



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
POSGRADO EN CIENCIAS DE LA SOSTENIBILIDAD
INSTITUTO DE ECOLOGÍA
POLÍTICA, GOBERNANZA E INSTITUCIONES

**CATEGORIZACIÓN DE ESTRATEGIAS DE PRODUCTIVIDAD AGRÍCOLA:
EL CASO DEL CAFÉ EN MÉXICO**

TESIS
QUE PARA OPTAR POR EL GRADO DE
MAESTRA EN CIENCIAS DE LA SOSTENIBILIDAD

PRESENTA:
KARLA ADRIANA PEÑA SANABRIA

TUTORA PRINCIPAL:

DRA. ANA ELENA ESCALANTE HERNÁNDEZ
LABORATORIO NACIONAL DE CIENCIAS DE LA SOSTENIBILIDAD. INSTITUTO DE ECOLOGÍA, UNAM

MIEMBROS DE COMITÉ TUTOR:

DR. CARLOS DÍAS ÁVALOS
INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y EN SISTEMAS, UNAM

DR. LUIS ANTONIO BOJÓRQUEZ TAPIA
LABORATORIO NACIONAL DE CIENCIAS DE LA SOSTENIBILIDAD. INSTITUTO DE ECOLOGÍA, UNAM

REVISORES:

DR. ALONSO AGUILAR IBARRA
INSTITUTO DE INVESTIGACIONES ECONÓMICAS, UNAM

DR. RODOLFO OMAR ARELLANO AGUILAR
FACULTAD DE CIENCIAS, UNAM



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

AGRADECIMIENTOS ACADÉMICOS

A la **Universidad Nacional Autónoma de México** por darme nuevamente la oportunidad de crecer profesional y académicamente, esta vez en la realización de mis estudios de maestría.

Al **Posgrado en Ciencias de la Sostenibilidad** y a todos los **profesores** que me dieron clases dentro de este, por brindarme una nueva visión del mundo y enseñarme herramientas para poder hacerle frente, de forma conjunta y organizada, a los problemas socioambientales que vivimos actualmente.

Al **Consejo Nacional de Ciencia y Tecnología (CONACyT)** por la beca otorgada que me permitió realizar este posgrado.

Al **Laboratorio Nacional de Ciencias de la Sostenibilidad** del Instituto de Ecología-UNAM, por permitirme realizar mi proyecto dentro de sus instalaciones y fungir como mi segundo hogar durante toda mi maestría.

Al proyecto “**Biodiversidad y ecología funcional de suelos: conservación y resiliencia en sistemas naturales, agroecológicos y agroforestales -CONACyT -Desarrollo Científico para Atender Problemas Nacionales.**

Al **Instituto Nacional de Estadística y Geografía (INEGI)** por brindarme el acceso a la información del Censo Agrícola, Ganadero y Forestal 2007 a través del **Laboratorio de Microdatos.**

A mi tutora, la **Dra. Ana E. Escalante Hernández**, por su apoyo y sus enseñanzas tanto académicas como personales durante un proceso, que fue para ambas: algo nuevo, muy interesante y también lleno de retos, y que supimos llevar a cabo manera conjunta, y del que aprendimos bastante.

A los miembros de mi Comité tutor, el **Dr. Carlos Díaz** y el **Dr. Luis Bojórquez**, por acompañarme y guiarme académicamente desde fases tempranas en la realización de este proyecto de investigación. Así también por sus palabras de aliento que me motivaron en cada uno de los exámenes tutorales.

Al **Dr. Juan A. Arias Del Angel**, por su amistad, sus ideas inspiradoras y todo su tiempo brindado en la realización de este proyecto. Su conocimiento en programación en el software libre R, fueron fundamentales para avanzar en el procesamiento de datos.

Al técnico académico **M.G. Fidel Serrano Candela**, por su tiempo y sus enseñanzas para poder utilizar el software libre QGIS.

A la **M.E. Alejandra Estrada Barón** por su apoyo y sugerencias para la realización de los análisis estadísticos en fases tempranas del proyecto.

Al **Dr. Alonso Aguilar Ibarra** y al **Dr. Rodolfo Omar Arellano Aguilar** por aceptar ser parte de mi jurado y por sus aportaciones en la elaboración del escrito final, que permitieron darle el último toque final a este proyecto.

AGRADECIMIENTOS PERSONALES

A mi familia nuclear, mi mamá *Clara*, mi abue *María de Jesús*, mi papá *Alberto*, mi hermana *Karina* y mi hermano *Andrés*, por todo el apoyo incondicional y la confianza que solo una familia; caótica pero muy unida; puede brindar. Así también a los más peludos de la casa (*Cajeta* y *Brad*) por hacer que todo fuera más pasadero y divertido durante los días de ansiedad y de estrés. Y a mi tortuga *Dixie* para que no piense que siempre la olvido.

A mi novio *Alí*, por estar tan cerca de mí y de mi universo personal, y por hacerme sentir amada y comprendida. Gracias por apoyarme en todo momento, en las buenas y en las malas. Gracias por estar. ¡Logramos otro paso académico juntos ☺!

A mis amigos y compañeros en Lancis: *Karen*, *Gloria*, *Marcelo*, *Valeria*, *Ale*, *Alberto*, *Nancy*, *Morena* y *Mafer*. Gracias por sus observaciones y sus críticas constructivas en los seminarios, porque en verdad aportaron mucho a este proyecto. Y gracias por las pláticas y las risas en las comidas, por los memes, las bromas y los consejos. Es difícil conocer personas tan chidas en tan poco tiempo. Y les agradezco enormemente el hacerme sentir en confianza, y el permitir que viviéramos y enfrentáramos juntos la pérdida de un amigo. En grupo todo ha sido un poco más fácil de sobrellevar. GRACIAS.

A mis amigos de la universidad, *Arely*, *Érika*, *Edgar*, *Elisa*, *Diego*, *Bill*, *Yoalli*, *Gonzalo*, *Hei* y *Moisés*, que a pesar de que en esta fase la distancia se sintió, logramos mantener nuestra amistad a flote. Gracias por los momentos que compartimos juntos, el tiempo dedicado a alguien vale mucho, ya sea presencialmente o por mensajes y llamadas. El cariño que les tengo es enorme.

A mis mejores amigas de toda la vida: *Mess* y *Vale*, que me han demostrado que hay personas que se sienten tan cerca, por más lejos que nos encontremos. Gracias *Mess* por hacer todo lo que ha estado en tus manos por hacerme sentir rodeada de cariño. Gracias *Val* por ser parte de mi vida, yo sé que vienen tiempos mejores para dejar de extrañarnos tanto. En verdad las valoro mucho, gracias por estar conmigo siempre, en toda esta vida llena de cambios y transformaciones.

A mis amigos y compañeros tan listísimos y chéveres del posgrado y con a los que felizmente seguiré tratado en los próximos años. A *Lau*, *Pao*, *Jaz*, *Enrique*, *Rodrigo*, *Balam*, *Aldo*... ¡a todos los miembros de la ANEECS! y en general a toda la “*Mafia de la Sostenibilidad*”, un honor compartir metas, sueños y energía para buscar mejorar este mundo.

A todas aquellas personas que han sido parte de mi vida en estos últimos años y que han externado su apoyo siempre que han tenido la oportunidad, como lo han sido la familia y a las parejas de mis seres queridos.

Y a todos mis familiares, amigos y conocidos que quizá no pongo explícitamente aquí, pero que serán gustosamente invitados a mi examen profesional porque los valoro como personas y me caen muy bien.

Esta tesis te la dedico a ti *Juan*.

Lamento mucho que nuestros caminos se separen en esta fase, pero valoro en demasía todos aquellos momentos que pasamos juntos. Te agradezco todo el apoyo que personalmente me brindaste en el poco tiempo que teníamos de conocernos. Gracias por las risas, los consejos y los mazapanes.

Nos dejas un tremendo vacío a todos aquellos que tuvimos la fortuna de conocerte, pero así también nos dejas un mar de lecciones, aprendimos de ti incluso en el final.

Siempre te recordaré, como alguien cómico, simple, muy inteligente, sumamente elocuente y creativo, bondadoso, una persona directa y honesta, pero a la vez muy sensible. Tan tú.

Prometo llevarte así en mi mente hasta que mi momento llegue y nos volvamos a encontrar.

Mientras tanto,

Hasta pronto.

Gracias por todo ☺

“Había nacido para ser feliz. Durante mi juventud nunca me había afligido la tristeza, y si en algún momento me sentía abatido, contemplar las maravillas de la naturaleza o estudiar lo que de sublime y excelente ha hecho el hombre siempre conseguía interesarme y animarme”

(Frankenstein: o el nuevo prometeo, Mary Shelley)

ANÁLISIS SOCIAMBIENTAL DE LA PRODUCTIVIDAD DEL CAFÉ EN MÉXICO



Karla Adriana Peña Sanabria
Tutora: Ana E. Escalante
Posgrado en Ciencias de la Sostenibilidad

• Su producción emplea a más de **500,000** productores de **14** entidades federativas.

• Alta prestación de **servicios ecosistémicos**

• Su producción ha caído **43.51%** en la última década



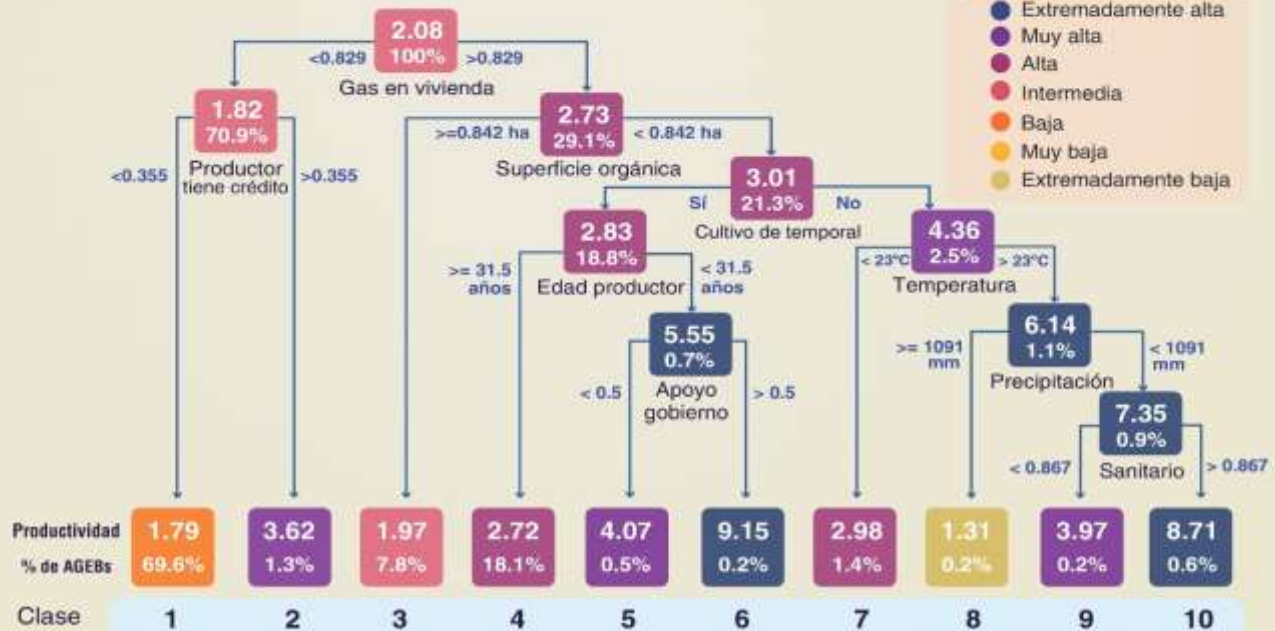
(Moguel & Toledo, 2004; SAGARPA, 2015; SAGARPA, 2017)



puesto que el **90%** de los predios cafetaleros se establecen bajo sombra.



RESULTADOS



El diagnóstico del estado del cultivo de café con base en la categorización de las variables asociadas con las distintas AGEs rurales en el territorio nacional, arroja como determinante estadística de la **productividad**: la **situación socio-económica de los productores, directamente relacionada con el nivel de desarrollo de la región y la infraestructura disponible.**

REFERENCIAS

Moguel, P., & Toledo V.M (2004). Convenir produciendo: Biodiversidad, café orgánico y jardinerías productivas. CONABIO. Biodiversitas 55:1-7.
SAGARPA (2015). Convención Internacional del Café México 2015. Carpeta de Difusión, 22.
SAGARPA (2017). Café Mexicano. Planeación Agrícola Nacional (2017-2030). Ciudad de México, México.

ÍNDICE

| | | |
|--------------|--|-----------|
| I. | RESUMEN..... | 8 |
| II. | INTRODUCCIÓN..... | 9 |
| | 1. Agricultura y retos globales..... | 9 |
| | 2. Agricultura sostenible..... | 10 |
| | 3. Agricultura de escala en el mundo..... | 12 |
| | 4. Agricultura en México..... | 13 |
| | ❖ Política agrícola: en búsqueda del incremento en la productividad..... | 13 |
| | 5. Estrategias estadísticas para el análisis integral del sector agrícola..... | 16 |
| III. | ANTECEDENTES..... | 17 |
| | 1. Generalidades del cultivo de café en México como caso de estudio..... | 17 |
| | 2. Estado actual de la producción de café en México..... | 18 |
| IV. | HIPÓTESIS..... | 19 |
| V. | OBJETIVO GENERAL..... | 19 |
| VI. | OBJETIVOS PARTICULARES..... | 20 |
| VII. | METODOLOGÍA..... | 20 |
| | 1. Consulta y procesamiento de datos en INEGI..... | 20 |
| | ❖ Acceso al Laboratorio de Microdatos..... | 20 |
| | ❖ Procesamiento de información del cultivo de café..... | 20 |
| | 2. Consulta y obtención de información espacial ambiental..... | 23 |
| | ❖ Categorías para tipos de climas y de suelos..... | 25 |
| | 3. Selección de las variables informativas..... | 28 |
| | ❖ Análisis de multicolinealidad..... | 28 |
| | ❖ Modelo de Regresión Lineal Múltiple..... | 28 |
| | ❖ Análisis Factorial Mixto..... | 29 |
| | 4. Clasificación de las AGEB rurales..... | 29 |
| | ❖ Árbol de regresión..... | 29 |
| | 5. Indicador de productividad..... | 30 |
| VIII. | RESULTADOS..... | 32 |
| | 1. Obtención de información relacionada al cultivo de café en México..... | 32 |
| | 2. Reducción de las variables a través de análisis estadísticos..... | 33 |
| | ❖ Análisis de Correlación..... | 33 |

| | | |
|--------------|--|-----------|
| ❖ | Análisis de Multicolinealidad..... | 36 |
| ❖ | Análisis Factorial Mixto (FAMD)..... | 37 |
| 3. | Árbol de regresión como modelo de clasificación de las AGEB rurales donde se cultiva café a partir de su productividad..... | 41 |
| 4. | Mapas..... | 42 |
| IX. | DISCUSIÓN..... | 45 |
| X. | CONCLUSIONES..... | 58 |
| XI. | PERSPECTIVAS..... | 60 |
| XII. | REFERENCIAS..... | 61 |
| XIII. | ANEXOS..... | 69 |
| 1. | Metodología..... | 69 |
| a. | Proceso de agregación..... | 69 |
| b. | Mapa de AGEBs rurales de México en dónde se cultiva café..... | 69 |
| c. | Análisis estadísticos..... | 70 |
| d. | Indicador de productividad..... | 73 |
| 2. | Resultados..... | 74 |
| a. | Variables obtenidas..... | 74 |
| b. | Gráficos de contribución del FAMD..... | 79 |
| c. | Diagnóstico del modelo..... | 81 |

I. RESUMEN

En México el cultivo de café representa una actividad fundamental en el sector agrícola debido a sus contribuciones económicas, culturales y ambientales. No obstante, en años recientes se ha reportado una tendencia decreciente en la productividad cafetalera. En este contexto resulta prioritario evaluar los factores que influyen en la productividad del cultivo de café.

Nosotros proponemos utilizar técnicas y modelos estadísticos para categorizar las estrategias de productividad de las AGEBs rurales en dónde se cultiva café, buscando aportar al desarrollo de un diagnóstico integral que determine los principales factores que afectan la productividad de este cultivo.

La información se obtuvo a través de la consulta y procesamiento de bases de datos del Censo Agrícola, Ganadero y Forestal 2007 de INEGI, y del procesamiento de capas vectoriales obtenidas del Portal de Geoinformación de CONABIO y de INEGI. Posteriormente se realizaron diferentes análisis multivariados para seleccionar las variables informativas de la base de datos. Finalmente se realizó un modelo de árbol de regresión como método de clasificación de las estrategias de productividad.

Como resultado el modelo de árbol de regresión mostró que la variable que mejor clasifica las AGEBs rurales mexicanas en dónde se cultiva café es “*gas en la vivienda*”, seguida por otras 8 variables principales que explican las diferencias de productividad.

Los resultados presentados en este trabajo arrojan como determinante estadística de la productividad, *la situación socio-económica de los productores*, relacionada con el nivel de desarrollo de la región y la infraestructura disponible.

Este modelo podría contribuir al diseño de programas de gobierno que buscan incentivar la productividad del sector cafetalero en México.

II. INTRODUCCIÓN

1. *Agricultura y retos globales*

La agricultura es una forma dominante de gestión de la tierra a nivel mundial, cubre casi el 40 % de la superficie terrestre y utiliza el 70 % del agua dulce del planeta (Altieri & Koohafkan, 2008; FAO, 2009). La actividad agrícola proporciona principalmente alimentos, forrajes y bioenergía, pero también contribuye en el secuestro de carbono, la gestión de cuencas hidrográficas y el hábitat para diversidad de organismos (FAO, 2005; Power, 2010). Además, la agricultura posee un valor no monetario para las personas que se dedican a esta actividad, ya que la consideran una forma de vida, un patrimonio, una identidad cultural y una relación ancestral de respeto con la naturaleza (FAO, 2005).

Históricamente, las actividades agrícolas se relacionan con un bajo desarrollo, bajo crecimiento económico y alta pobreza (FAO, 2005). Son datos ejemplares de este fenómeno el hecho de que, en los países en desarrollo, el sector agrícola emplea alrededor del 70% de los trabajadores, mientras que en los países desarrollados sólo el 3 % (ILO, 2018).

No obstante, dado el crecimiento acelerado de la población humana, la distribución inequitativa de los recursos y la tendencia hacia el consumismo exacerbado, se ha aumentado la presión sobre los recursos naturales y la producción de alimentos y otros productos derivados de la agricultura de manera mundial (Shriar, 2000). Teniendo como consecuencia un gran interés en implementar acciones que ayuden a incrementar la producción agrícola.

De acuerdo con la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO, por sus siglas en inglés), el suministro de alimentos tendrá que crecer otro 60 % para satisfacer las demandas de un mundo poblado por 9,000 millones de habitantes en el año 2050. Al mismo tiempo, los sistemas de producción agrícola enfrentarán una creciente competencia por los recursos con otros sectores (FAO, 2013). En este sentido, las prácticas agrícolas que utilizan fertilizantes químicos y plaguicidas han contribuido significativamente a aumentar la productividad agrícola. Sin embargo, también han aumentado la contaminación química, la degradación del suelo, la pérdida de biodiversidad y han comprometido la calidad del suelo y del agua, ocasionando en su conjunto un impacto negativo en la sostenibilidad ambiental (Singh, *et al.*, 2018). Así también, en las últimas décadas se han perdido tierras agrícolas que antes eran productivas debido a fenómenos producidos por la gestión insostenible de la tierra (desertificación, salinización, erosión del suelo), así como debido a eventos de urbanización

(Nellemann, *et al.*, 2009) y migración de la población rural, que se lleva el conocimiento agrícola histórico a las ciudades.

Por lo descrito anteriormente, existe la necesidad de sistemas productivos agrícolas sostenibles, en cuanto a que mantengan el potencial de la producción futura y no infrinjan un daño inaceptable a largo plazo en otros componentes del ecosistema (NRC, 1993).

