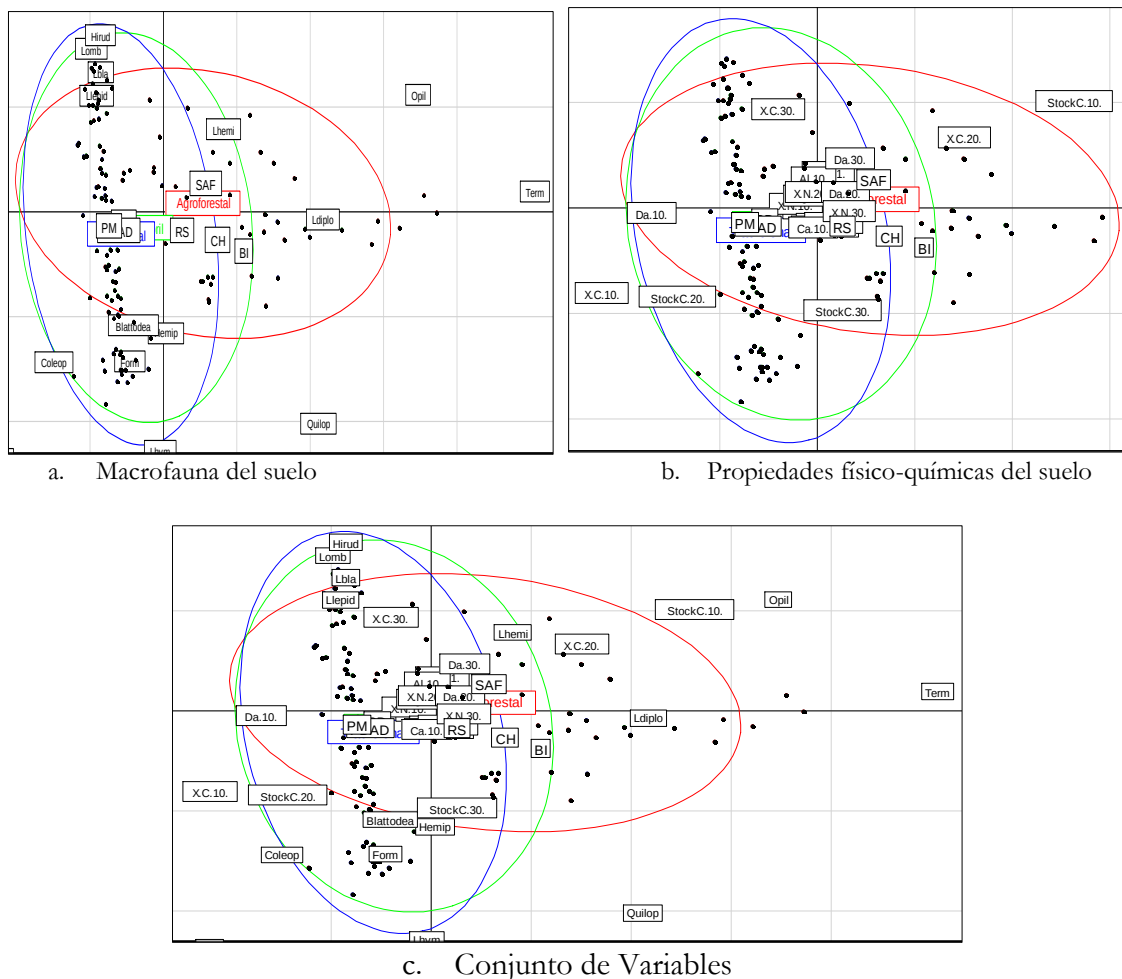


Figura 10.Valores propios ACC

Según el análisis de los valores propios se toman los primeros dos factores que reúnen el 70.5% de la varianza total, ver figura 10. Por consiguiente, los resultados se muestran en el primer plano factorial.



c. Conjunto de Variables
Figura 11. Análisis de Correlación Canónica

En la figura 11 hace referencia a la relación de las variables de la macrofauna del suelo y las propiedades físico – químicas del suelo con los sistemas de producción y usos del suelo con la finalidad de caracterizar los sistemas de producción en Caquetá, se observa que el Sistema Agroforestal lo conforman cultivos agroforestales con ganadería y pasto (SAF) y el cultivo de caucho (CH), asimismo está representado por las siguientes especies: Opiliones, termitas y larvas de diplopoda (milpiés) y hemíptera (chinchés). El Sistema Silvopastoril se singulariza por una diversidad de usos del suelo RS, PM, PDAD, BI, y un poco de CH (cultivo de caucho) y PD (pasto sin forraje), es decir, rastrojo que está compuesto por vegetación de más de 10 años, pastizal mejorado correspondiente a cultivos nativos y ganado extensivo sin árboles forrajeros, pastizal con árboles dispersos y bosque intervenido por extracción de madera; sus especies más representativas son las lombrices, coleópteros (escarabajos), blattodea (cucarachas), Formidae (hormigas), Hemípteros (chinchés) y algunas larvas de hemípteros. El Sistema Tradicional compuesto principalmente por pastizal degradado (PD) o pastos nativos sin presencia de árboles forrajeros y una cobertura arbórea de menos del 10% del total de la superficie, es decir, el uso que le dan a la tierra son sembrados de pasto, las especies de macrofauna que representan este sistema son las lombrices, los hirudínea (sanguijuelas), hormigas, lombrices, larvas de cucarachas y lepidóptera (orugas).