2. Agricultura sostenible

La sostenibilidad de un sistema agroalimentario se puede entender como la capacidad de asegurar que los niveles de suficiencia, estabilidad y autonomía alcanzados no implican un deterioro de los recursos naturales que afecte negativamente el sistema el largo plazo y con ello la seguridad alimentaria de las generaciones futuras (Morón & Schejtman, 1997).

De esta manera la agricultura sostenible se puede definir como aquella que cumple los siguientes requisitos: (a) promueve la suficiencia alimentaria, b) conserva los recursos naturales y protege el medio ambiente, c) es viable económicamente, y d) es social y culturalmente aceptable (Rodríguez y Balmón, 2011).

Con el objetivo de impulsar una agricultura sostenible globalmente, en 2015 la Asamblea General de las Naciones Unidas, en la cual participaron 193 países miembros, aprobó la ***Agenda 2030 para el Desarrollo Sostenible*** con 17 objetivos de Desarrollo Sostenible (ODS), entre los cuáles el ODS2 aborda el incremento de la producción de alimentos de una manera sostenible, para hacer frente a la inseguridad alimentaria y la malnutrición en el mundo (FAO, 2015a). En la tabla 1 se desglosa el ODS2 y se resaltan las metas relacionadas con el aumento de la productividad agrícola.

Tabla 1. Objetivo del Desarrollo Sostenible 2

| ODS2 | Descripción | Metas |
|-------------|--|--|
| Hambre Cero | Poner fin al hambre, lograr la seguridad alimentaria y la mejora de la nutrición y promover la agricultura sostenible. | <p>2.1 Para 2030, poner fin al hambre y asegurar el acceso de todas las personas, en particular los pobres y las personas en situaciones vulnerables, incluidos los lactantes, a una alimentación sana, nutritiva y suficiente durante todo el año.</p> <p>2.2 Para 2030, poner fin a todas las formas de malnutrición, incluso logrando, a más tardar en 2025, las metas convenidas internacionalmente sobre el retraso del crecimiento y la emaciación de los niños menores de 5 años, y abordar las necesidades de nutrición de las adolescentes, las mujeres embarazadas y lactantes y las personas de edad.</p> <p>2.3 Para 2030, <u>duplicar la productividad agrícola y los ingresos de los productores de alimentos en pequeña escala</u>, en particular las mujeres, los pueblos indígenas, los agricultores familiares, los pastores y los pescadores, entre otras cosas mediante un acceso seguro y equitativo a las tierras, a otros recursos de producción e insumos, conocimientos, servicios financieros, mercados y oportunidades para la generación de valor añadido y empleos no agrícolas.</p> <p>2.4 Para 2030, <u>asegurar la sostenibilidad de los sistemas de producción de alimentos y aplicar prácticas agrícolas resilientes que aumenten la productividad y la producción</u>, contribuyan al mantenimiento de los ecosistemas, fortalezcan la capacidad de adaptación al cambio climático, los fenómenos meteorológicos extremos, las sequías, las inundaciones y otros desastres, y mejoren progresivamente la calidad del suelo y la tierra.</p> <p>2.5 Para 2020, mantener la diversidad genética de las semillas, las plantas cultivadas y los animales de granja y domesticados y sus especies silvestres conexas, entre otras cosas mediante una buena gestión y diversificación de los bancos de semillas y plantas a nivel nacional, regional e internacional, y promover el acceso a los beneficios que se deriven de la utilización de los recursos genéticos y los conocimientos tradicionales y su distribución justa y equitativa, como se ha convenido internacionalmente.</p> <p>2.a <u>Aumentar las inversiones, incluso mediante una mayor cooperación internacional, en la infraestructura rural</u>, la investigación agrícola y los servicios de extensión, el desarrollo tecnológico y los bancos de genes de plantas y ganado a fin de mejorar la capacidad de producción agrícola en los países en desarrollo, en particular en los países menos adelantados</p> <p>2.b Corregir y prevenir las restricciones y distorsiones comerciales en los mercados agropecuarios mundiales, entre otras cosas mediante la eliminación paralela de todas las formas de subvenciones a las exportaciones agrícolas y todas las medidas de exportación con efectos equivalentes, de conformidad con el mandato de la Ronda de Doha para el Desarrollo.</p> <p>2.c Adoptar medidas para asegurar el buen funcionamiento de los mercados de productos básicos alimentarios y sus derivados y facilitar el acceso oportuno a información sobre los mercados, en particular sobre las reservas de alimentos, a fin de ayudar a limitar la extrema volatilidad de los precios de los alimentos.</p> |

3. Agricultura de escala en el mundo

A nivel mundial la agricultura implementada en menos de 50 ha (pequeña y mediana escala) produce entre 51% y 77% del volumen de los mayores grupos de alimentos para consumo humano: cereales, frutas, legumbres, raíces y tubérculos, y vegetales. Por otro lado, los cultivos de azúcar y aceite, suelen ser producidos por la agricultura a gran escala (>50 ha) (Herrero et al., 2017).

La agricultura a pequeña escala es practicada por familias (incluyendo uno o más hogares) que llevan a cabo principalmente trabajo familiar y que de ese trabajo se derivan gran parte de sus ingresos, en especie o en efectivo (HLPE, 2013). A los agricultores y elaboradores, pastores, artesanos, pescadores, comunidades que dependen estrechamente de los bosques, pueblos indígenas y trabajadores agrícolas en general, que trabajan la agricultura a pequeña escala se les conoce como pequeños productores (CSA,2015).

Actualmente dos mil millones de pequeños productores contribuyen sustancialmente a la seguridad alimentaria a nivel local, regional y nacional; contribuyendo con la producción del 80 % del suministro de alimentos en el mundo (Luna & Wilson, 2015). No obstante, muchos de ellos se encuentran en condiciones de inseguridad alimentaria y malnutrición ya que presentan dificultades en el acceso a los mercados y no logran obtener beneficios que favorezcan medios de vida sanos con una dieta variada y equilibrada (CSA, 2015, Altieri & Koohafkan, 2008). Así mismo, los pequeños agricultores luchan debido a su acceso limitado a insumos (por ejemplo, crédito, tecnología, información) mientras trabajan en tierras de baja productividad ubicadas a grandes distancias de los mercados de producción a través de un sistema de carreteras inadecuado y de alto costo (Luna & Wilson, 2015). Las dificultades se agravan si se trata de mujeres, o de jóvenes debido a la falta de recursos financieros, oportunidades y de conocimientos especializados y capacidades (CSA, 2015).

La viabilidad comercial de la agricultura a pequeña escala representa un desafío crítico para el sistema económico mundial (Luna & Wilson, 2015).

Por lo anterior, queda manifiesta la importancia de los pequeños productores en los esfuerzos para aumentar la producción de alimentos y satisfacer la futura demanda mundial, por lo que las medidas que tengan como objetivo aumentar la productividad de una manera sostenible, tienen que enfocarse necesariamente en estos pequeños productores.

4. Agricultura en México

México, al igual que el resto del mundo, también enfrenta la necesidad de aumentar la producción nacional de alimentos de una manera sostenible para satisfacer la demanda de una población que, de acuerdo con estimaciones del Instituto Nacional de Estadística, Geografía e Informática (INEGI), se duplicará en los próximos años para alcanzar 230 millones de personas en el 2050 (Curiel, 2013). De acuerdo con Baldivia e Ibarra (2017) para 2050 se necesitarán producir casi 289.87 millones ton/año de productos agrícolas en México. Considerando que no existe más tierra de cultivo, el rendimiento de los cultivos debe aumentar de 12.6 ton/ha en 2010 hasta 15.4 ton/ha en 2050. Por otro lado, se estima que el área cultivable podría disminuir por el cambio de uso de suelo para construir viviendas (INEGI-SAGARPA, 2015). Y se suma a este escenario, estimaciones por expertos sobre los efectos del cambio climático que podrían reducir hasta 25% la producción agrícola del país en 2080 (Curiel, 2013). En este escenario, se ha dicho que para aumentar la producción agrícola, más que expandir nuevas tierras de cultivo, los productores se deben enfocar en mejorar el rendimiento por unidad de superficie o productividad, y para lograr esto, se necesita incorporar a la agricultura todas las prácticas agronómicas que ayudan a incrementar la productividad dentro de las que se incluye el uso de: (1) cultivos mejorados; (2) fertilizantes; (3) irrigación, (4) plaguicidas para controlar malezas, plagas y enfermedades; (5) rotación de cultivos; y (6) sistemas de labranza (Sosa Baldivia & Ruíz Ibarra, 2017).

❖ Política agrícola: en búsqueda del incremento en la productividad

México ha transitado por una serie de políticas que han buscado incrementar la producción agrícola a lo largo del tiempo. No obstante, es importante mencionar que existen diferentes perspectivas analíticas sobre la evolución del sector agropecuario en las últimas décadas. Es posible ubicar dos posturas principales opuestas: la primera plantea la existencia de una crisis prolongada del campo y de la economía campesina por falta de fuentes de empleo; mientras que la segunda destaca el éxito del proceso de modernización tecnológica, reestructuración productiva, incremento de la producción y de la productividad de las empresas y de las exportaciones. Los primeros privilegian el análisis de la microeconomía que refleja el bienestar de la población y los segundos se preocupan por la buena salud de la macroeconomía que mide la fortaleza de las empresas. Ambos fenómenos coexisten y son complementarios, por lo que estas dos posturas reflejan parte de la realidad y sus discrepancias tienen que ver con sus perspectivas sobre cuál debe ser el futuro del campo (Rosenweig, 2005; de Grammont, 2010).

Dicho esto, a continuación, se mencionan los eventos que diversos autores consideran importantes para entender el panorama actual de apoyo al campo por parte del Estado mexicano.

En los **años cuarenta**, el gasto agrícola por parte del gobierno se enfocó en grandes inversiones para la infraestructura de riego, así como en insumos y créditos subsidiados que beneficiaron principalmente a granjas comerciales del norte de México (Fox y Haight, 2010). Fue en el sexenio de Ávila Camacho (**1940-1946**) que se dio un impulso decisivo a la introducción de tractores y, en general, a la modernización de los instrumentos de trabajo agrícola: se otorgaban subsidios de hasta 50 % en el precio de la maquinaria agrícola y una parte sustancial de los créditos agrícolas a largo plazo se dirigía a la compra de tractores (Hewitt, 1978, citado por Masera, 1999).

En el periodo de **1940 a 1970** en México se adoptó el modelo de la *Revolución Verde* a fin de obtener mayores rendimientos en la agricultura. De esta manera, el apoyo a la irrigación, el uso de fertilizantes y semillas mejoradas de maíz y trigo influyeron directamente en el crecimiento en la productividad agrícola en el periodo de **1950 a 1965**, con un ritmo promedio anual muy superior al de la década de 1940. Adicionalmente existieron otros fenómenos que impulsaron la producción agropecuaria, como el crecimiento en la demanda de alimentos a raíz del aumento en el ingreso de los mexicanos y las condiciones favorables de demanda internacional durante la posguerra, el conflicto en Corea y el comienzo de la Guerra Fría, los cuales promovieron las exportaciones mexicanas, especialmente las de algodón en los años cincuenta (Sánchez, 1985; Kuntz, 2010).

De **1966 a 1979** se observó un declive del crecimiento agrícola, que no fue generalizado pero que culminó con el papel de la agricultura como uno de los pilares del desarrollo económico de México, y se explica principalmente por la insuficiente producción de maíz y de frijol, así como al desplome de la oferta de algodón, provocada por la reducción de su demanda internacional, a causa del aumento del uso de fibras artificiales en la industria textil (Kuntz, 2010).

Durante el periodo de **1971 y 1982** se presentó un mayor gasto del gobierno para el desarrollo rural, las reformas comenzaron a ampliar el acceso al crédito subsidiado, a insumos, a precios de apoyo y a inversiones en infraestructura rural, incorporando así a un mayor número de campesinos al sistema existente (Fox y Haight, 2010). Alrededor del año **1980** se pretende alcanzar la autosuficiencia alimentaria; sin embargo, un año después, se desmorona el precio internacional del petróleo, ocasionando que el país entrara en recesión económica. Es a partir de esto, que con la intención de dinamizar el sector

agrícola México empieza a adoptar una política Neoliberal¹ y se da una mayor apertura a las importaciones (García, 2010).

En **1986**, México entró al Acuerdo General de Aranceles y Comercio (GATT) para intentar lograr mayores intercambios comerciales. En **1991** crea la dependencia de Apoyos y Servicios a la Comercialización (ASERCA) para ayudar a aquellos productores comerciales que tenían dificultades para comercializar sus cosechas y después para distribuir los pagos compensatorios a los productores de granos en general, incluyendo a los productores campesinos de subsistencia. El programa más conocido de ASERCA es el Programa de Apoyos Directos al Campo (Procampo), un pago de transferencia directa asignado por hectárea y que estuvo oficialmente abierto a todos los productores que habían cultivado granos durante el periodo inmediato anterior al proceso de registro de 1993 a 1994 (Fox y Haight, 2010).

En **1994** entró en vigor el Tratado de Libre Comercio de América del Norte (TLCAN) formado por Canadá, Estados Unidos y México, cuyos principales objetivos han sido incrementar el comercio, la inversión y fortalecer la competitividad internacional, por lo que la política mexicana agrícola se adecuó a las directrices de dichos tratados (Sánchez Cano, 2014). Sin embargo, la mayor parte del sector productivo mexicano, compuesto por pequeños y micro-empresarios, no pudo modernizarse, lo que conllevó a que durante el primer año de funcionamiento del tratado quebraran cientos de unidades económicas (Huerta, 1994).

En **1995**, el gobierno del presidente Zedillo crea *Alianza para el Campo* que ha tenido propósitos productivos y de sanidad de los alimentos, que incluyen al Programa de Desarrollo Rural para cubrir a productores en zonas rurales marginadas (Kuntz, 2010).

En **2001** se creó el programa transversal llamado *Programa Especial Concurrente (PEC) para el Desarrollo Rural Sustentable* cuya función básica es el impulso del sector agropecuario, no sólo en lo que compete a la productividad del mismo, sino en la calidad de vida de la población que depende de las actividades agropecuarias (DOF, 2001; CEF, 2018).

Cabe destacar que los programas de apoyo al campo que se han concentrado en municipios de bajos ingresos han sido pocos; por ejemplo, de los programas de la Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación (SAGARPA), solo *Procampo*, *PROGAN* (pagos de transferencia para ganaderos) y *Fomento al Café* tienen una presencia significativa en dichos municipios. Mientras

¹ **Neoliberal** se refiere a que apoya una gran cantidad de libertad para los mercados, con poco control o gasto por parte del gobierno, y bajos impuestos. (*Diccionario de Cambridge*, En línea en: <https://dictionary.cambridge.org/dictionary/english/neoliberal>).

El **neoliberalismo** es una teoría política y económica que tiende a reducir al mínimo la intervención del Estado (*Diccionario de la Real Academia Española*, en línea en: <https://dle.rae.es/?id=QOIXItD>)

tanto, los programas para la agricultura que promueven la “competitividad” tienden a estar limitados a los productores del norte de México que ya están capitalizados (Fox y Haight, 2010).

De acuerdo al panorama histórico arriba mencionado, consideramos que los esfuerzos del gobierno en materia agrícola se han basado en diagnósticos parciales sobre un panorama agrícola complejo, y por lo tanto existe la necesidad de realizar diagnósticos integrales que informen la toma de decisión en materia de políticas públicas para incentivar la productividad del sector agrícola en México.

5. Estrategias estadísticas para el análisis integral del sector agrícola

Hoy en día una gran cantidad de información que podría ser útil para mejorar la toma de decisiones se encuentra en almacenes de datos, por lo que es de suma importancia lograr extraer y analizar dicha información almacenada. Gracias al crecimiento de los ordenadores y a la facilidad de adquirir y procesar grandes bancos de datos, se ha estimulado el desarrollo de la llamada *ciencia de los datos*.

La ciencia de los datos es la *ciencia de aprender de los datos*; esto es, estudia los métodos involucrados en el análisis y procesamiento de datos y propone tecnología para mejorar los métodos de una manera basada en evidencia (Donoho, 2015). La ciencia de datos es una disciplina científica que está influenciada por la informática, la ciencia de la computación, las matemáticas, la investigación de operaciones, la estadística y las ciencias aplicadas (Hayashi, 1998, Welhs & Ickstadt, 2017).

La estadística es una de las disciplinas más importantes para proporcionar herramientas y métodos para encontrar estructura y proporcionar una visión más profunda de los datos, y es la disciplina más importante para analizar y cuantificar la incertidumbre. Así también, la estadística tiene gran impacto en dos de los pasos más importantes en la Ciencia de los Datos, que es el encontrar la estructura de los datos y el hacer predicciones (Welhs & Ickstadt, 2017).

Una parte de la información puede analizarse y entenderse con estadísticas simples, pero la mayor parte requiere técnicas estadísticas más complejas y multivariadas para convertir estos datos en conocimiento. Centros de investigación relacionados con la industria, el gobierno y las universidades de todo el mundo están haciendo un uso generalizado de estas técnicas (Peña, 2002; Hair *et al.*, 2014; Sánchez, *et al.*, 2018).

El *análisis multivariado* se refiere a todas las técnicas estadísticas que analizan simultáneamente múltiples mediciones en individuos u objetos bajo investigación. Por lo tanto, cualquier análisis

simultáneo que tiene por objeto el estudio estadístico de más de dos variables; medidas en elementos de una población; puede considerarse libremente un análisis multivariado (Hair *et al.*, 2014).

Cuando buscamos describir una realidad compleja en la cual sabemos que existen muchas variables, podemos intentar simplificar identificando las variables que indican, describan y resuman dicha realidad (Peña, 2002). Disponer de variables indicadoras que resuman una “realidad de datos” permite representarlas gráficamente y comparar distintos conjuntos de datos o instantes en el tiempo, y simplifican el análisis al permitir trabajar con un número menor de variables, y si se pueden interpretar, se puede mejorar nuestro conocimiento de la realidad estudiada. De esta forma, el análisis multivariado proporciona métodos objetivos para conocer cuántas variables indicadoras, que a veces se denominan factores, son necesarias para describir una realidad compleja y determinar su estructura (Peña, 2002).

Las técnicas de análisis multivariado tienen aplicaciones en todos los campos científicos, y comenzaron desarrollándose para resolver problemas de clasificación en Biología, posteriormente se extendieron para encontrar variables indicadoras y factores en Psicometría, Marketing y en las Ciencias sociales y han alcanzado una gran aplicación en Ingeniería y Ciencias de la computación como herramientas para resumir la información y diseñar sistemas de clasificación automática y de reconocimiento de patrones (Peña, 2002). Así también se han desarrollado modelos, soportados en diferentes áreas del conocimiento, tales como las matemáticas, la estadística, la computación, entre otras, para develar la información subyacente en grandes bases de datos, en donde la identificación de la información inmersa en ellos se convierte en un proceso más complejo. (Peña, 2002; Ye, 2014; Medrano, 2017, Sánchez, *et al.*, 2018).

Así es que, tanto la estadística exploratoria como la multivariada y la elaboración de modelos a partir de esta, puede posicionarse como una herramienta útil para realizar un diagnóstico integral que permita analizar el sector agrícola en México.

III. ANTECEDENTES

1. Generalidades del cultivo de café en México como caso de estudio.

En el presente trabajo hemos elegido al cultivo de café en México como caso de estudio, debido a que la cafecultura representa una actividad económica fundamental en el sector agrícola en nuestro país. Su producción emplea a más de 500,000 productores de 14 entidades federativas y 480 municipios (SAGARPA, 2017). Por otra parte, su prestación de servicios ecosistémicos es importante no sólo por el área cubierta, sino también porque los cultivos de café frecuentemente están cercanos a áreas prioritarias para la conservación de la biodiversidad, de hecho, se le considera un cultivo de relevancia ambiental,

puesto que el 90% de los predios cafetaleros se establecen bajo sombra (Moguel & Toledo, 1996, SAGARPA, 2015).

En México existen las condiciones climáticas ideales para el cultivo del café, con zonas montañosas del sureste del país que se encuentran a altitudes mayores a 900 metros sobre el nivel del mar, así como temperaturas que van de los 13 a los 26 °C. La variedad que se produce es la denominada “arábiga” (*coffea arábica*) y del tipo robusta (*coffea canephora*). Del total de la superficie de café, el 94.5 % corresponde a las variedades arábicas y un 5.5 % a las robustas, estas últimas ocupadas para la industria de café soluble (FIRA, 2016; SAGARPA, 2017).

La producción de café en México está distribuida en 14 estados cafetaleros: Colima, Chiapas, Estado de México, Guerrero, Hidalgo, Jalisco, Morelos, Nayarit, Oaxaca, Puebla, Querétaro, San Luis Potosí, Tabasco y Veracruz. No obstante, la mayor parte de la producción de café en México se genera en el trópico mexicano, dado que esta región presenta condiciones agroecológicas apropiadas para el desarrollo de cultivos perennes, de modo que se concentra principalmente en Chiapas, Veracruz y Oaxaca, con más del 75% de productores; superficie sembrada, y volúmenes de producción (Robles, 2011; SAGARPA, 2015).

Se estima que en México de 25% a 35% de los predios producen café en sistemas especializados de sombra (policultivo comercial y monocultivo semisombreado), y solamente 10% lo hacen a pleno sol. Por lo que la mayor parte del café; dos terceras partes, se produce en los llamados sistemas tradicionales bajo sombra (rusticano y de policultivo) con una mínima afectación del ecosistema forestal y con un uso nulo o mínimo de agroquímicos (Moguel & Toledo, 2004).

Cabe señalar que la producción del café persiste la heterogeneidad productiva que integra a pequeños, medianos y grandes productores de café, el 80% de la superficie, se concentra en unidades productivas de menos de 10 ha.

2. Estado actual de la producción de café en México.

La producción de café en México se realiza en una superficie equivalente al 3.3 % del área total sembrada en México, por lo que durante 2014 ocupó la sexta posición en este rubro después del maíz grano, los pastos, el sorgo grano, el frijol y la caña de azúcar (SIAP-SAGARPA,2016). En la última década la producción de café ha caído 43.51% (SAGARPA, 2017). En la Figura 1 se muestra la disminución en la producción de café a nivel nacional. En el año 2001 México llegó a ser el quinto país productor de café

más importante (CEFP,2001), no obstante, hoy se encuentra en el onceavo lugar a nivel de producción mundial (SIAP-SAGARPA,2017). Las causas propuestas de esta crisis son la presencia del hongo de la roya y la plaga de la broca, el envejecimiento de los cafetales, la falta de programas y de políticas públicas efectivas, así como los bajos precios internacionales del café. Situación que induce a los productores de café a abandonar la actividad para migrar a otras actividades económicas (Temis et al., 2011, Ortega y Ramírez, 2013).



Fig. 1. Producción de café en México del año 20{04 al 2016. Tomado del Atlas Agroalimentario 2017

Con base a el panorama reciente del cultivo de café en México mencionado arriba, la necesidad de crear un diagnóstico integral que proporcione conocimiento para la toma de decisiones en materia de políticas públicas enfocadas a aumentar la productividad del sector cafetalero mexicano de manera sostenible, se vuelve de suma relevancia. En este sentido se plantean a continuación la hipótesis y los objetivos del presente trabajo.

IV. HIPÓTESIS

Los modelos estadísticos permiten la identificación de patrones para el establecimiento de prioridades en el diseño de políticas públicas que fomenten la caficultura sostenible.

V. OBJETIVO GENERAL

Elaborar un modelo estadístico que integre información social, económica y ambiental que permita la identificación de categorías de productividad en cultivo de café en México en el año 2007.

VI. OBJETIVOS PARTICULARES

- Obtener información relacionada al cultivo de café en México a través de la consulta y procesamiento de datos socioeconómicos de caficultores del país en el Censo Agropecuario 2007 y datos ambientales del Portal de Geoinformación de CONABIO y de INEGI.
- Aplicar métodos estadísticos adecuados para diagnosticar y plantear recomendaciones de gestión en los sistemas cafetaleros mexicanos.
- Construir un mapa para comunicar la información generada.

VII. METODOLOGÍA

1. Consulta y procesamiento de datos en INEGI

❖ *Acceso al Laboratorio de Microdatos*

Para obtener información sobre del cultivo de café en México se solicitó al Instituto Nacional de Estadística y Geografía (INEGI) la consulta y procesamiento de los resultados del Censo Agropecuario 2007 (VII Censo Agrícola, Ganadero y Forestal). La solicitud estuvo enmarcada por un convenio existente entre INEGI y el Instituto de Ecología de la UNAM que permitió el trámite de acceso al Laboratorio de Microdatos.

❖ *Procesamiento de información del cultivo de café*

En las instalaciones del Laboratorio de Microdatos de INEGI se llevó la agregación de la información del cultivo de café a nivel de *AGEB rural*²; por ser la unidad geográfica mínima en la que se permite extraer los datos para proteger información personal de los productores. La información del Censo Agropecuario se proporciona por INEGI en *unidad de producción*³. Para el proceso de agregación y manejo de los datos se utilizó el software R (Versión 3.4.4).

² **Área Geostadística Básica Rural (AGEB rural):** Subdivisión de las áreas geostadísticas municipales que se ubican en la parte rural, cuya extensión territorial es variable y se caracteriza por el uso del suelo de tipo agropecuario o forestal. Contiene localidades rurales y extensiones naturales como pantanos, lagos. Desiertos y otros. Tanto las AGEB rurales como urbanas constituyen la unidad básica del Marco Geostadístico Nacional (*Manual de Cartografía. Censos Económicos 2014, INEGI*).

³ **Unidad de producción:** Conjunto formado por los terrenos, con o sin actividad agropecuaria o forestal en el área rural o con actividad agropecuaria o forestal en el área urbana, ubicados en un mismo municipio los animales que se posean o críen por su carne, leche, huevo, piel, miel o para trabajo, independientemente del lugar donde se encuentren; así como los equipos, maquinarias y vehículos destinados a las actividades agrícolas, pecuarias o forestales; siempre que durante el

A continuación, se desglosa el procesamiento de datos llevado a cabo en INEGI. Y en la Figura 3, se muestra un diagrama con el procedimiento resumido:

- a) Se creó una *tabla de identificación* con las unidades de producción en las que se realiza el cultivo de café, con la finalidad de filtrar únicamente la información de dichas unidades, para ello se utilizaron dos bases de datos del Censo Agropecuario 2007: “TRD_TERRENOS” y RD_AGRICULTURA_PERENNES_CAFE”.
- b) La *tabla de identificación* generada se utilizó para seleccionar las unidades de producción en dónde se lleva a cabo el cultivo de café para cada una de las 19 bases de datos que se utilizaron del Censo Agropecuario 2007, en dónde se encontraban las variables de interés, generándose de esta manera una *tabla de variables*, principalmente relacionadas al manejo del cultivo y características socioeconómicas del productor. En la Tabla 2 se muestran los nombres de las bases de datos consultadas para crear tanto la *tabla de identificación* como la *tabla de variables*. La totalidad de la información del Censo Agropecuario que se encuentra en el Laboratorio de Microdatos de INEGI se pueden consultar en documento de Excel “Descriptor de Archivos (FD)” en <https://www.inegi.org.mx/programas/cagf/2007/default.html#Microdatos>.

Tabla 2. Nombres de las bases de datos consultadas del Censo Agropecuario 2007

| 1 | TRD_TERRENOS | 12 | TRD_OTRAS_SUPERFICIES |
|----|----------------------------------|----|---------------------------------------|
| 2 | TRD_TENENCIA | 13 | TRD_SUPERFICIE_BOSQUE_SELVA |
| 3 | TRD_DERECHOS | 14 | TRD_TECNOLOGIA_FORESTAL |
| 4 | TRD_AGRICULTURA_PERENNES | 15 | TRD_VENTA_FORESTAL |
| 5 | TRD_INTERCALADOS_PERENNES | 16 | TRD_CREDITO_SEGURO_APOYOS |
| 6 | TRD_TECNOLOGIA_AGRICOLA | 17 | TRD_ORGANIZACION_MANEJO_UP |
| 7 | TRD_AGRICULTURA_BAJO_CONTRATO | 18 | TRD_ORGANIZACION_ENTRE_PRODUCTORES |
| 8 | TRD_NO_SEMB_SUPERF_CAUSA | 19 | TRD_CAPACITACION_ASISTENCIA_TECNICA |
| 9 | TRD_RIEGO_SUPERF_TIPO_FUENTE | 20 | TRD_ACTIVIDAD_PROBLEMATICA_PRINCIPAL |
| 10 | TRD_CONSTRUCCIONES_INSTALACIONES | 21 | TRD_CARACTERISTICAS_SOCIODEMOGRAFICAS |
| 11 | TRD_DESTINO_PROD_AGRICOLA | | |

periodo de marzo a septiembre del 2007, todo esto se haya manejado bajo una misma administración (*Definiciones básicas del Cuestionario Censal del Censo Agropecuario 2007*).

Como resultado se obtuvo una tabla con las unidades de producción de café en filas y las variables acomodadas en columnas, a la que llamamos *tabla de variables*.

- c) La tabla de variables sirvió como insumo para la agregación de los datos a nivel de AGEB rural y anteriormente reportados por unidad de producción. La agregación se llevó a cabo utilizando el promedio para las variables numéricas y la frecuencia para cada una de las respuestas de las variables categóricas; y como resultado quedo una *tabla de agregación* (En la Figura A1 de la sección ANEXOS se muestra un ejemplo del proceso de agregación de los datos).
- d) Finalmente, se solicitó la revisión de confidencialidad de la *tabla de agregación* al Laboratorio de Microdatos para poder extraer la información del cultivo de café de las instalaciones de INEGI y así continuar con el análisis y procesamiento de los datos en el Laboratorio Nacional de Ciencias de la Sostenibilidad, del Instituto de Ecología, UNAM.

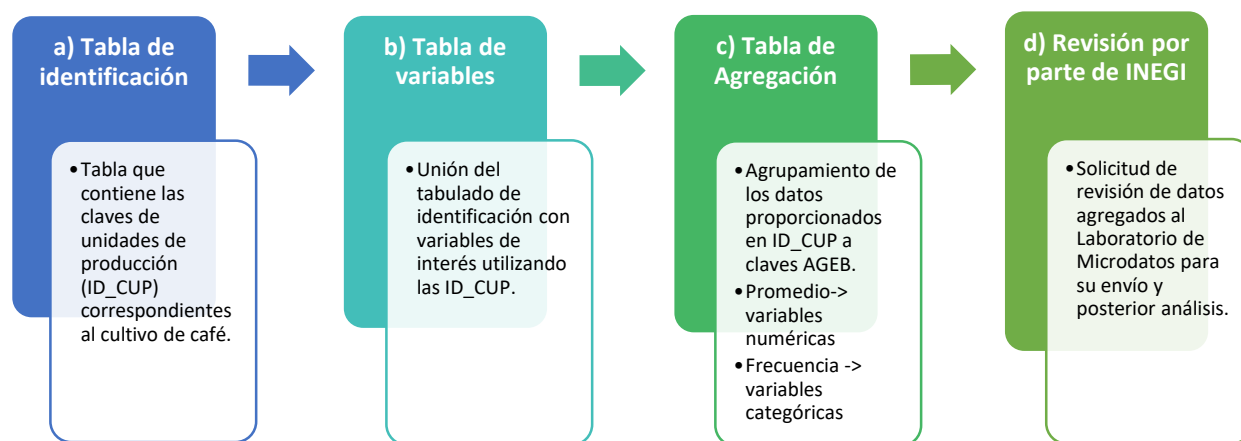


Figura 3. Secuencia de pasos para el procesamiento de datos en INEGI.

En la siguiente liga https://github.com/LANCIS-escalante-lab/TESIS_cafe_Karla se encuentra el repositorio con el script utilizado para todo el procesamiento de datos llevado a cabo en las instalaciones de INEGI, llamado “1_PROCESAMIENTO_DATOS_INEGI”.

2. Consulta y obtención de información espacial ambiental

Se integró información de variables ambientales a la base de datos (tabla de agregación) utilizando capas con datos vectoriales⁴ de mapas disponibles en las plataformas en línea de INEGI (<https://www.inegi.org.mx/>) y CONABIO (<http://www.conabio.gob.mx/informacion/gis/>). Los mapas utilizados fueron los siguientes:

- Conjunto de Datos Vectorial Edafológico, Escala 1:250 000 Serie II (continuo nacional)
- Climas
- Pendiente
- Curvas de nivel para la República Mexicana
- Precipitación anual en México (1910-2009)
- Temperatura media anual en México (1910-2009)
- Uso Potencial del suelo
- Marco geoestadístico 2007 (Censos Agropecuarios 2007)

El procesamiento de información geográfica de las capas vectoriales se realizó con los softwares QGIS (Versión 2.18.25) y R (Versión 3.4.4) y consistió en obtener datos ambientales que corresponden a las AGEBs rurales de café.

- ❑ Las variables obtenidas fueron: suelo, pendiente, textura del suelo, uso potencial del suelo, clima, temperatura y precipitación.

⁴ El *Dato Vectorial* proporciona una manera de representar “objetos espaciales” del mundo real dentro de un ambiente SIG y para ello utilizan geometría (puntos, polilíneas y polígonos). Los objetos espaciales vectoriales tienen *atributos*, que consiste en texto o información numérica que describe los objetos espaciales (en este caso los atributos corresponden a las variables). Más información en: https://docs.qgis.org/2.14/es/docs/gentle_gis_introduction/vector_data.html

A continuación, se desglosa el procesamiento de la información ambiental. Y en la Figura 4 se muestra el diagrama a manera de resumen.

- a) Se unió la capa vectorial “AGEB_rural_2007” del *Marco geoestadístico 2007 (Censos Agropecuarios 2007)* con una copia de la *tabla de agregación* (creada previamente con datos del Censo Agropecuario 2007 en INEGI). Esto se realizó a partir de la clave de identificación de las AGEB rurales (CVEGEO). De esta forma se pudieron localizar espacialmente en un mapa de la República Mexicana las AGEB rurales en dónde se cultiva café (el mapa se encuentra en la Figura A2 de la sección ANEXOS). Esto dio como resultado la capa vectorial nombrada *capa de identificación*.
- b) Se cargaron en el programa QGIS las capas vectoriales que poseían las variables ambientales de interés. Y se unieron una por una a la *capa de identificación*, esto a través de las herramientas del software. Para ello, inicialmente se convirtieron a capas de datos ráster⁵ y posteriormente se utilizó el complemento de análisis llamado “*Estadísticas de zona*” para calcular el *promedio*⁶ o la *mayoría*⁷ de los valores correspondientes a los píxeles que están adentro de cada de las AGEB rurales. Esto último dependiendo de si se trataba de una variable era numérica (ej. precipitación y temperatura) o categórica (ej. textura del suelo, clima). Como resultado, se obtuvo una tabla de atributos (base de datos) compuesta por la *capa de identificación* y todas las variables ambientales.
- c) Se exportó la tabla de atributos del software QGIS en formato .csv y se añadieron las columnas correspondientes a las variables ambientales a la *tabla de agregación*, para crear la *tabla final* o base de datos final, que poseía la totalidad de las variables a analizar.

⁵ Los datos ráster se componen de una matriz de píxeles (también llamadas celdas). Cada píxel representa una región geográfica, y el valor en ese píxel representa alguna característica de dicha región. Más información en: https://docs.qgis.org/2.14/es/docs/gentle_gis_introduction/raster_data.html

⁶ *Promedio*: la media aritmética de los valores de píxeles involucrados en el polígono de la capa vectorial.

⁷ *Mayoría*: obtiene el valor de píxel más representado en el polígono de la capa vectorial.

Para más información: https://docs.qgis.org/2.14/es/docs/user_manual/plugins/plugins_zonal_statistics.html

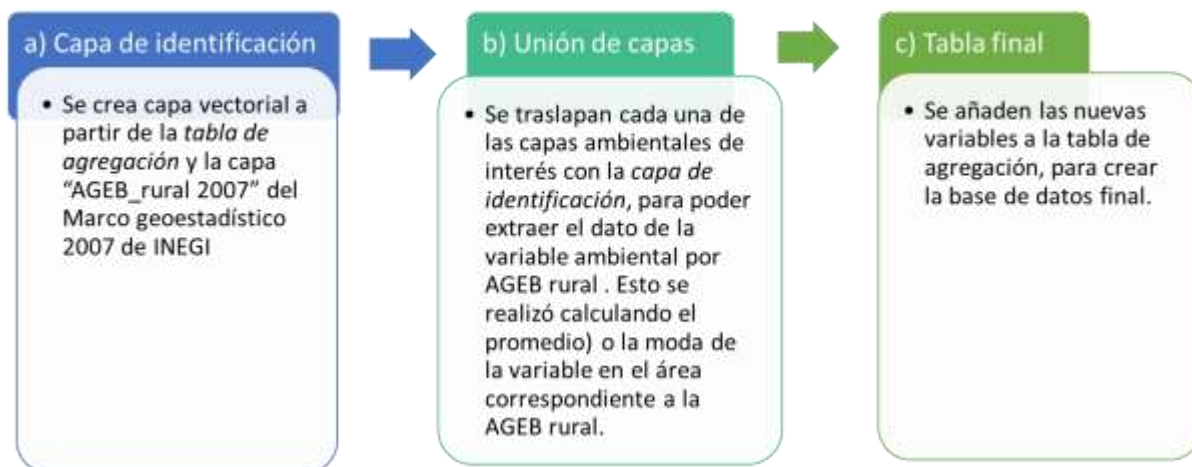


Figura 4. Diagrama de pasos realizados durante el procesamiento de la información ambiental con el programa QGIS.

❖ *Categorías para tipos de climas y de suelos*

Para poder realizar los análisis estadísticos en el Software R, las variables ambientales categóricas: *clima* y *suelo* fueron agrupadas para reducir el número de alternativas que presentaban cada una (tipos de climas y tipos de suelos). El script de R utilizado en este paso, se encuentra en el documento "2_AGRUPACIÓN_CLIMA_SUELO" dentro del repositorio: https://github.com/LANCIS-escalante-lab/TESIS_cafe_Karla.

Para la variable *clima* se presentaban inicialmente 37 tipos de clima en las AGEB rurales dónde se cultiva café, de los 71 que se manifiestan en México de acuerdo a la clasificación de Köppen, modificada por Enriqueta García. Estos 37 tipos de clima fueron agrupados en categorías superiores, con base en la temperatura y posteriormente en la humedad, por lo que al final se obtuvieron 11 tipos de clima (Tabla. 3).