También, se puede apreciar que la mayoría de especies se encuentran muy influenciadas por la acumulación de carbono presente en el suelo, así como con la estructura física del mismo, además, entre mayor porosidad del suelo, mayor humedad y esto conlleva a mejores ciclos de los nutrientes del suelo que hacen posible la riqueza y abundancia de la macrofauna en el suelo, lo que permite mayor proliferación de las especies. Así, se concluye que el carbono es un elemento esencial que permite la activación de otros nutrientes del suelo y hace posible la abundancia de la macrofauna.

El Análisis de Correlación Canónica identificó que los sistemas Tradicional y Silvopastoril tienen una alta asociación, es decir, que presentan características similares en cuanto a la abundancia de especies y propiedades del suelo. Asimismo, el sistema agroforestal comparte algunas características o propiedades del suelo, sin embargo, está caracterizado por la presencia de carbono, aluminio, amonio, potasio, humedad y demás propiedades, lo que lo hace un suelo más rico en nutrientes y especies, además, posee más árboles y más posibilidades de proveer techo y alimento, ya que tiene mayor cobertura vegetal, más fresco, más húmedo.

Igualmente, en la prueba de permutación de monte carlo que se utilizó para determinar la relación lineal entre las variables de macrofauna y las físico – químicas, se obtuvo un $p - \text{valor} = 0.04 < 0.05$, lo que indica que las especies y las propiedades del suelo están relacionadas linealmente como se mostró anteriormente.

6. CONCLUSIONES

En este trabajo se propone y se desarrolla el análisis de dos conjuntos de datos de variables biológicas y físico – químicas en el departamento de Caquetá en el año 2008. El análisis de estas variables permite observar la relación existente entre el conjunto de datos, corroborando que entre mayor número de especies, mejores condiciones de la estructura del suelo en el Caquetá.

En el ACP se evidenció que la Densidad Aparente disminuye cuando el potasio, aluminio y la humedad aumentan, debido a que estas propiedades están altamente correlacionadas, y hacen que haya un adecuado equilibrio de la estructura de suelo en cuanto a su capacidad de filtración de sustancias y mayor estabilidad. También, el sistema Agroforestal está determinado por variables medidas a nivel del suelo, mientras que los sistemas Silvopastoril y Tradicional, lo caracterizan las variables medidas a mayor profundidad, debido a la relación del uso de la tierra.

Se determinó con el Análisis de Correlación Canónico que los sistemas Silvopastoril y Tradicional poseen características similares, puesto que comparten propiedades físico – químicas como el nitrógeno, la densidad aparente y el de mayor importancia el carbono, igualmente el sistema silvopastoril tiene diferentes usos del terreno, entre éstos los que caracterizan el sistema tradicional. También se encontró que el sistema agroforestal es el sistema que más contribuye a la salud del suelo, puesto que la forma del uso de la tierra hace que su suelo sea rico en nutrimentos y por ende es un lugar apropiado para la abundancia de macrofauna, las especies más representativas fueron los murgañes, termitas, orugas, saltamontes cochinillas y milpiés. Es así como se determina que la macrofauna edáfica está altamente influenciada con la estructura del suelo, puesto que entre mayor sea su abundancia, se encontrará un suelo más equilibrado, ello fue confirmado al realizarse la prueba de permutación de monte carlo, la cual definió la relación lineal entre las variables de macrofauna y las físico – químicas, puesto que el $p - valor$ fue inferior a 0.05. Además, el tener un suelo saludable, rico en nutrientes contribuye a la vida de otras especies de animales como aves, peces, mamíferos, anfibios y reptiles, constituyendo un mejor hábitad.

7. RECOMENDACIONES

El análisis de correlación canónica indaga acerca de la relación existente entre las variables ambientales (propiedades físicas y químicas del suelo) con algunas variables sobre especies de animales (macrofauna del suelo). No obstante, se propone realizar el estudio de este conjunto de datos con otro método de análisis multivariado y hacer comparación de los resultados obtenidos, y así mejorar las apreciaciones.

Realizar investigaciones que relacionen la macrofauna del suelo y el restablecimiento de la estructura del suelo con la finalidad de observar como a medida que pasa el tiempo si se realizan prácticas adecuadas de manejo del suelo, su estructura se ve altamente mejorada.

Realizar un análisis de correlación canónica con las variables que más contribuyen a la formación de las componentes.

BIBLIOGRAFÍA

ANDERSON, T W. Asymptotic theory for Principal Component Analysis. Columbia (USA). En: The Annals of Mathematical Statistics, Vol. 34, No. 1. (Mar., 1963); p. 122 – 148.

ANDERSON, J.P.E. Soil Respiration. In: Methods of Soil Analysis, Partz, Chemical and Microbiological Properties. Wisconsin (USA). En: American Society of Agronomy, No. 9 (1982); p. 831 – 871.