Tabla 3. Categorías de tipos de clima presentes en las AGEB en dónde se cultiva café en México.

| | Código de clima | Categoría de Temperatura | Categoría de Humedad | Categorías |
|----|-----------------|--------------------------|----------------------|-------------------------|
| 1 | (A)C(fm) | semicálido | húmedo | 1) Semicálido húmedo |
| 2 | (A)C(m) | semicálido | húmedo | |
| 3 | (A)C(m)(f) | semicálido | húmedo | |
| 4 | (A)C(w1) | semicálido | subhúmedo | 2) Semicálido subhúmedo |
| 5 | (A)C(w2) | semicálido | subhúmedo | |
| 6 | (A)C(wo) | semicálido | subhúmedo | |
| 7 | (A)C(wo)x' | semicálido | subhúmedo | |
| 8 | A(f) | cálido | húmedo | 3) Cálido húmedo |
| 9 | Am | cálido | húmedo | |
| 10 | Am(f) | cálido | húmedo | |
| 11 | Aw1 | cálido | subhúmedo | 4) Cálido subhúmedo |
| 12 | Aw1(x') | cálido | subhúmedo | |
| 13 | Aw2 | cálido | subhúmedo | |
| 14 | Aw2(x') | cálido | subhúmedo | |
| 15 | Awo | cálido | subhúmedo | |
| 16 | Awo(x') | cálido | subhúmedo | |
| 17 | BS1(h')w | cálido | secos | 5) Cálido seco |
| 18 | BSo(h')(x') | cálido | secos | |
| 19 | BSo(h')w | cálido | secos | |
| 20 | BW(h')w | cálido | secos | |
| 21 | BS1h(x') | semicálido | secos | |
| 22 | BS1hw | semicálido | secos | 6) Seco semicálido |
| 23 | BSohw | semicálido | secos | |
| 24 | BS1k(x') | templado | secos | |
| 25 | BS1kw | templado | secos | 7) Seco templado |
| 26 | BSokw | templado | secos | |
| 27 | C(f) | templado | húmedo | 8) Templado húmedo |
| 28 | C(m) | templado | húmedo | |
| 29 | C(m)(f) | templado | húmedo | |
| 30 | C(w1) | templado | subhúmedo | 9) Templado subhúmedo |
| 31 | C(w1)x' | templado | subhúmedo | |
| 32 | C(w2) | templado | subhúmedo | |
| 33 | C(w2)x' | templado | subhúmedo | |
| 34 | C(wo) | templado | subhúmedo | |
| 35 | Cb'(m)(f) | semifrío | húmedo | 10) Semifrío húmedo |
| 36 | Cb'(w2) | semifrío | subhúmedo | 11) Semifrío subhúmedo |
| 37 | Cb'(w2)x' | semifrío | subhúmedo | |

Por otra parte, la variable *suelo* tenía 24 tipos de suelos para las AGEb en dónde se cultiva café, de los 48 tipos de suelo que se presentan en el territorio mexicano, de acuerdo al Sistema Internacional Base Referencial Mundial del Recurso Suelo (WRB, por sus siglas en inglés). Los niveles de la variable *suelo* se redujeron a 7 categorías utilizando la “Guía simplificada de los grupos de suelo de referencia de WRB” (FAO, 2015b) (Tabla 4).

Tabla 4. Categorías utilizadas de tipos de suelo presentes en las AGEb en dónde se cultiva café en México.

| | Código | Tipo de Suelo | Categorías |
|-----------|---------------|----------------------|---|
| 1 | AC | Acrisol | 1) Suelo con subsuelo enriquecido con arcilla |
| 2 | AL | Alisol | |
| 3 | LV | Luvisol | |
| 4 | LX | Lixisol | |
| 5 | HS | Histosol | 2) Suelos con capa orgánica delgada |
| 6 | CL | Calcisol | 3) Suelos con acumulación de sales moderadamente solubles o sustancias no salinas |
| 7 | DU | Durisol | |
| 8 | AR | Arenosol | 4) Suelos con poca o nula diferenciación de perfiles |
| 9 | CM | Cambisol | |
| 10 | FL | Fluvisol | |
| 11 | RG | Regosol | |
| 12 | CH | Chernozem | |
| 13 | KS | Kastañozem | 5) Suelos con una pronunciada acumulación de materia orgánica en el suelo mineral |
| 14 | PH | Phaeozem | |
| 15 | UM | Umbrisol | |
| 16 | AN | Andosol | 6) Suelos que se distinguen por su química Fe/Al |
| 17 | GL | Gleysol | |
| 18 | NT | Nitisol | |
| 19 | PL | Planosol | |
| 20 | PT | Plintosol | |
| 21 | LP | Leptosol | 7) Suelos con limitaciones para el crecimiento de las plantas |
| 22 | SC | Solonchak | |
| 23 | SN | Solonetz | |
| 24 | VR | Vertisol | |

3. Selección de las variables informativas

Para la construcción de modelos se requiere elegir cuidadosamente las variables que van a ser utilizadas, es por ello que un paso crucial en el análisis de grandes bases de datos es la *reducción de las variables* (Kathait, 2017). En este sentido las técnicas estadísticas se posicionan como herramientas para gestionar eficazmente la selección de variables, lo que conduce a un modelo parsimonioso con la mayor precisión de predicción posible

De esta manera, las variables utilizadas para el modelo final, fueron seleccionadas de la base de datos total con diferentes análisis estadísticos. A estas variables las llamamos **variables informativas**; y con informativas nos referimos a que son variables que cumplen con los siguientes criterios: 1) que estén menos correlacionadas con cualquiera de las otras variables, y 2) que contribuyan a explicar la variabilidad general del sistema cafetalero.

A continuación, se enlistan y describen los análisis de exploración de datos que fueron utilizados en el procedimiento de reducción de las variables. El procesamiento se realizó en el software R (Versión 3.4.4).

❖ *Análisis de correlación*

Para medir la fuerza o el grado de asociación entre cada combinación de dos variables numéricas de la base de datos, se calculó la correlación a través de la técnica no paramétrica: *Coficiente de Spearman o rho de Spearman* (Rodríguez-Salazar *et al.*, 2000). Se eligió este coeficiente debido a que se asume en este caso que algunas variables no presentan una distribución normal.

La *rho* de Spearman varía entre -1 y 1, los límites indican una correlación máxima y 0 la ausencia de una relación monótona (Rodríguez-Salazar *et al.*, 2000).

En este análisis se identificaron aquellos pares de variables que presentaban una correlación “fuerte” (≥ 0.7) y significativa ($p < 0.05$) (Ortega *et al.*, 2009), para posteriormente eliminar aquellas variables que estaban significativamente más correlacionadas con otras variables de la base de datos, y con esto poder realizar la primera reducción de variables.

❖ *Análisis de multicolinealidad*

Se realizó para conocer si dentro de las variables numéricas existen algunas que presenten *multicolinealidad* (relaciones estadísticas de correlación aproximadamente lineales) con respecto a las otras variables numéricas de la base de datos. La prueba utilizada para la detección de la

multicolinealidad fue la del ***Factor de inflación de varianza (VIF)***. El VIF nos indica el grado en el que la varianza del estimador de mínimos cuadrados se eleva (o infla) debido a que la variable está correlacionada con al menos una de las otras variables de la base de datos. En la práctica, partir de 5 se considera que existe multicolinealidad (Esteban, 2008).

Después de identificar las variables que presentaban mayor multicolinealidad en la base de datos, se llevó a cabo su eliminación y con ello, el segundo proceso de reducción de variables.

❖ ***Análisis Factorial Mixto***

Dado que el Análisis Factorial Mixto es un análisis multivariado que permite el tratamiento simultáneo de variables de distinta naturaleza, tanto cualitativa como cuantitativa (Zubcoff, 2017), éste fue utilizado para conocer cuáles variables describen mejor a todas las AGEB rurales. Para ello, se evaluó la contribución de todas las variables en conjunto (numéricas y categóricas), y también se evaluaron por separado.

Finalmente, con los resultados del FAMD y del modelo de regresión lineal múltiple, se realizó una segunda reducción de variables.

4. Clasificación de las AGEB rurales

❖ **Árbol de regresión**

En el presente trabajo se busca clasificar a todas las AGEBs rurales en dónde se cultiva café, a través de la predicción de la variable objetivo o de respuesta: *productividad*. Por lo tanto, se utilizó un árbol de regresión como modelo de clasificación y predicción (Willians, 2011). Como producto de este análisis, se muestra de manera representativa en una figura de árbol, aquellas variables que logran describir las diferencias en la productividad de todas las AGEBs.

La variable *productividad* fue calculada de la siguiente manera:

$$\text{Productividad}_{(n)} = \text{vol. cosechado}_{(n)} / \text{sup. producida}_{(n)}$$

n= AGEB rural

La representación gráfica de un árbol consta de nodos, hojas y ramas. El nodo principal o raíz es la variable a partir del cual se inicia el proceso de clasificación. Los nodos internos corresponden a cada una de las preguntas acerca de la variable en particular. Cada posible respuesta a los cuestionamientos se

representa mediante un nodo hijo. Las ramas que salen de cada uno de estos nodos se encuentran etiquetadas con los posibles valores de la variable. Los nodos finales o nodos hoja corresponden a una decisión, la cual coincide con uno de los niveles de la variable objetivo (Willians, 2011).

5. Indicador de Productividad

Con la finalidad de poder medir y comparar la productividad explicada por las categorías creadas por el árbol de regresión, se clasificó a la productividad en siete categorías lingüísticas, creando así un *indicador de productividad*.

El indicador de productividad se llevó a cabo utilizando la *Ley de Weber-Fechner* (Bojórquez-Tapia et al., 2009), así como considerando el número de categorías lingüísticas propuestas por *Miller*, que son 7 ± 2 (Miller, 1956). En la sección 1.d de los ANEXOS se encuentra más información sobre la Ley Weber-Fechner y el número de Miller.

Finalmente se eligió que fueran 7 las categorías de productividad:

1. Extremadamente alta
2. Muy alta
3. Alta
4. Intermedia
5. Baja
6. Muy baja
7. Extremadamente baja

La aplicación de la Ley de Weber-Fechner consistió en:

1) calcular el estímulo cero:

$$e_0 = \frac{\text{ValorMax} - \text{ValorMin}}{f^v}$$

Dónde:

e_0 = estímulo cero

v = número de clases o categorías totales

f = Factor de progresión

y 2) obtener cada uno de los cortes de las categorías:

$$e_n = (f)^n \cdot e_0$$

Dónde:

n = número de la categoría que se calcula

En la Figura 5. se muestra un diagrama de flujo para sintetizar la metodología. Mientras que en la *sección 1.c* de los ANEXOS, se definen y describen de manera más detallada los análisis estadísticos utilizados.

Así también en el repositorio https://github.com/LANCIS-escalante-lab/TESIS_cafe_Karla se encuentra el script utilizado para los análisis estadísticos realizados, en un archivo nombrado “3_ANALISIS_ESTADISTICOS”.



Figura. 5. Diagrama de flujo del Método

VIII. RESULTADOS

1. Obtención de información relacionada al cultivo de café en México

A partir del procesamiento de datos del Censo Agropecuario 2007 en las instalaciones del INEGI, así como del procesamiento de capas vectoriales con información ambiental, se obtuvieron un total de **167 variables**, las cuáles fueron agrupadas en seis grupos: 1) características sociodemográficas del productor; 2) manejo de cultivo; 3) procesamiento, venta y consumo; 4) organización del productor; 5) ambientales. En las *Tablas A1 a A5* de la sección 2.a de ANEXOS se muestran todas las variables desglosadas por categoría, así como las abreviaturas que fueron utilizadas para cada una de ellas durante el análisis en el software R y que serán utilizadas también a lo largo del presente texto para fines prácticos. A continuación, en la Tabla 5 se muestran tres variables por grupo a modo de ejemplo.

Tabla 5. Ejemplos de las variables pertenecientes a los seis grupos.

| Grupo | Nombre de la variable | Abreviatura |
|--|--|---------------|
| Grupo 1 <i>Características sociodemográficas del productor</i> | Obtuvo algún crédito/préstamo | crédito |
| | Contaba con seguro | seguro |
| | Apoyo por parte de los programas de gobierno | apoyo_gob |
| Grupo 2 <i>Manejo del cultivo</i> | Uso de fertilizantes químicos | fert |
| | Realiza rotación de cultivos | rotacion_cult |
| | Superficie de cultivo orgánico | sup_org |
| Grupo 3 <i>Procesamiento, transporte, venta y consumo</i> | Contaba con beneficiadora de café o cacao | b_cafe |
| | Venta a una empaedora o agroindustria | v_agroi |
| | Toda la producción a consumo familiar | con_fam_td |
| Grupo 4 <i>Relacionadas a la organización del productor</i> | Organización con otros productores para obtener algún apoyo o servicio | org_p_apoy |
| | Organización individual para realizar actividades agropecuarias/forestales | org_indiv |
| | Participación de familiares del productor para realizar labores agropecuarias/forestales | part_fam |
| Grupo 5 | Agricultura es la principal actividad | p_act_agr |
| | No sembró por falta de dinero o apoyos | falta_dinero |

| | | |
|---|--|--------------|
| <i>Relacionadas con la actividad y problemática principal</i> | Pérdida de fertilidad del suelo principal problema | pro_p_fertil |
| Grupo 6 <i>Características ambientales</i> | Temperatura media anual | temperatura |
| | Tipo de clima | clima |
| | Tipo de suelo | suelo |

2. Reducción de las variables a través de análisis estadísticos

❖ Análisis de Correlación

Se calculó la correlación por cada par de variables numéricas con el *coeficiente de Spearman* (ρ), creándose una matriz de correlación (datos no mostrados). Así también, se realizó un gráfico de correlación para identificar visualmente las combinaciones de variables con las asociaciones más fuertes en la matriz de correlación; en color azul oscuro se muestran las correlaciones positivas y en color rojo oscuro las correlaciones negativas (Figura 6). Como resultado se obtuvieron 64 variables correlacionadas fuertemente ($\rho \geq 0.7$) y de manera significativa ($p < 0.05$) en forma de 35 pares de variables (Tablas 6 y 7). De estos 35 pares, 9 se agruparon en tríos (Tabla 7), debido a que eran variables muy correlacionadas entre sí y que pertenecían al mismo grupo de variables.

Posteriormente se procedió a elegir una variable por cada grupo de variables fuertemente correlacionadas, para así evitar sobrestimar las relaciones de las variables en la base de datos. De manera que se seleccionó una por cada par de variables de la Tabla 6, y una variable por cada grupo de tres pares de la Tabla 7; a excepción del último trío, en donde no se seleccionó ninguna variable debido a que dos de ellas fueron utilizadas para crear la variable productividad (vol_cosechado y sup_plantada) y la restante está muy correlacionada con estas dos.

Finalmente se eliminaron 35 variables como resultado de este primer análisis, teniendo así una base de datos de 132 variables totales (no incluyendo la variable productividad).

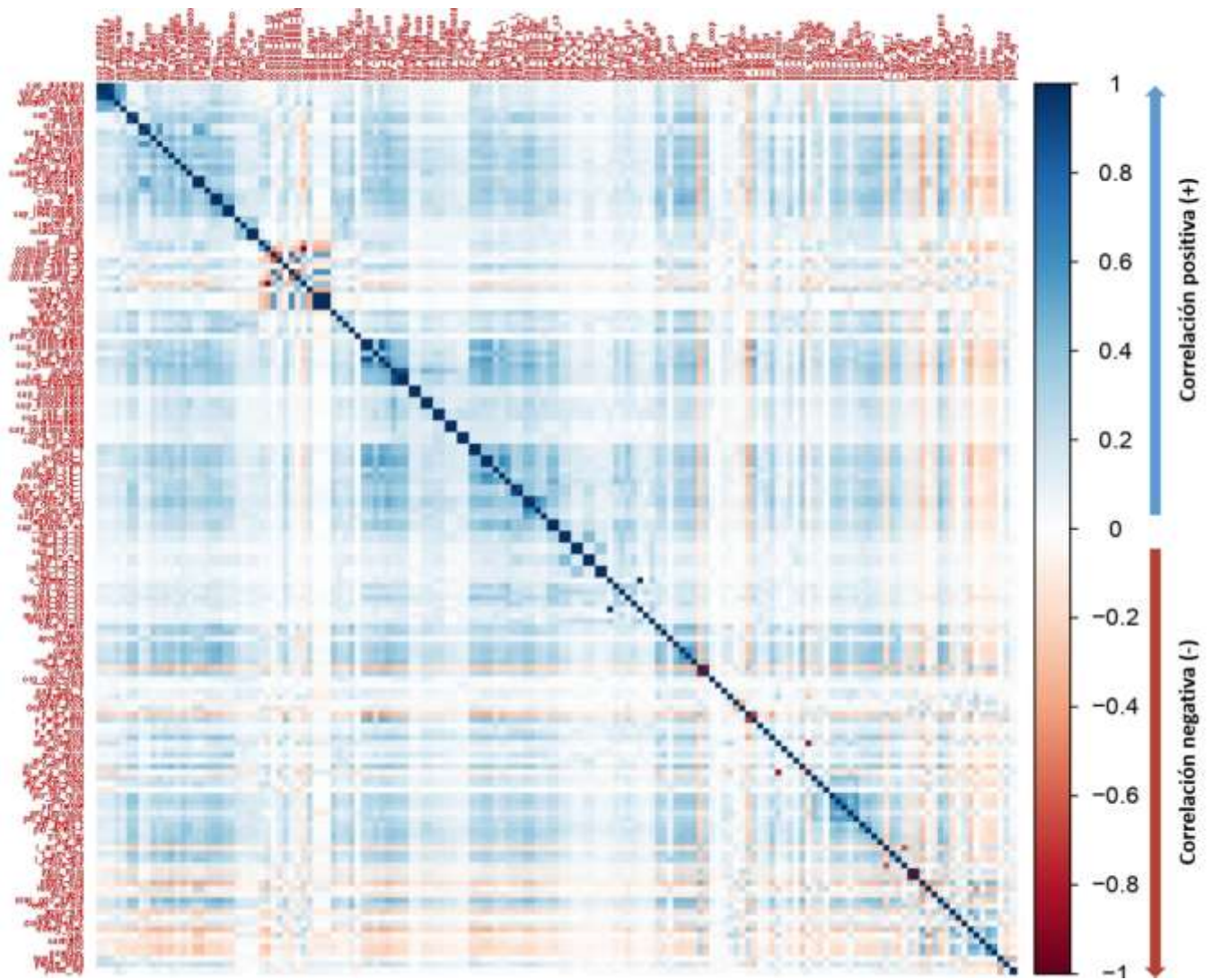


Figura 6. Gráfico de la matriz de correlación de Spearman para todas las variables numéricas registradas. En azul se muestran las correlaciones positivas y en azul las correlaciones negativas, el tono corresponde a el grado de correlación señalado por el indicador que se muestra a la derecha de la imagen.

Tabla 6. Pares de variables fuerte ($p \geq 0.7$) y significativamente ($p < 0.05$) correlacionadas. Las variables seleccionadas se muestran con un asterisco (*).

| Variables correlacionadas (pares) | | Variables correlacionadas (pares) | |
|-----------------------------------|----------------|-----------------------------------|-----------------|
| sup_intercal | Intercal * | sup_bosque | bosque * |
| sup_no_semb | sin_semb * | arb_terr_prod_1 | plant_spp_reg * |
| sup_descanso | en_descanso * | sup_desm_agr | desmonte_sb * |
| sup_vivero | vivero * | sup_aclareo_sb | aclareo_b_1 |
| sup_invernadero | invernadero * | sup_h_q_sb | herb_q_sb * |
| rotacion_cult | podas * | sup_h_o_sb | herb_o_sb * |
| consum_anim_pt | con_fam_pt * | sup_i_q_sb | insec_q_b * |
| sup_ensalitrada | erosionada * | sup_i_o_sb | insec_org_sb * |
| sup_erosionada | erosionada * | s_otratec_sb | otra_tec_sb * |
| sup_cub_agua | cub_agua * | consum_fam_td | venta * |
| sup_contaminada | contaminada * | org_indiv | org_fam * |
| sup_o_sn_veg | otra_sin_veg * | p_act_agr | p_act_anim * |
| sup_selva | selva * | uso_camion | uso_otro_med * |

Tabla 7. Pares de variables fuerte ($p \geq 0.7$) y significativamente ($p < 0.05$) correlacionadas ordenados por grupos de tres. Las variables seleccionadas se muestran con un asterisco (*). Y de las variables en ***cursivas*** no se seleccionó ninguna debido a que fueron utilizadas para crear variable productividad.

| Variables correlacionadas (por grupos de 3 pares) | |
|---|----------------------|
| venta_may | venta_comer |
| venta_may | venta_agroi * |
| venta_comer | venta_agroi |
| sup_enmontada | sup_pastos |
| sup_enmontada | enmontada * |
| sup_pastos | enmontada |
| sin_veg | sup_arenales |
| sin_veg | arenal_pedregal * |
| sup_arenales | arenal_pedregal |
| <i>sup_plantada</i> | <i>sup_producida</i> |
| <i>sup_plantada</i> | <i>vol_cosechado</i> |
| <i>sup_producida</i> | <i>vol_cosechado</i> |

❖ Análisis de Multicolinealidad

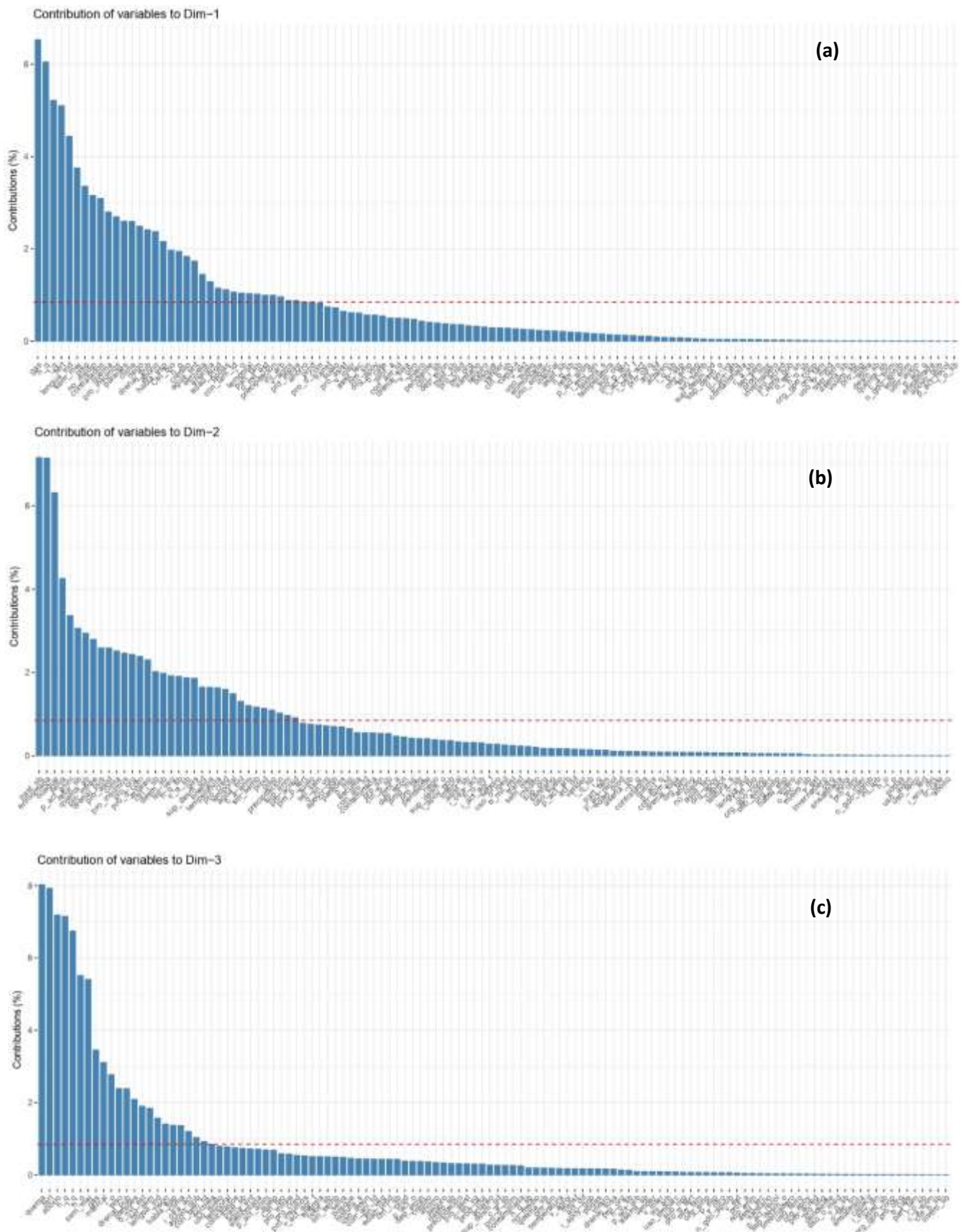
Se realizó la detección de multicolinealidad de las variables numéricas de la base de datos y se encontraron 17 variables que presentaban multicolinealidad (Tabla 8).

Tabla 8. Variables en las que se detectó multicolinealidad. Se muestran las variables seleccionadas con un asterisco (*).

| | Variables Grupo 2 | Variables Grupo 3 | Variables de Grupo 5 |
|----|-------------------|-------------------|----------------------|
| 1 | sin_veg * | con_fam_nd | sin_semb * |
| 2 | arenal_pedregal | con_fam_td * | no_intereso |
| 3 | erosionada | con_fam_pt | falta_dinero |
| 4 | cub_agua | | mal_temporal |
| 5 | | | no_hubo_quien |
| 6 | | | estaba_invadida |
| 7 | | | suelo_p_fertil |
| 8 | | | suelo_erosionado |
| 9 | | | en_descanso |
| 10 | | | o_causa_ns |

Como podemos observar en la Tabla 8, el análisis detectó variables de tres grupos diferentes (Grupo 2, Grupo 3 y Grupo 5). Con la finalidad de eliminar la multicolinealidad en el sistema, se eligió una única variable de cada grupo. Para ello se revisó el Cuestionario del Censo Agropecuario 2007 y se encontró que en el Grupo 2 la variable “sin_veg” (tiene superficie sin vegetación) englobaba las otras variables de su grupo, ya que correspondían a subpreguntas de la pregunta no.41. La misma situación se encontró para la variable “sin_semb” que englobaba a las demás variables de su grupo (pregunta no.16). Mientras que, en el Grupo 3 las tres variables correspondían a las diferentes opciones de la pregunta no.33, por lo que eran mutuamente excluyentes, por lo tanto, con la información de una sola variable era suficiente, en este caso se eligió a la variable “consum_fam_td” (toda la producción a consumo familiar).

Con los resultados generados por este análisis se realizó la segunda reducción de variables, eliminándose en total 14 variables. Finalmente, a partir de este análisis se obtuvo una base de datos con 118 variables totales.



(a)

(b)

(c)

Figura 8. Gráficos del porcentaje de contribución de las variables para cada una de las tres primeras dimensiones. (a) Porcentaje de contribución de la dimensión 1. (b) Porcentaje de contribución de la dimensión 2. (c) Porcentaje de contribución de la dimensión 3. La línea roja punteada corresponde a la media esperada del porcentaje de contribución de las variables si las contribuciones fueran uniformes.

A partir de los gráficos de contribución se seleccionaron aquellas variables cuyo porcentaje de contribución era mayor que la media esperada para cada una de las diez dimensiones principales. Obteniendo como resultado 90 variables totales seleccionadas (incluyendo la variable *productividad*). En la Tabla 9, se muestran las variables seleccionadas. De manera que, con los resultados de este análisis multivariado se llevó a cabo la última reducción de las variables de las bases de datos.

Tabla 9. Variables con mayor contribución en las diez dimensiones principales.

| | Variables | Dim 1 | Dim 2 | Dim 3 | Dim 4 | Dim 5 | Dim 6 | Dim 7 | Dim 8 | Dim 9 | Dim 10 |
|----|---------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| 1 | Gas | x | x | x | | | | | | | |
| 2 | h_q | x | | x | | | | | | | |
| 3 | i_q | x | | x | | | | | | | |
| 4 | Fert | x | | x | | | | | | | |
| 5 | Lengua_ind | x | | x | | | | x | | | x |
| 6 | Sem_mej | x | | x | | | | | | | |
| 7 | Piso | x | x | x | | | | x | x | | |
| 8 | No_medio | x | | | | | | x | | | |
| 9 | Contratados | x | | | | | | | | | |
| 10 | Quema | x | | x | | | | | | | |
| 11 | Pro_insumos | x | x | | x | | x | | | | |
| 12 | Clima | x | x | x | x | x | x | x | x | x | x |
| 13 | Paredes | x | | x | | | | x | x | x | |
| 14 | Riego | x | | x | | | | x | | | |
| 15 | Drenaj_pub | x | x | x | | | | | x | x | x |
| 16 | Abono | x | | x | | | | | | | |
| 17 | Habla_esp | x | | x | | x | | x | | x | |
| 18 | Ctrl_bio | x | | | | | | | | | |
| 19 | H_o | x | | x | | | | | | | |
| 20 | Crédito | x | | | | x | | x | | | |
| 21 | Agua_ent | x | x | | | | | x | x | | |
| 22 | Escuela | x | | x | | | | x | | | |
| 23 | Energía | x | x | | | | | x | x | | |
| 24 | Edad_prod | x | | | | | | x | x | x | |
| 25 | Podas | x | | | | | | | | | x |
| 26 | Temporal | x | x | | | | | x | | | x |
| 27 | l_o | x | | x | | | | | | | |
| 28 | Precipitación | x | x | | x | x | x | x | | x | |
| 29 | O_nvl_est | x | | x | | | | | | x | |
| 30 | Part_fam | x | | | | | | | | | |
| 31 | V_agroi | x | | | | | x | | | | x |
| 32 | Inj_arb | x | | | | x | x | | | | x |
| 33 | con_fam_td | x | | x | | x | | | | x | x |
| 34 | Prod | x | x | | | | | x | | | |
| 35 | Ahorro | x | | | | | | | | | |
| 36 | Pro_infrae | x | x | | x | | | | x | | |
| 37 | Past_sb | | x | | x | | | | x | x | x |
| 38 | enmontada | | x | | x | | | | | | |
| 39 | Bosque | | x | | x | | | | | | x |
| 40 | Selva | | x | | x | | | | | | x |

Continuación Tabla 9. Variables con mayor contribución en las diez dimensiones principales.

| | Variables | Dim 1 | Dim 2 | Dim 3 | Dim 4 | Dim 5 | Dim 6 | Dim 7 | Dim 8 | Dim 9 | Dim 10 |
|----|--------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| 41 | Refor_sb | | x | | x | | | | | | x |
| 42 | Quema_sb | | x | | | | | | | | x |
| 43 | Pro_asist | | x | | x | | x | | x | | |
| 44 | Pro_d_com | | x | | x | | x | | x | x | |
| 45 | Pro_org | | x | | x | | | | x | | |
| 46 | Aclareo | | x | | x | | | | | | |
| 47 | Ctrl_i_sb | | x | | x | | | x | | | |
| 48 | Desm_sb | | x | | | | | | x | x | x |
| 49 | H_q_sb | | x | | x | 0 | | | x | x | |
| 50 | C_a_sb | | x | | | | | | | | x |
| 51 | Sin_veg | | x | | | | | | x | | |
| 52 | Pro_cred | | x | | x | | | | x | | |
| 53 | Sup_desm_gd | | x | | | | | | x | x | x |
| 54 | Temperatura | | x | | | x | x | | | | |
| 55 | C_a_enm | | x | | | | | | | | |
| 56 | Sup_enm | | x | | | | | | | | |
| 57 | Pro_acred | | x | | x | | | | | | |
| 58 | Sin_semb | | x | | | | | | | x | x |
| 59 | Intercal | | | x | | | | | | x | |
| 60 | Apoy_gob | | | x | | | | x | | x | |
| 61 | Pro_p_fert | | | | x | | x | x | | | |
| 62 | Pro_sin | | | | x | x | | x | | | |
| 63 | Uso_embarc | | | | x | x | x | | x | x | |
| 64 | Ctrl_b_sb | | | | x | x | x | x | x | | x |
| 65 | V_ext | | | | x | x | x | | | | |
| 66 | Asist_sb | | | | x | x | x | x | x | | x |
| 67 | B_cafe | | | | x | x | x | | x | | |
| 68 | I_act_agr_f | | | | x | x | | | | x | x |
| 69 | Suelo | | | | x | x | x | x | x | x | x |
| 70 | Vivero | | | | | x | x | | | | |
| 71 | V_inter | | | | x | | | | | | |
| 72 | Uso_suelo | | | | | x | | | | x | x |
| 73 | Sel_sem | | | x | | x | x | | | x | x |
| 74 | P_act_arb | | | | x | x | | x | x | x | |
| 75 | Pendiente | | | | | x | | x | x | | x |
| 76 | I_otra_act | | | x | | x | | | | x | |
| 77 | Textura_s | | | | | x | x | | | x | x |
| 78 | Drenaje_fosa | | | | | x | | | | | |
| 79 | Sup_org | | | | | | x | | | | |
| 80 | P_ssp | | | | | | | x | x | | x |
| 81 | Sanitario | | | | | | | x | | | x |
| 82 | Prof_a | | | | | | | x | | | |
| 83 | Dep_econ | | | | | | | x | | x | |
| 84 | O_sn_vg | | | | | | | | x | x | x |
| 85 | contaminada | | | | | | | | x | x | x |
| 86 | capacit | | | | | | | | x | | |
| 87 | Asist_tec | | | | | | | | x | | |
| 88 | I_apoy_gob | | | | | | | | | x | |
| 89 | V_otro | | | | | | | | | | x |
| 90 | P_act_o | | | | | | | | | | x |

3. Árbol de regresión como modelo de clasificación de las AGEB rurales donde se cultiva café a partir de su productividad

Se realizó el modelo de árbol de regresión considerando como variable de respuesta la *productividad* y como variables explicativas las 89 variables restantes de la base de datos.

Como resultado el árbol de regresión arrojó **9 variables principales** que explican las diferencias de productividad en las AGEBs rurales en dónde se cultiva café. También se generaron 10 categorías (nodos finales del árbol), y cada categoría es explicada por los valores específicos de cada una de las variables que la constituyen (nodos internos). En la Figura 9 se muestra el árbol de regresión, en sus nodos finales se señala la productividad y el porcentaje de AGEBs rurales que son representadas por cada una de las categorías formadas. Así también, con la finalidad de visualizar de una manera más sencilla la magnitud de las diferencias en cuanto a productividad, se muestran todos los nodos coloreados de acuerdo al indicador de productividad creado, en una gama de colores de claros a oscuros, de menor a mayor productividad respectivamente.

De manera que, cada categoría constituye un escenario específico que ocurre en esas AGEBs rurales, y que está representado por la combinación de variables y sus respectivos valores para cada rama del árbol.

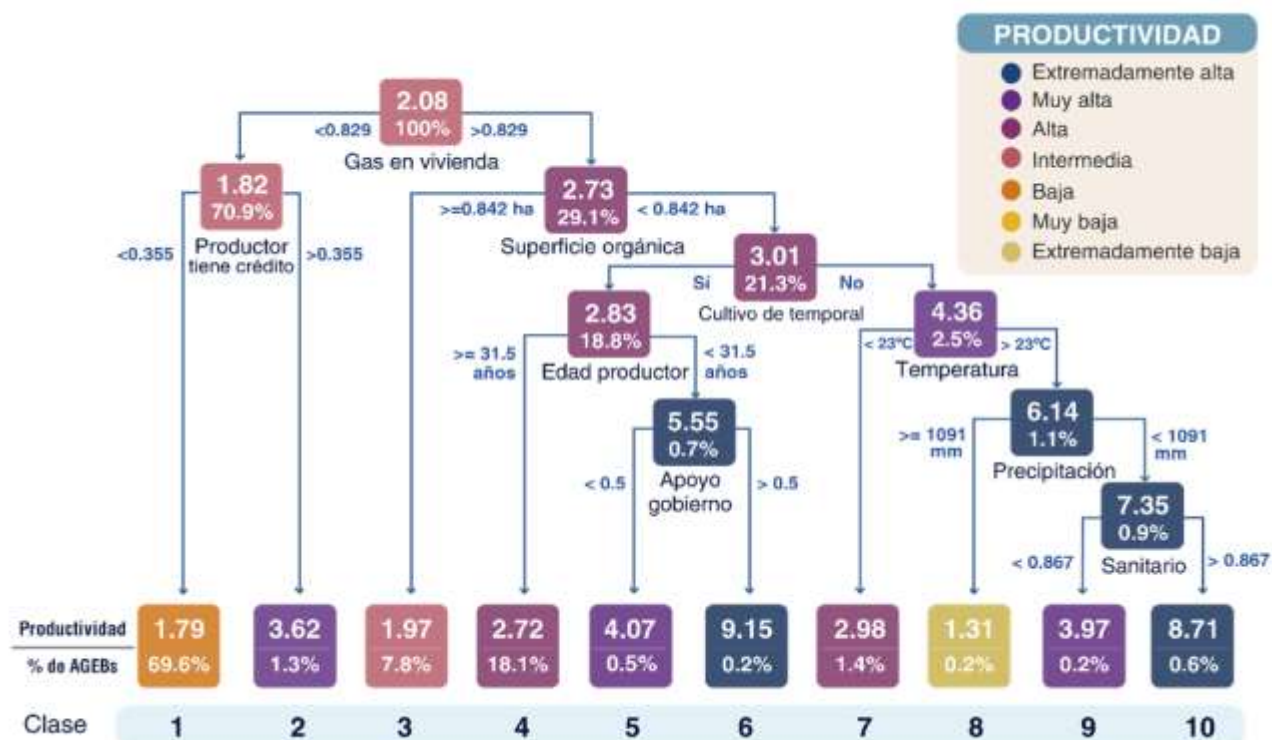


Figura 9. **Árbol de regresión** que muestra las 9 variables que explican la productividad de las AGEB rurales en dónde se siembra café. En cada nodo se señala el valor promedio de la **productividad** que se explica por la condición de esa variable (parte superior) y el **porcentaje de AGEBs** que representa (parte inferior). Los nodos terminales representan cada una de las 10 categorías formadas por el árbol. En colores se muestran coloreados todos los nodos dependiendo su nivel de productividad en el índice.

Finalmente, para evaluar la eficiencia del modelo de árbol de regresión se realizó un gráfico de residuos estandarizados (Figura A6 de ANEXOS). La mayoría de los residuos se ajustaron al valor cero, lo que significa que el modelo tiene un buen ajuste a lo que se observa en el sistema cafetal.

4. Mapas

Finalmente, se realizaron dos mapas para ubicar a todas las AGEB rurales en dónde se siembra café y que pertenecen a cada una de las clases creadas con el árbol de regresión. En la Figura 10 se muestra en a) el mapa con las AGEBs rurales pertenecientes a categoría 1, y en b) las AGEBs rurales pertenecientes a las nuevas categorías restantes.

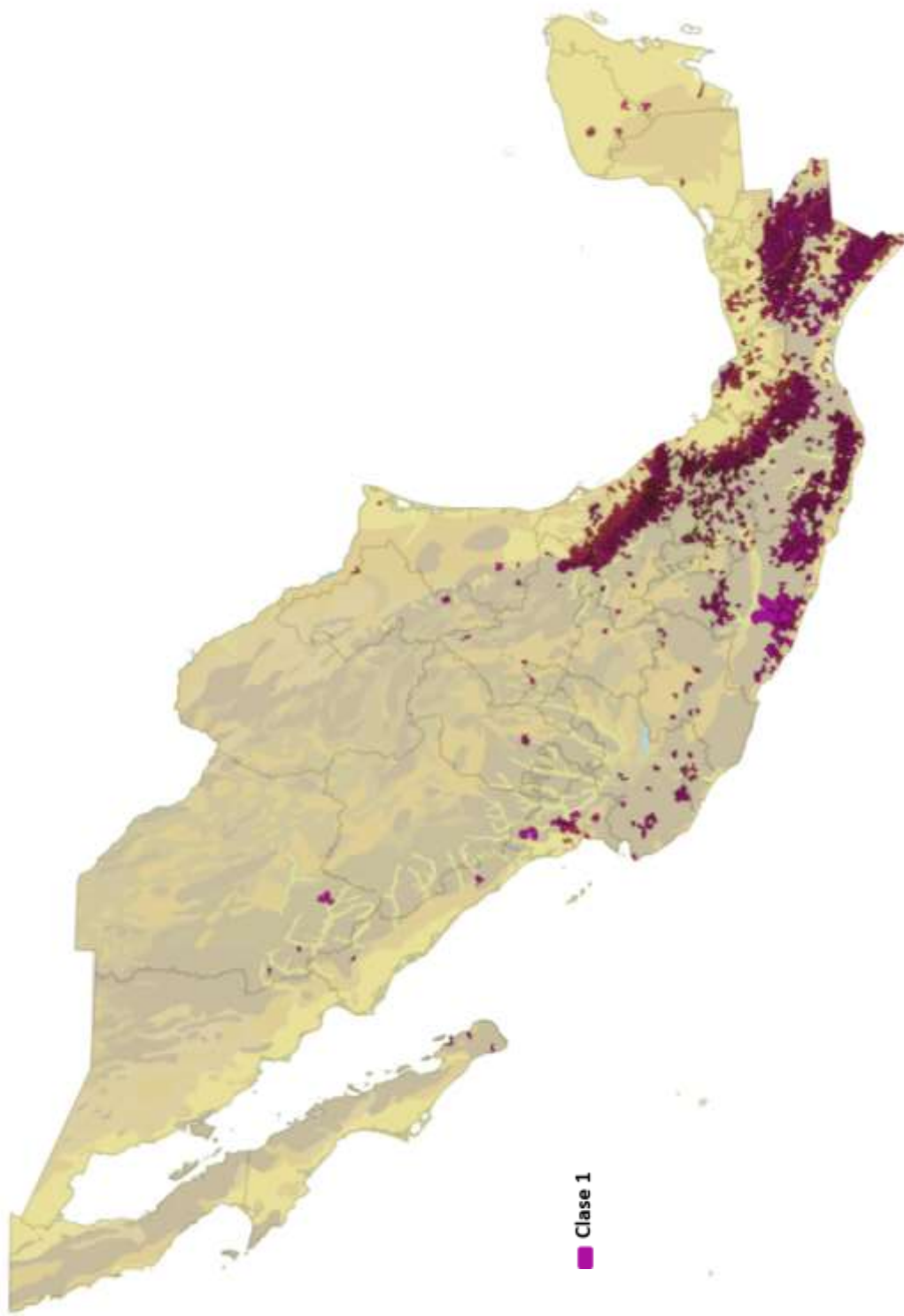


Figura 10. Mapa de la República mexicana en dónde se señalan los AGEB rurales que conforman la **Clase 1** generada por el árbol de regresión. El mapa presenta como fondo la capa "El relieve como atractivo natural" descargada del catálogo de metadatos geográficos de CONABIO.