BLANCHART, E., *et al.* Regulation of soil structure by geophagous earthworm activities in humid savannas of Côte d'Ivoire. Soil Biology and Biochemistry (1997); p. 149 – 172.

BRYUN, L.A. Ants as bioindicators of soil function in rural environments. Agriculture, Ecosystems and Environment. Armidale (Australia). En: Agriculture, Ecosystems & Environment, Vol 74, Issues 1–3. (June 1999); p. 425 – 441.

COCHRANE; SANCHEZ, F.L. Estudios del uso potencial de la tierra para la evaluación y transferencia de agrotecnología basada en germoplasma: un informe intermedio. In Suelos Ecuatoriales. Medellín (Colombia), Memorias del seminario internacional sobre clasificación del uso potencial de la tierra en el trópico latinoamericano, (1981); p. 29 y anexos.

DE LAS SALAS, Gonzalo. suelos y ecosistemas forestales con énfasis en américa tropical. San José: IICA, (1987); p. 234 – 365.

DIAZ, Luis Guillermo. Estadística Multivariada: Inferencia y Métodos. Bogotá (2007); p. 197 – 376.

FEIJOO, Alexander. Impacto del uso de la tierra en áreas de laderas sobre comunidades de macrofauna del suelo (Caldono, Cauca, Colombia). Palmira, 2001, 196 p. Trabajo de tesis

(Doctor en Ciencias Agrarias con énfasis en Suelos). Universidad Nacional de Colombia. Facultad de Ciencias Agropecuarias. Área de Suelos.

INSTITUTO COLOMBIANO DE NORMAS TECNICAS Y CERTIFICACION. Compendio de normas técnicas colombianas sobre documentación: presentación y elaboración de trabajos y tesis de grado. Bogotá: ICONTEC. 2008 P. 104. (NTC 1486; NTC 1487, NTC 1160; NTC 1308; NTC 1307).

LAURANCE WF. Do edge effect occur over large spatial scales? Trends Ecol Evolut. 2000;15:134-135.

LAVELLE, Patrick. Faunal activities and soil processes: Adaptive strategies that determine ecosystem function. Paris (Francia). Advances in Ecological Research, Vol 27 (1997); p. 93-132.

LEBART, Ludovic y MORINEAU, Alain. Statistique Exploratoire Multidimensionnelle. Paris: DUNOT (1995); p. 32 – 220.

LOZANO-ZAMBRANO *et al.* Hormigas: Relaciones especies-área en fragmentos de bosque seco tropical. En: Neotrop Entomol, Vol 38 (2009); p. 44 – 54.

MALDONADO, G. y VELÁSQUEZ, J. “Determinación de la capacidad de carga y la ganancia de peso de Bovinos en pastoreo de Gramíneas del piedemonte amazónico de Colombia”. Pasturas Tropicales, Vol 16(1994); p. 2 – 8.

MALHI, y BALDOCHI. The carbon balance of tropical, temperate and boreal forests. Plant, Cell and Environment, Vol 22 (1999); p. 715-740.

MONTGOMERY, Douglas. Diseño y análisis de experimentos. México: LIMUSA (2004); p. 60 – 161.

MORENO, Claudia. Manual de métodos para medir la biodiversidad. Universidad Veracruzana. Zaragoza: CYTED, UNESCO Y SEA (2001); p. 19 – 59.

MOYA , Nabor, *et. al.* Comunidad de macroinvertebrados en ríos de La Cuenca Chipiriri con condiciones de referencia y alteradas, estudio preliminar para la bio-evaluación de arroyos. En: Ciencia Abierta Vol 30 (2006); p. 1-27.

SHESKIN, David. Handbook of Parametric and Nonparametric Statistical Procedures. Washington, D.C.: CHAPMAN & HALL (2007); p. 604 – 688.

RUIZ, Nuria. Mise au point d'un systeme de bioindication de la qualite du sol base sur l'etude des peuplements de macroinvertebres. Paris, 2004, 321 p. Trabajo de tesis (Doctor de la Universidad de Paris). Universite Pierre et Marie Curie.

TAPIA, Sandra. Macro-invertebrados do solo e estoques de carbono e nutrientes em diferentes tipos de vegetação de terra firme na amazônia peruana. Manaus, 2004, 135 p. Trabajo de tesis (Doctor en Ciencias Biológicas). Universidad Federal del Amazonas. Área de concentración en Ecología.

VELÁSQUEZ, E. Bioindicadores de calidad de suelo basados en poblaciones de macrofauna y su relación con características funcionales del suelo. Palmira, 2004, 181 p. Trabajo de tesis (Doctor en Ciencias Agropecuarias con énfasis en Suelos). Universidad Nacional de Colombia.

VERTEL, Melba. Comparación entre el análisis canónico de correspondencias y el análisis factorial múltiple en tablas de frecuencias – variables continuas. Bogotá, 2010, 53 p. Trabajo de tesis (Maestría en Ciencias Estadística). Universidad Nacional de Colombia.