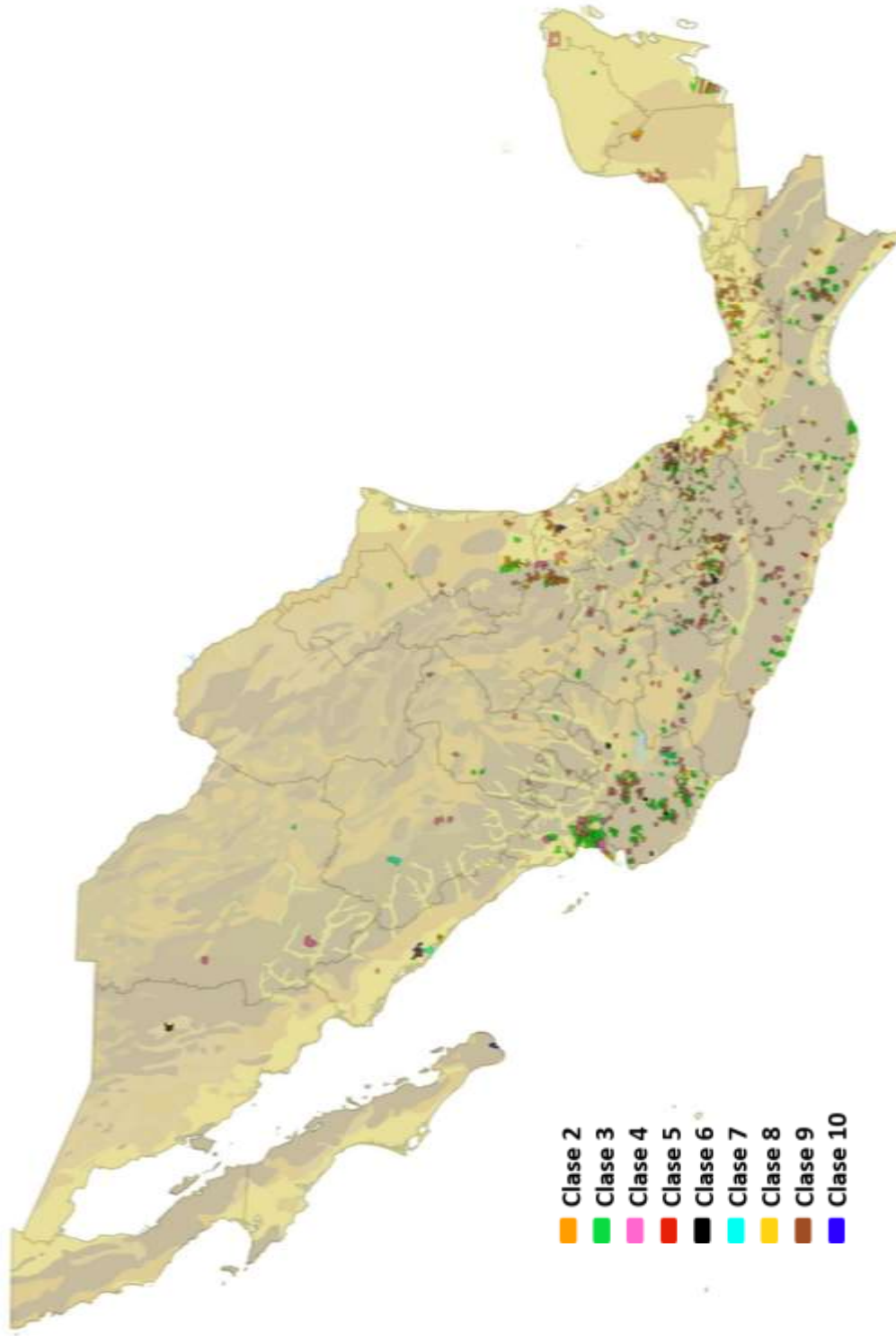


Figura 11. Mapa de la República mexicana en dónde se señalan las AGEB rurales que conforman la **Clase 2-10** generadas por el árbol de regresión. El mapa presenta como fondo la capa "El relieve como atractivo natural" descargada del catálogo de metadatos geográficos de CONABIO.

IX. DISCUSIÓN

El diagnóstico del estado del cultivo de café con base en la categorización de las variables asociadas con las distintas AGEBs rurales en el territorio nacional, arroja como determinante estadística de la productividad, la situación socio-económica de los productores, directamente relacionada con el nivel de desarrollo de la región y la infraestructura disponible.

Este diagnóstico se basa en los resultados del árbol de regresión, o el modelo estadístico que resulta del análisis de los datos del censo agropecuario 2007 (Figura 9), y que a continuación se discutirá puntualmente en cuanto a las variables que mejor explican las categorías de productividad.

❖ Infraestructura doméstica (gas)

La variable “gas” (que el productor cuente con gas para cocinar en su vivienda) resultó ser la variable que explica mejor la diferencias en productividad de todo el sistema cafetalero mexicano para el año 2007. En esta primera bifurcación del árbol, un 29.1% de las AGEBs tiene más del 0.829 de frecuencia correspondiente a gas en la vivienda, con una productividad= 2.73 ton/ha (alta), esto significa, que menos del 30% de todo el sistema cafetalero tiene de forma certera (frecuencia cercana a 1) gas en la vivienda para cocinar. Mientras que la gran mayoría (70.9%) tiene una frecuencia menor a este valor y con una productividad= 1.82 ton/ha (intermedia). Esto nos sugiere que un aumento en la frecuencia de poseer gas en la vivienda está explicando el cambio de una productividad *intermedia* (1.80-2.29 ton/ha) a una productividad *alta* (2.29-3.27 ton/ha). Este resultado es el hallazgo más relevante de este trabajo, ya que una variable socioeconómica es la que divide en dos partes a todas las AGEBs en dónde se siembra café dada su productividad, y nos muestra que la mayoría de los productores cafetaleros no tienen asegurado gas para cocinar en su vivienda. Este dato concuerda con la realidad que se vive en el sector cafetalero mexicano, en dónde el 90% de la superficie cosechada se ubica en zonas de alta y muy alta marginación. Según cifras del Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL), se estima que cerca del 50% de la población en los diez principales municipios productores de café en el país vive en condición de pobreza (SIAP-SAGARPA, 2016). Y como se señaló previamente en los *Antecedentes* de este trabajo, el 80% de la superficie del cultivo de café, se concentra en unidades productivas de menos de 10 ha (Moguel y Toledo, 2004). El 65% de la superficie cafetalera del país la manejan pequeños productores con menos 2 ha y son ellos los que generan la mitad de la producción total. De manera que, el café en México se produce a una pequeña escala, en dónde los pequeños productores generalmente pertenecen a comunidades indígenas o mestizas (Moguel & Toledo, 1999); el

66% de los productores cafetaleros habla al menos una lengua indígena (Palomares, González & Mireles, 2012).

Si la mayoría de los cafetaleros mexicanos se encuentran en condiciones de pobreza, sus necesidades básicas no están cubiertas, y los subsidios a la producción por parte del gobierno sólo se transformarán en subsidios al consumo, ya que los productores lo gastarán en alimentos, en vez de gastarlos en fertilizantes, semillas mejorada y otros insumos agrícolas (Martínez, 2006). Así mismo, hay que recordar que muchos de los programas de apoyo al campo (como lo era *Procampo*), son pagos que se asignan por hectáreas (Kuntz, 2010), y si las superficies de cultivo de la mayoría de los cafetaleros son menores de 10 ha, el apoyo económico que reciben es lógicamente menor que todos aquellos grandes productores comerciales que poseen áreas de cultivo mucho mayores, con sistemas de producción tecnificados y cuya producción se exporta a otros países. Y este escenario es el que se vive en toda la agricultura mexicana, en dónde la mayoría de los agricultores son pequeños agricultores que comúnmente operan en el umbral de la subsistencia y condiciones de pobreza, combatiendo otros obstáculos como la exclusión del mercado, de manera que se encuentran marginalizados social y económicamente (Quintero, 2017). Mientras que una pequeña fracción de productores son agricultores empresariales que cuentan con mayor tecnología y recursos económicos. En México más del 40% de los agricultores mexicanos viven en la pobreza extrema y de los 50 millones de pobres que hay, 30 millones viven en zonas rurales (Sánchez Cano, 2014). El 57% de las personas que viven en el sector rural no tiene acceso a los servicios básicos de vivienda, cómo en este caso el gas para cocinar. Y el 63% de los habitantes de las comunidades rurales no cuenta con el ingreso suficiente para comprar una canasta básica de alimentos (DOF, 2014).

De manera que, se requiere de una política integral de reformas en el sector rural, y coincidimos con Martínez (2007) en las siguientes propuestas: incrementar medidas económicas orientadas a abaratar todos los costos de producción (insumos, transporte, energía y comercialización); la investigación dirigida a la tecnología agrícola; inversión en servicios (camino, escuelas, educación, salud y alimentación); y que exista un crecimiento paralelo en los otros sectores económicos para que así se aproveche la mano de obra que sobra en el campo y de esa misma manera se reduzca el desempleo rural, y también para fortalecer el mercado interno de los productos de origen agropecuario.

❖ Créditos y préstamos

En aquellas AGEBs con una frecuencia de gas en la vivienda < 0.829 , la variable socioeconómica “**crédito**” (si él o la productora obtuvo algún crédito o préstamo) explica si la productividad aumenta dos niveles en la escala del indicador de productividad. Específicamente en aquellas AGEBs rurales con una frecuencia de crédito < 0.355 , que se entiende que sus productores no tuvieron acceso a dicho crédito, presentan una productividad *Baja* de 1.79 ton/ha (*Clase 1*) y que corresponde el 69.6% de todo el sistema cafetalero. Mientras que, aquellas AGEBs en donde sus productores obtuvieron un crédito o préstamo con frecuencia > 0.355 , poseen en promedio una productividad *Muy alta* (3.62 ton/ ha) y representan únicamente al 1.3% del sistema (*Clase 2*). Este resultado, nuevamente ejemplifica otro de los retos económicos que enfrentan la mayoría de los cafetaleros mexicanos que corresponden a pequeños productores. Con este dato se manifiesta que todos aquellos productores que no tienen asegurado el gas para cocinar en su vivienda y tampoco obtuvieron un crédito o préstamo, obtuvieron como promedio el segundo valor más bajo de productividad representando en el árbol de regresión.

En este contexto, se sabe a partir de los resultados del censo Agropecuario 2007, que solo el 4.2% de las unidades de producción (UP) nacionales recibieron crédito o seguro. Y en los municipios cafetaleros es todavía menor la cobertura, sólo un 2.5% de las (UP) tienen acceso a estos instrumentos financieros. Por lo que el acceso a financiamiento para las actividades agropecuarias en México resulta limitado (Robles, 2011).

En un estudio se encontró que sólo 37% de la población económicamente activa operaba en el circuito de los servicios de la banca comercial, 15% contaba con acceso limitado a la banca social, 13% era atendido por la banca de desarrollo y el 35% por ciento restante estaba marginado del mercado (Ruíz 2002). Lo que demuestra que el acceso a los créditos y préstamo no es equitativo para todos los productores y dado el resultado obtenido en el árbol de regresión, pertenecer al porcentaje sin acceso o con acceso limitado representa una desventaja para aumentar su productividad agrícola.

Por otro lado, en el 2004 el Programa de Apoyo para Acceder al Sistema Financiero Rural (PAASFIR) apoyó a 26 de 32 organizaciones que presentaron su solicitud a la SAGARPA y a la Financiera Rural; 17 eran reconocidas como Intermediarios Financieros Rurales (IFR), 2 fueron organismos integradores y las otras 7 dispensoras de crédito. De las 26 asociaciones favorecidas, la mitad eran de ejidatarios y la otra mitad de propietarios privados, el 51% de ellos producía en áreas de riego, 33% en temporal y 16% en zonas de pastizal; las personas beneficiadas fueron alrededor de 85 mil y la mayoría de ellas se concentraron en el norte del país (FND. 2005). En este sentido, existe una disparidad regional en la

ayuda financiera en México. De acuerdo con cifras de SAGARPA-FAO, en el Noroeste del país el 28% de las unidades tuvieron un crédito, mientras que en el Sureste solamente el 2.9% lograron obtener financiamiento. La razón de esta diferencia se explica por la alta correlación que existe entre el nivel de ingresos y el acceso a crédito; en la región Noroeste el ingreso promedio por Unidad Económica Rural (UER) es de \$214,831 pesos, mientras que en el Sureste es de \$27,694 pesos (DOF, 2014). Aunada a esta diferencia de apoyo financiero regional, se sabe que los caficultores empresariales, que representan únicamente el 0.4% de los productores de café, poseen un adecuado acceso a la estructura de subsidios a la producción, al crédito institucional y comercial y son los dueños de la mayor parte de la infraestructura industrial (Hernández-Trujillo, 2016). Estos datos respaldan el resultado generado por el modelo, la gran mayoría de los productores de café no cuentan con el acceso a un crédito o préstamo, y aquellos pocos que si lo obtuvieron se vieron beneficiados con una productividad en sus terrenos de cultivo.

De manera que, es necesario aumentar la cobertura de los créditos y préstamos, para que se encuentren disponibles y de manera equitativa en toda la República Mexicana. Pero principalmente, hay que buscar aumentar la calidad de vida de los cafetaleros, porque difícilmente un pequeño productor que vive en condiciones de extrema pobreza, va a lograr conseguir los requisitos que le solicitan para poder acceder a un crédito o préstamo. Nuevamente con esta variable socioeconómica (crédito y préstamo) se resalta la urgencia de crear reformas rurales que reestructuren y mejoren la vida en el campo mexicano.

❖ **Variables de manejo de cultivo**

La variable que mejor divide al 29.1% de las AGEBS cafetaleras cuyos productores poseen gas en su vivienda (>0.829), resulta ser la “**superficie orgánica**” (cantidad de superficie de cultivo orgánico). De manera que el 7.8% del total de AGEBS rurales, presentan una superficie con cultivo orgánico mayor o igual a 0.842 ha, con un promedio de productividad *intermedia* de 1.97 ton/ ha (**Clase 3**). No obstante, aquellas que poseen menos de 0.842 ha de cultivo orgánico presentan una mayor productividad: 3.01 ton/ha (*productividad alta*) y corresponden al 21.3% del sistema cafetalero. Lo que coincide con lo que se ha reportado en agroecosistemas en México y Costa Rica con cultivos de café orgánico bajo sombra, cuya productividad fue un 20-30% menor en comparación con el café convencional sombreado y un 40% menor que el café sin sombra (Van der Vossen, 2005). Los autores también mencionan que los ingresos netos fueron de entre un 25-50% más bajos que en las fincas manejadas convencionalmente. Uno de los factores que se ha considerado como limitante de la productividad en los cultivos orgánicos es el nitrógeno (N). Los sistemas de agricultura orgánica tienen el potencial de suministrar grandes cantidades de N a los cultivos en crecimiento mediante la incorporación de residuos de cultivos, abonos y composta,

sin embargo, los niveles óptimos de N disponible rara vez se logran en la práctica y existe una pobre sincronización de la disponibilidad de N y la demanda de cultivos (Berry et al.,2002).

Ahora bien, lo que este resultado nos indica son los retos de productividad que experimentan todos aquellos productores orgánicos de café. De modo que es importante que los valores agregados para sus productos de venta sean los justos, de acuerdo a los beneficios ambientales que proveen este tipo de cultivos, como lo es la protección del suelo y del agua de la contaminación química, que se produce en la agricultura convencional a causa de los agroquímicos utilizados.

Los cafés que promueven la sostenibilidad ambiental, social y económica a largo plazo, y que están certificados por un tercero acreditado, se conocen como *cafés sostenibles*. Los cafés orgánicos, cultivados a la sombra (o respetuosos con el medio ambiente) y de comercio justo⁸ son todos cafés sostenibles, que generan primas⁹ de precios (Giovannucci y Koekoek 2003, ITC 2011).

El café orgánico mexicano se ha diferenciado y ha incursionado en mercados muy especializados (Palomares, González & Mireles, 2012). Actualmente, México se coloca como segundo productor mundial de café orgánico, con un volumen de producción de 350 mil sacos de 60 kilos. Los estados de Chiapas, Oaxaca, Veracruz y Puebla son los principales estados productores de café orgánico. Y es importante señalar que México se coloca como el principal proveedor mundial, exportando cada año más de 38 mil toneladas a países europeos (CEDRSSA, 2018).

Cabe señalar que el valor del café orgánico cuando es certificado internacionalmente ha logrado mejorar el nivel de vida de muchas familias campesinas en México. El sobreprecio de este producto (que puede ser de hasta 30% con respecto al no orgánico) se debe al valor extra que se añade a todos los conocimientos de la caficultura tradicional, a evitar el uso de insumos sintéticos procedentes del exterior e intensificar el uso de mano de obra, bajo normas y reglamentos establecidos por agencias certificadoras. De esta manera, el café orgánico se integra a procesos organizativos democráticos y autogestionados de organizaciones de pequeños productores, aspectos sociales que además han impulsado la creación del comercio justo. Los principales productores orgánicos son en su mayoría grupos muy bien organizados, muchos de ellos en comunidades indígenas. Es por ello, que el café orgánico y el comercio justo, se

⁸ **Comercio justo:** Forma alternativa de comercialización en el mundo desarrollado de productos y servicios procedentes de países en vías de desarrollo o subdesarrollados a un precio justo y en situación de dignidad entre productor y consumidor con el fin de mejorar las condiciones de estos países y luchar contra la pobreza (En línea en: <https://economipedia.com/definiciones/comercio-justo.html>).

⁹ **Prima:** Recompensa económica que se le otorga a alguien a modo de premio o incentivo (En línea en: <https://definicion.de/prima/>).

piensa que son grandes alternativas para que México tenga un crecimiento económico y por fin la pobreza de estas zonas marginadas disminuya (Najera, 2002; Palomares, González & Mireles, 2012).

No obstante, es importante tener en cuenta que para que un productor ingrese al mercado del café orgánico, es necesario que cuente con una certificación que lo acredite, y para aplicar a ella, debe cumplir con estrictas normas de producción, bajo el control de inspectores de una Agencia Certificadora Internacional. Y el proceso de certificación y las inspecciones tienen un costo, por lo que es recomendable que los caficultores se asocien para compartir los costos (Palomares, González & Mireles, 2012).

En un estudio en dónde se comparan regímenes de comercio de café en Chiapas, Luna & Wilson (2015) encontraron que los pequeños productores pertenecientes a dos regímenes basados en cooperativas (Comercio justo e Integración vertical)¹⁰ presentan varias ventajas en comparación con los productores convencionales no afiliados a ninguna cooperativa, como lo es que aprecian precios más altos por el café que venden a la cooperativa y que tienen acceso servicios y programas gubernamentales. Y adicionalmente, en el caso específico de la cooperativa de comercio justo presentan mayor disponibilidad de crédito. No obstante, una de las desventajas que mencionaron los productores de café de esta cooperativa es que los ingresos se los pagan en cuotas durante el año de cosecha (sistema de pago a plazos). Mientras que los productores independientes (no afiliados a ninguna cooperativa) tienen la ventaja de un pago completo e inmediato en efectivo y la disponibilidad de crédito del coyote o intermediario. Sin embargo, las principales desventajas que se encontraron para estos productores independientes son los precios bajos y variables a los que se enfrentan, y que no poseen un comprobante de venta que les permita participar en programas como el del *Fomento Productivo* de AMECAFE (Asociación Mexicana de la Cadena de Producción de Café) o del *Fondo de Compensación de Precios de Café*, que aumenta los ingresos de los agricultores a través de reembolsos a los compradores de café (Luna & Wilson, 2015).

Lo anterior cobra relevancia al tomar en cuenta que las políticas y programas que promueven el desarrollo económico de la agricultura a pequeña escala enfrentan la realidad competitiva de los mercados locales, regionales y globales (Luna & Wilson, 2015).

¹⁰ **Integración vertical:** Estrategia de crecimiento de algunas empresas en dónde éstas entran en actividades relacionadas con el ciclo de producción de un producto o servicio. Así, desde una fase inicial en el desarrollo y producción de un producto hasta el abastecimiento de los clientes hay diferentes procesos que pueden ser aprovechados por la misma firma. (En línea en: <https://economipedia.com/definiciones/integracion-vertical.html>).

De manera que, el apoyo económico por parte de programas gubernamentales para lograr obtener las certificaciones resulta ser de gran ayuda para todos aquellos pequeños productores que cultivan café orgánico y que se asocian entre ellos buscando ser recompensados por sus prácticas de manejo que favorecen el cuidado del medio ambiente y de esta manera que contribuyen a la sostenibilidad del sistema socioecológico.

Por otro lado, la variable “**temporal**” fue la que permitió clasificar en el modelo a aquellas AGEBs rurales que poseen gas para cocinar (>0.829) y que tienen sembrados menos de 0.842 ha de cultivo orgánico. Esta variable se refiere a si existe superficie de temporal en los terrenos de los productores, si la respuesta es *Sí*, entonces se explica el 18.8% de las AGEBs cafetaleras que presentan un promedio de productividad *Alta* de 2.83 ton/ha. En cambio, si la respuesta es *No*, las AGEBs que se explican por esta serie de condicionantes resultan ser únicamente el 2.5% del sistema total, con un valor promedio de productividad de 4.36 ton/ha (*Muy alta*). Estos resultados tienen sentido, si se deduce que ese 2.5% de las AGEBs cafetaleras que no tienen superficie de temporal, corresponden por lo tanto a superficies con sistemas de riego de agua. Por lo que, es muy probable que se traten de sistemas cafetaleros tecnificados con un sistema de producción a pleno sol. Lo que conllevaría a un mayor uso de insumos como fertilizantes y otros agroquímicos como plaguicidas y herbicidas, que pueden estar influyendo positivamente en la productividad de los cafetos. De hecho, existen estudios que han demostrado que, bajo un manejo intensivo del cultivo, el café a menudo produce rendimientos mucho más altos que un sistema bajo sombra (DaMatta y Rodríguez, 2007). E incluso esta fue la principal razón por la cual la producción bajo sombra se ha abandonado casi por completo en Brasil desde la década de 1950, en donde se ha explicado la baja productividad como una consecuencia probable de la competencia excesiva por el agua entre las especies de árboles y el café, así como por el sombreado excesivo (DaMatta y Rena, 2002). También, es importante considerar que, en superficies tecnificadas e intensificadas a pleno sol, se siembra el café en forma de monocultivo en densidades que doblan o incluso quintuplican las plantas sembradas por hectárea en comparación con los cultivos bajo sombra (Manion et al., 1999), por lo que también es de esperarse que la productividad sea mayor en este tipo de sistemas de producción, ya que hay un mayor número de plantas por superficie de área.

Finalmente, el porcentaje que el modelo de árbol de regresión proporciona con esta variable (2.5%) coincide con el porcentaje reportado de café producido con sistemas de riego en México, correspondiente al 3% del total (Robles, 2011). Y en este sentido, cabe destacar que estas condiciones específicas en donde la productividad presenta de los valores más altos que se pueden alcanzar en el cultivo de café, están disponibles para una fracción muy pequeña del total de productores cafetaleros mexicanos. No

obstante, la productividad del 18.8% de las AGEBs de cultivo de temporal es alta (2.83 ton/ha) y comparada con la productividad baja (1.79 ton/ha) de las AGEBs en donde tienen gas en la vivienda <0.829 y crédito <0.355 , se encuentra dos niveles de productividad por encima, lo que deja en evidencia que el cultivo de temporal (que describe la mayoría de los cultivos de nuestro país) puede llegar a ser una buena alternativa, si el productor no tiene una situación económica desfavorable. Y con esto, vale la pena recordar todos los beneficios ambientales que el café cultivado bajo sombra conlleva; como la conservación de la biodiversidad y los servicios ecosistémicos que proveen a la sociedad (López & Caamal, 2009). De manera que, más que incitar a hacer un cambio en el sistema de producción (de sombra a pleno sol) disminuyendo así la sostenibilidad del sistema, lo que se debe hacer es identificar los problemas a los que los pequeños productores se enfrentan, para poder aumentar su productividad, y encontrar las soluciones a estos.

❖ Edad del productor

La “**edad del productor**” fue la variable que permitió dividir aquellas AGEBs rurales que sí tienen superficie de temporal, esta variable resultó ser una de las variables más interesantes que salieron del árbol de regresión, ya que sugiere que una característica sociodemográfica como lo es la edad es importante para explicar las diferencias de productividad del sistema de café en México. El punto de corte de esta variable son los 31 y medio años; si el o la productora tiene 31.5 años o más, entonces se explica un 18.1% de las AGEBs cafetaleras totales, con una productividad *Alta* de 2.72 ton/ha (**Clase 4**). En cambio, si el productor o la productora tiene menos de 31.5 años se explica únicamente el 0.7% del sistema, pero la productividad resulta tener un valor *extremadamente alto*: 5.55 ton/ha. Este dato, sugiere que en ese pequeño porcentaje (menor al 1%) de AGEBs rurales en donde se siembra café, *los y las productores más jóvenes están manejando de una manera tal sus cultivos que su productividad resulta ser de las más altas en todo el sistema*. La disminución de la productividad en adultos mayores sí se ha mencionado en la literatura y se ha explicado por la disminución de la fuerza física del productor después de la mediana edad, aunado a que la planeación a futuro se reduce, e incluso que las decisiones y acciones para manejar el cultivo se vuelven más conservadoras, por lo que es menos probable que los agricultores de mayor edad adopten las nuevas tecnologías que se necesitan para incrementar la productividad agrícola de forma sostenible (Tauer, 1984; Woodsong, 1994; FAO, 2014). Asimismo, se ha reportado que la edad del jefe de hogar es un factor significativo que influye en la decisión de un agricultor de adoptar estrategias de diversificación de cultivos. Cuanto mayor es el agricultor, menos probabilidades tiene de recoger el agua, diversificar los cultivos o plantar árboles con un nivel de significación del 1% (Kinuthia *et al.*, 2018). De manera que, el manejo de los sistemas agrícolas sí depende de la edad de los

productores. Y este dato toma relevancia, considerando que el promedio de edad de los productores de café que contestaron el cuestionario del Censo Agropecuario 2007, es de 51.6 ± 12 años, por lo que la mayoría de los productores de café en México son adultos mayores. No obstante, también es importante señalar que muchas veces la experiencia de los agricultores mayores conduce a combinaciones de insumos más eficientes, lo que hace que una unidad de trabajo sea más efectiva; recompensando los aspectos negativos que su envejecimiento genera (Taur, 1984).

❖ Apoyos de gobierno

La variable “**Apoyo del gobierno**” (si el productor obtuvo algún apoyo por parte de los Programas del Gobierno) fue aquella que permitió clasificar a las AGEBs cafetaleras que son manejadas por productores menores de 31.5 años. El 0.2 % de estas AGEBs recibió apoyo de parte de un programa gubernamental con una frecuencia menor a 0.5 y una productividad promedio de 4.97 (*Muy alta*), mientras que el 0.5% de las AGEBs restante recibió apoyo por parte del gobierno con una frecuencia mayor a 0.5 y tiene una productividad promedio de 9.15 (*Extremadamente alta*), que es el valor promedio más alto de productividad explicado por las categorías generadas por el modelo de árbol de regresión (en este caso, correspondiente a la *Clase 6*). Este resultado muestra que el hecho de poder obtener un apoyo económico por parte de los Programas de Gobierno resulta un factor importante que tiene repercusiones positivas en la productividad del cultivo de café, por lo menos en lo que respecta a los productores jóvenes (<31.5 años) que tienen cultivos de temporal, una superficie orgánica menor a 0.842 ha y que poseen gas en su vivienda para cocinar con una frecuencia mayor a 0.829.

En este sentido, resulta pertinente mencionar los desafíos que presentan los jóvenes productores y que fueron identificados durante el proyecto “Facilitar el Acceso de los Jóvenes Rurales a las Actividades Agrícolas” por parte del conjunto MIJARC/FAO/FIDA en 2011: **1)** el acceso insuficiente de los jóvenes al conocimiento, a la información y a la educación; **2)** el acceso limitado de los jóvenes a la tierra; **3)** el acceso inadecuado a los servicios financieros; **4)** la dificultad para acceder a empleos verdes; **5)** el acceso limitado de los jóvenes a los mercados; y **6)** la participación limitada de los jóvenes en el diálogo sobre políticas (FAO, 2014). Conociendo los obstáculos por los que se enfrentan la mayoría de los agricultores jóvenes, resulta más fácil comprender por qué los productores jóvenes de la clase 6 del árbol de regresión, se ven beneficiados cuando reciben un apoyo extra (en cuestión monetaria o de insumos) por parte de un

programa gubernamental, al grado de representar la productividad más alta del sistema de AGEBs rurales cafetaleras mexicanas. No obstante, no hay que olvidar el hecho que esta situación particular explica únicamente el 0.2% de todo el sistema. Por lo que es muy probable que la mayoría de los cafetaleros jóvenes de México se encuentren representados en el 70.9% de las AGEBs rurales que tienen una frecuencia menor del 0.829 de gas para cocinar, aumentando así el nivel de obstáculos económicos que viven. Y con esto, vale la pena enfatizar la problemática de la juventud rural que trae consigo desempleo y pobreza; situación que se agrava si se trata de mujeres rurales jóvenes, ya que culturalmente se espera que ellas ayuden con las tareas domésticas y, por lo tanto, poseen menos oportunidades para asistir a actividades de aprendizaje, participar en las actividades de las organizaciones rurales y acceder a los mercados, sumado a que tradicionalmente las mujeres solo pueden acceder a la tierra únicamente por medio de su relación con un familiar masculino (FAO, 2014).

Actualmente se necesita lograr el involucramiento de la juventud en el sector agrícola, ya que con el crecimiento de la población mundial los jóvenes van a jugar un papel fundamental para garantizar que exista un futuro con seguridad alimentaria para ellos mismos y para las futuras generaciones (FAO, 2014). Es por ello, que es importante que tanto los programas como las políticas públicas mexicanas deben de tratar de resolver las necesidades de esta población juvenil de pequeños productores que siembran tanto café como cualquier otro tipo de cultivo. Porque si no se considera así, los jóvenes rurales continuarán abandonando el campo y migrando a zonas urbanas en busca de mejores oportunidades económicas, poniendo en peligro de la seguridad alimentaria de todos en un futuro no muy lejano.

❖ Variables ambientales

Por otro lado, para aquellas AGEBs que no tienen terrenos con superficie de temporal, la variable que mejor las clasifica es “**temperatura**” (temperatura media anual); si la temperatura es menor a 23 °C se explica el 1.4% de las AGEBs cafetaleras con un promedio de productividad de 2.98 ton/ha (*alta*) (**Clase 7**), mientras que, si la temperatura resulta mayor a los 23°C, entonces se explica el 1.1% de las AGEBs cafetaleras que tuvieron un promedio de productividad de 6.14 ton/ha (*extremadamente alta*). De manera que, la temperatura resulta ser la primera variable ambiental que aparece en el árbol de regresión, lo que nos demuestra que las condiciones climáticas sí tienen una repercusión en la productividad del cultivo de café. En México el café se cultiva en regiones con temperaturas de 13 a 26 °C (Sagarpa, 2017). Se sabe que la temperatura óptima (media anual) para la especie *Coffea arabica* (café arábico) es de 15 a 24°C y de 24 a 30°C para la especie *Coffea canephora* (café robusta) (Clifford, 1985). El cultivo de café robusta corresponde tan solo al 5.5% de la producción de café en México. Se sabe que esta especie es

más resistente a las plagas, que contiene 2.5 por ciento de cafeína (el doble que *C. arabica*) y que su sabor más es áspero (razón por la cual se dice que tiene una calidad menor). La variedad robusta se siembra de manera tecnificada a pleno sol y se utiliza principalmente para la industria de café soluble. Una vez planteadas estas características, si deducimos que se trata de cultivos en dónde se utiliza tecnología de irrigación para el riego de los cafetos, entonces estaremos hablando de que el 1.1% de las AGEBs que presentan en promedio una productividad extremadamente alta (6.14 ton/ha) tienen cultivos de la especie robusta, y por lo tanto una temperatura mayor a 23°C va a favorecer la producción de esta especie de café, de acuerdo a sus características biológicas.

Si nos seguimos en esta rama del árbol, con el 1.1% de AGEBs cuya temperatura es mayor a 23 °C, nos damos cuenta que la variable siguiente que divide a este conjunto de observaciones, es nuevamente una variable ambiental: la “**precipitación**” (precipitación media anual). Si la precipitación es mayor o igual a 1091 mm por año, entonces se explica el 0.2% de las AGEBs cafetaleras totales con un promedio de productividad de 1.31 ton/ha (*extremadamente bajo*), y que corresponde al valor más bajo de productividad representado en la clasificación presentada por el modelo (**Clase 8**). Por otra parte, si la precipitación es menor a 1091 mm por año, entonces se explica el 0.9% del sistema cafetalero, con una productividad *extremadamente alta* de 7.35 ton/ha. Este resultado nos muestra los dos niveles extremos en el indicador de productividad, no obstante, resulta difícil de interpretar si únicamente nos centramos en las condiciones de precipitación óptimas para las dos especies principales de café que se siembran en México, que para *C.arabica* es de 1500-2000 mm y para *C.canephora* de 2000-2300 mm (Clifford, 1985), ya que no parece tener una explicación biológica, sobre todo considerando que los cultivos de deben tratarse de variedades de café robusta, cuya precipitación óptima está en un rango de lluvias abundantes. Más bien, lo que esta variable muestra es que los productores de las fincas cafetaleras que tienen un sistema de producción intensivo y tecnificado con instalaciones de riego, han recurrido a estas tecnologías debido a las que las condiciones climáticas en dónde se desarrollaba el cultivo lo ameritaban. Revisando específicamente a las AGEBs que corresponden a esta ruta de variables (el 0.9% del total) nos percatamos que la mayoría de ellas se encuentran en el norte del país, en estados como Baja California Sur, Sinaloa, Nayarit y Tamaulipas, así como en el norte de Veracruz, Zacatecas, Michoacán, Morelos. En dónde, en Nayarit y B.C.S presentan un rango de precipitación media anual de 400-500 mm/año. De manera que, podemos constatar que la variable precipitación lo que nos describe son las condiciones climáticas inadecuadas para el cultivo de café, que representan un desafío para los productores que viven en estas zonas del país. Y solo aquellos que poseen la capacidad económica para contrarrestar los retos climáticos, recurriendo a tecnologías de irrigación, son los únicos que pueden

cosechar los beneficios de una productividad extremadamente alta (y que representa el 0.9% del total de AGEBs cafetaleras). No obstante, no es posible explicar de igual manera el 0.2% restante, en dónde llueve más de 1091 mm anuales, y en dónde la productividad obtenida es extremadamente baja. Con ayuda del mapa, se pueden ubicar estas AGEBs rurales en Oaxaca, Yucatán y Nayarit, en dónde la precipitación es de 1500-2000 para Oaxaca y de 1000-1200 para los otros dos estados, característica que podría llegar a ser insuficiente para la especie robusta, solo si el cultivo fuera de temporal, pero si en este caso el cultivo es de riego, se desconocen las causas que podrían de esta problemática, y valdría la pena analizarlo más a profundidad. Sin embargo, también es importante considerar que el porcentaje de AGEBs rurales explicado por el modelo en este punto es muy bajo (0.2%) hay posibilidad de que estemos presenciando también un sobreajuste del modelo.

❖ Alcances y limitaciones del modelado estadístico de sistemas socio-ambientales

Los modelos son abstracciones o estructuras intelectuales que creamos para estudiar algún fenómeno de la realidad. En este sentido cada modelo no es la realidad o es una “mentira” ya que es una representación imperfecta de la realidad, pero se espera y se busca que el modelo se parezca a la realidad en los aspectos más importantes y que solo difiera en que es más manejable. Para confirmar que esto efectivamente sucede, se analiza si los resultados son propiedades de la realidad o del modelo (Scheffer & Carpenter, 2003; Levins, 2004; Lewontin & Levins, 2007).

Es deseable trabajar con modelos manejables que maximicen la generalidad, el realismo y la precisión hacia los objetivos superpuestos, pero no idénticos en comprender, predecir y modificar los sistemas; no obstante, no es posible satisfacer todos estos criterios a la vez. Y es por ello que tenemos que elegir qué criterios enfatizar de acuerdo con el problema y luego cambiar a otro modelo (Lewontin & Levins, 2007).

En el caso particular de este proyecto, es importante considerar primeramente que el modelo de árbol de regresión utilizó como insumos a variables que describen el cultivo de café en México en una sola temporalidad ya que la mayoría de la información se refiere al año agrícola 2006 – 2007, que fue cuando se realizó el Censo Agrícola, Ganadero y Forestal 2007. Así mismo, hay que tener en cuenta que la escala es grande ya que los datos proporcionados son a nivel nacional, porque nuestro objetivo era ver patrones o tendencias generales con respecto a la productividad del cultivo de café en México para ese momento. Pero estamos conscientes en que se pueden estar perdiendo detalles y particularidades específicas del sistema cafetalero a nivel regional y local. Cuando un modelo no genera predicciones puntuales para los parámetros de salida, pierde *precisión*. Así también, un modelo pequeño, puede ser más *preciso* ya que hay menos variables para medir, pero los procesos dominantes pueden aparecer como entradas externas

al sistema. En cambio, cuando ampliamos el modelo, aunque podemos estar perdiendo precisión, se gana *realismo*, porque notamos que los procesos opuestos ya no son externos, sino que surgen dentro del sistema más inclusivo (Lewontin & Levins, 2007). Si un modelo se aplica a más sistemas del mundo real, es *general* y si tiene en cuenta más variables independientes que se sabe que tienen un efecto, es *realista* (Orzack & Sober, 1993). De manera que, el modelo de árbol de regresión aquí presentado puede considerarse que aborda los criterios de *generalidad* y *realismo*. *General* en el sentido de que podría utilizarse para otro tipo de cultivos y *realista* porque se utilizan varias variables independientes que abordan distintas características que influyen en el manejo del cultivo de café (ej. variables socioeconómicas, variables ambientales).

De forma particular el modelo de regresión presenta ventajas, pero también desventajas en su utilización. A continuación, se comparten las que se mencionan en la literatura (Orellana, 2018):

➤ **Ventajas**

- Fácil de entender.
- Útil en exploración de datos: al identificar importancia de ciertas variables a en una base de datos con muchas variables.
- Menos limpieza de datos: los outliers y valores faltantes no influyen el modelo.
- El tipo de datos no es una restricción: se utilizan variables categóricas y numéricas.
- Es un método no paramétrico (i.e., no hay suposición acerca del espacio de distribución y la estructura del clasificado).

➤ **Desventajas**

- Sobreajuste.
- Pérdida de información al categorizar variables continuas.
- Precisión: métodos como SVM (*Support Vector Machines*), y clasificadores tipo ensamblador a menudo tienen tasas de error 30% más bajas que CART (*Classification and Regression Trees*)
- Inestabilidad: un pequeño cambio en los datos puede modificar ampliamente la estructura del árbol. Por ejemplo, la eliminación de varias observaciones, podría conducir a cambios radicales en el árbol de decisión: aumento o disminución de la complejidad del árbol, cambios en la división de las variables y los valores.

Y otra desventaja que también se menciona es que todas las divisiones realizadas por el árbol son perpendiculares a los ejes (X y Y), por lo tanto, si los datos tienen una estructura muy compleja, podría

ser que CART no capte la estructura correcta. Para capturar la estructura de los datos, el algoritmo de división generará muchas divisiones (nodos) en el borde. No obstante, al final, CART creará un árbol grande en donde casi cada observación en el borde estará en un nodo separado. Pero a pesar del gran árbol, la clasificación se realizará correctamente y con todas las observaciones que pertenezcan (Timofeev, 2014).

Así mismo, es importante señalar que las relaciones entre las variables presentadas en el modelo de árbol de regresión son relaciones estadísticas y no causales. Por lo que queda abierta la línea de investigación para explorar de qué tipo son las asociaciones que aquí reportamos, con la ayuda de experimentación diseñada explícitamente para estos fines.

Finalmente, el conocer las desventajas del árbol de regresión, así como las limitaciones de los modelos en general, nos permite comprender el por qué no debemos buscar explicar la totalidad un sistema complejo como lo son los sistemas agroalimentarios únicamente a través de los resultados de uno sólo modelo. Sino que hay que utilizarlos como herramientas de la investigación creadas para ampliar el entendimiento de estos sistemas socio ambientales.

X. CONCLUSIONES

En este trabajo utilizamos como modelo estadístico el árbol de regresión, con la finalidad de identificar aquellas variables que permiten la categorización de las AGEBs rurales en dónde se lleva a cabo el cultivo de café en México para el ciclo de cosecha 2006-2007, con base a la productividad obtenida. Y a continuación se resumen los resultados más relevantes:

- ❖ Las variables *“gas para cocinar en la vivienda”* y *“obtención de crédito o préstamo”*; que contienen información sobre las características socioeconómicas del productor; son las variables principales que dividen las diferencias de productividad en el sistema cafetalero en México. Y son las que explican el mayor porcentaje de las AGEBs rurales del país en dónde se cultiva café, a diferencia de otro tipo de variables como las relacionadas al tipo de manejo o a las características ambientales en dónde se desarrolla el cultivo.
- ❖ Las variables relacionadas al manejo de cultivo *“cantidad de superficie de cultivo orgánico”* y *“cultivo de temporal”* son las que mejor explican el sistema cafetalero después de las variables socioeconómicas antes señaladas. No obstante, su relación es negativa con la productividad, a mayores hectáreas de cultivo orgánico, y en presencia de un cultivo de temporal, la productividad

es menor, que en comparación con los cultivos convencionales (no orgánicos) y con sistema de riego. Sin embargo, la sostenibilidad del sistema socioambiental de los cultivos de café orgánico y/o de todos aquellos que se encuentran bajo un sistema de producción tradicional de sombra, es mucho mayor que sus contrapartes que resultan tener una mayor productividad, por lo que hay preferir el escenario que involucra bienestar social, ambiental y económico por encima del puramente económico (productividad del cultivo).

- ❖ La *edad del productor* tiene una influencia en la productividad de las AGEBs cafetaleras. Los productores jóvenes y que recibieron “*apoyo de programas gubernamentales*” explicaron la productividad más alta observada en el sistema cafetalero mexicano analizado. Resultado que invita a priorizar los apoyos gubernamentales hacia el grupo vulnerable al que pertenecen los pequeños productores jóvenes.
- ❖ Las variables ambientales “temperatura media anual” y “precipitación media anual” explicaron un porcentaje pequeño de la productividad de AGEBs rurales que presentan cultivos con instalaciones para riego, y que muy probablemente se utilicen para la industria de café soluble, muchas de las cuales se localizan en la zona norte del país.

Si bien este estudio puede acotarse a la información analizada y a la temporalidad de los datos, los resultados presentados nos permiten tener un acercamiento respecto a los principales factores que influyen en la productividad de la caficultura mexicana; que es una de las actividades económicas más importantes del sector agrícola y que conlleva una alta prestación de servicios ecosistémicos debido al sistema de producción que predomina en nuestro país (cultivo bajo sombra).

Con este análisis en dónde se integra información social, ambiental y económica, buscamos aportar conocimiento que pueda ayudar a la identificación de patrones que reflejan los datos del cultivo de café, y que podrían ser de utilidad para los tomadores de decisiones en el gobierno, al momento de establecer prioridades en el diseño de políticas públicas que busquen aumentar la productividad del sector cafetalero en México.

Finalmente, sugerimos la utilización de este tipo de modelos estadísticos para analizar otro tipo de cultivos relevantes en nuestro país con información diversa (socioeconómica, características del manejo, organización de productores, características ambientales, etc.). Con el objetivo de contribuir de esta forma al desarrollo de diagnósticos integrales, en dónde los resultados obtenidos permitan retroalimentar y fortalecer esos mismos u otros diagnósticos integrales (a diferentes escalas), buscando lograr que la

agricultura transite hacia trayectorias cada vez más sostenibles, en dónde no solo se potencie su valor económico, sino también el cultural y el ambiental.

XI. PERSPECTIVAS

A partir de lo elaborado en la presente tesis de maestría, se plantean las siguientes perspectivas para trabajos futuros relacionados:

- Analizar, procesar y crear un modelo de árbol de regresión con los datos generados por el futuro **Censo Agrícola, Ganadero y Forestal de INEGI**, para comparar los resultados que se obtengan de este con los del presente trabajo, con la finalidad de observar los cambios en la categorización de las AGEBs rurales a lo largo del tiempo. Ya que la única información sobre el campo, más actual y disponible es de la **Encuesta Nacional Agropecuaria 2017**, pero que no sustituye al Censo Agropecuario, debido a que presenta una menor cobertura y la información que contiene es menos amplia (CEDRSSA, 2019).
- Añadir información actualizada y relacionada a la degradación y desertificación del suelo, para futuros diagnósticos integrales relacionados a la productividad de cultivos agrícolas. Ya que para fines de este proyecto no fue posible incorporar información del mapa de “**Degradación del suelo en la República Mexicana Escala 1:250 000**” cuya fuente es Semarnat 2004, como otra variable ambiental (degradación del suelo), porque no posee datos suficientes para la mayoría de las AGEBs rurales del país. Y hoy en día gracias a la información presentada por el documento “**Línea base nacional de degradación de tierras y desertificación**” se sabe que el 63,2% del país presenta algún grado de degradación edáfica (CONAFOR-UACH, 2013). Así mismo, se tiene conocimiento que la degradación del suelo provoca alteraciones en el nivel de fertilidad del suelo y consecuentemente en su capacidad de sostener una agricultura productiva (FAO,1994), por lo que es importante considerar este dato para futuros análisis.
- Llevar a cabo este ejercicio con el modelo de árbol de regresión con otros cultivos de gran relevancia para el sector agrícola mexicano, como aquellos que conforman los 38 cultivos estratégicos definidos previamente por SAGARPA en la **Planeación Agrícola Nacional 2017-**

2030. Con el objetivo de identificar los principales factores que influyen en la productividad y que permiten clasificar a las AGEBs en las que se siembran dichos cultivos.

XII. REFERENCIAS

A

- Abdi, H., & Valentin, D. (2007). Multiple correspondence analysis. *Encyclopedia of measurement and statistics*, 2, 651-66.
- Alin, A. (2010). Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(3), 370-374.
- Altieri, M. A., & Koohafkan, P. (2008). *Enduring farms: climate change, smallholders and traditional farming communities* (Vol. 6). Penang: Third World Network (TWN).
- Ayala Ortiz, D. A., Schwentesius Rindermann, R., & Gómez Cruz, M. Á. (2008). La ecocondicionalidad como instrumento de política agrícola para el desarrollo sustentable en México. *Gestión y política pública*, 17(2), 315-353.

B

- Berdegú, J.A. y Fuentealba, R. 2011. Latin America: the state of smallholders in agriculture. Documento presentado en la Conferencia del FIDA titulada “New Directions for Smallholder Agriculture” (Nuevas direcciones para la agricultura en pequeña escala), 24-25 de enero de 2011. Roma, FIDA.
- Bojórquez-Tapia, L. A., Cruz-Bello, G. M., Luna-González, L., Juárez, L., & Ortiz-Pérez, M. A. (2009). V-DRASTIC: using visualization to engage policymakers in groundwater vulnerability assessment. *Journal of Hydrology*, 373(1-2), 242-255.
- Robles, B. H. (2011). Los productores de café en México: problemática y ejercicio del presupuesto. Washington DC: Woodrow Wilson International Center for Scholars. Mexican Rural Development Research Reports, 14.
- Berry, P. M., Sylvester-Bradley, R., Philipps, L., Hatch, D. J., Cuttle, S. P., Rayns, F. W., & Gosling, P. (2002). Is the productivity of organic farms restricted by the supply of available nitrogen?. *Soil Use and Management*, 18, 248-255.

C

- CEDRSSA (2018). Reporte. El café en México. Diagnóstico y Perspectiva. Cámara de Diputados LXIII

Legislatura. Palacio Legislativo de San Lázaro.

CEDRSSA (2019). Resultados de la Encuesta Nacional Agropecuaria 2017. Centro de Estudios para el Desarrollo Rural Sustentable y la Soberanía Alimentaria. Cámara de Diputados LXIV Legislatura. Palacio Legislativo de San Lázaro, Ciudad de México.

CEFP (2018). Programa Especial Concurrente para el Desarrollo Rural Sustentable, 2012-2018. Cámara de Diputados LXII Legislatura. Palacio Legislativo de San Lázaro, Ciudad de México.

Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T., Saisana, M., Saltelli, A., ... & Tarantola, S. (2008). Creating composite indicators with DEA and robustness analysis: the case of the Technology Achievement Index. *Journal of the Operational Research Society*, 59(2), 239-251.

Clifford, M. N. and Willson, K.C. (Editors): *Coffee; botany, biochemistry and production of beans and beverage*. London, Croom Helm, 1985.

Curiel, R. (2013). MasAgro por la seguridad alimentaria y el desarrollo agrícola sustentable en México. *Claridades Agropecuarias (México)*, 237, 9-18.

CONAFOR-UACH. 2013. Línea base nacional de degradación de tierras y desertificación. Informe final. Comisión Nacional Forestal (CONAFOR) y Universidad Autónoma Chapingo (UACH). Zapopan, Jalisco, México.

Contreras, P. E. Q., Avila, D. D., Mendoza, S. L. H., & González, C. G. (2016). Desarrollo económico del sector agropecuario en México a 20 años de la firma TLCAN. *Revista: TECSISTECATL* ISSN, 1886, 8452.

D

DaMatta FM, Loos RA, Silva EA, Ducatti C, Loureiro ME (2002) Effects of soil water deficit and nitrogen nutrition on water relations and photosynthesis of pot-grown *Coffea canephora* Pierre. *Trees Struct. Funct.* 16:555-558.

DaMatta FM, Ronchi CP, Sales EF, Araújo JBS (2007) O café conilon em sistemas agroflorestais. In: Ferrão RG, Fonseca AFA, Bragança SM, Ferrão MAG, De Muner LH (eds), *Café Conilon*, pp.377-389.

de Grammont, H. C. (2010). La evolución de la producción agropecuaria en el campo mexicano: concentración productiva, pobreza y pluriactividad. *Andamios*, 7(13), 85-117.

DOF (2001). *Ley de Desarrollo Rural Sustentable*. Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación. México.

DOF (2014). *Programa Institucional 2013-2018 de la Financiera Nacional de Desarrollo Agropecuario, Rural, Forestal y Pesquero*. Secretaría de Gobernación. México.

Donoho, D. (2015). 50 years of data science. In *Tukey centennial workshop* (pp. 1-41). NJ: Princeton.

E

Esteban, M. (2008). Estadística actuarial: regresión lineal. SARRIKO-ON. (En línea en: <http://www.sarriko-online.com/cas/fichas/2008/ficha0308.htm>).

F

FAO (1994). Erosión de suelos en América Latina. Santiago, Chile. (En línea en: <http://www.fao.org/3/t2351s/T2351S06.htm#Relaci%C3%B3n%20entre%20erosi%C3%B3n%20y%20p%C3%A9rdida%20de%20fertilidad%20del%20suelo>)

FAO. (2005). Agricultura y diálogo de culturas: nuestro patrimonio común.
<http://www.fao.org/3/a0015s/a0015s00.htm>

FAO. (2011). El estado mundial de la agricultura y la alimentación 2010-2011. Las mujeres en la agricultura: Cerrar la brecha de género en aras del desarrollo. FAO, Roma.

FAO (2013). Climate smart agriculture sourcebook.

FAO (2014). Los jóvenes y la agricultura: Desafíos clave y soluciones concretas. (En línea en: <http://www.fao.org/3/a-i3947s.pdf>).

FAO (2015a). Construyendo una visión común para la agricultura y alimentación sostenibles. Principios y enfoques. Roma.

FAO (2015b). World reference base for soil resources 2014. Internacional soil classification system for naming soils and creating legends for soil maps. World Soil Resources Reports 106. Roma.

FAO-SAGARPA. (2012c). Agricultura familiar con potencial productivo en México. FAO-Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación, México.

Financiera Rural. (2005). Evaluación externa 2004. Evaluación del componente de apoyo para la constitución de IFR del PAASFIR.

FIRA (2016). Panorama Agroalimentario. Café 2016. Dirección de Investigación y Evaluación Económica y Sectorial. Fideicomisos Instituidos en Relación con la Agricultura.

Fox, J., & Haight, L. (2010). La política agrícola mexicana: metas múltiples e intereses en conflicto. Subsidios para la desigualdad. Las políticas públicas del maíz en México a partir del libre comercio, 9-54.

Freudenberg, M. (2003): Composite indicators of country performance: a critical assessment. OCDE Science.

G

Giovannucci, D. and F.J. Koekoek. 2003. The state of sustainable coffee: A study of twelve major markets. Philadelphia, PA: International Institute for Sustainable Development.

H

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). Multivariate data analysis: Pearson new international edition. Pearson Higher Ed.

- Hayashi, Chikio (1 January 1998). "What is Data Science? Fundamental Concepts and a Heuristic Example". In Hayashi, Chikio; Yajima, Keiji; Bock, Hans-Hermann; Ohsumi, Noboru; Tanaka, Yutaka; Baba, Yasumasa (eds.). *Data Science, Classification, and Related Methods. Studies in Classification, Data Analysis, and Knowledge Organization*. Springer Japan. pp. 40–51. doi:10.1007/978-4-431-65950-1_3.
- Hernández-Trujillo, J. M. (2016). Cortadores de café en México. El inframundo del trabajo decente. *Ra Ximhai*, 12(4), 93-110.
- Herrero, M., Thornton, P. K., Power, B., Bogard, J. R., Remans, R., Fritz, S., ... & Watson, R. A. (2017). Farming and the geography of nutrient production for human use: a transdisciplinary analysis. *The Lancet Planetary Health*, 1(1), e33-e42.
- HLPE. 2013. Investing in smallholder agriculture for food security. A report by the High Level Panel of Experts on Food Security and Nutrition of the Committee on World Food Security, Rome.
- Huerta, A. 1994. La política neoliberal de estabilización económica en México. Límites y alternativas. Instituto de Investigaciones Económicas, Universidad Nacional Autónoma de México, UNAM.
- INEGI-SAGARPA, 2015, "Nota técnica Encuesta nacional agropecuaria 2014", en Conociendo el campo de México, Instituto Nacional de Estadística Geografía y Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación, Aguascalientes, México, pp. 20.

I

- International Labour Office. (2018). *World employment and social outlook: trends 2015*. Geneva: International Labour Organization.
- International Trade Center (ITC). 2011. Trends in the trade of certified coffees. Technical Paper. Geneva, Switzerland.

K

- Kathait S. (2017). Variable Reduction: An art as well as science. Data Science Central. (En línea en: <https://www.datasciencecentral.com/profiles/blogs/variable-reduction-an-art-as-well-as-science>).
- Kinuthia K, J., Shadrack K, I. & Nakhone L (2018). Factors Influencing Farmer's Choice of Crop Production Response Strategies to Climate Change and Variability in Narok East Sub-county, Kenya. *Journal of Natural Resources and Development*. Volume 8, Pages:69 – 77. Doi: 10.5027/jnrd.v8i0.07
- Kuntz-Ficker, S. (coordinadora) (2010). *Historia económica general de México : de la Colonia a nuestros días*. México, D.F.: El Colegio de México : Secretaría de Economía. (2a. reimpresión, 2015). 834 p.

L

- Levins, R. (2004). Sorpresas, errores y dudas. *Revista Cubana de Salud Pública*, 30(3). (En línea en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-34662004000300006).
- Lewontin, R. C., & Levins, R. (2007). *Biology under the influence: Dialectical essays on ecology, agriculture, and health* (No. 57.02 LEW).

Luna, F., & Wilson, P. N. (2015). An economic exploration of smallholder value chains: Coffee transactions in Chiapas, Mexico. *International Food and Agribusiness Management Review*, 18: 85-106.

M

Masera, O. C. (1990): *Crisis y mecanización de la Agricultura Campesina*, México D.F., Edit. El Colegio de México.

Macías, A. M. (2013). Introducción. Los pequeños productores agrícolas en México. *Carta económica regional*, (111-112).

Manion, M., G. Dicum, N. Luttinger, G. Richards, J.J. Hardner, & T. Walker. 1999. *The Scale and Trends of Coffee Production Impacts on Global Biodiversity*. Paper prepared by Industrial Economics, Inc. for the Center for Applied Biodiversity Science, Conservation International, Washington, D.C.

Mansfield, E. R., & Helms, B. P. (1982). Detecting Multicollinearity. *The American Statistician*, 36(3a), 158–160. doi:10.1080/00031305.1982.10482818

Martínez G, M.V (2007). Intervención estatal en el campo mexicano. En *Kurczyn V, P(coordinadora). Panorama internacional de derecho social : culturas y sistemas jurídicos comparados*. Instituto de Investigaciones Jurídicas, UNAM, México, D.F.

Martínez O, RM., Tuya P., LC, Martínez O, M., Pérez A, A. & Cánovas, AM. (2009). El Coeficiente de Correlación de los rangos de Spearman. Caracterización. *Revista Habanera de Ciencias Médicas*, 8(2).

Medrano G, J.A (2017). Análisis Exploratorio de Datos como Herramienta de Solución de Problemas. *Revista de la Escuela de Estudios de Posgrado*, Vol. 8 No. 1 ISSN 2518-4725.

Menéndez F (2002). Residuos, Residuos studentizados y valores DFFIT. Su uso en Regresión Lineal Simple y Múltiple. Departamento de Sociología. Universidad de la República. (En línea en: <http://tabarefernandez.tripod.com/coco2.pdf>).

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 63(2), 81.

Moguel, P. & V. Toledo. 1996. El café en México, ecología, cultura indígena y sustentabilidad. *Ciencias* 43:40-51.

Moguel, P., & Toledo V.M (2004). Conervar produciendo: Biodiversidad, café orgánico y jardineras productivos. *CONABIO. Biodiversitas* 55:1-7

Morón, C., & Schejtman, A. (1997). *Situación de la seguridad alimentaria en América Latina*. Universidad de Chile-FAO.

N

Najera, O. (2002). El café orgánico en México. *Cuadernos de desarrollo Rural*, (48).

Nellemann, C. (Ed.). (2009). *The environmental food crisis: the environment's role in averting future food crises*:

a UNEP rapid response assessment. UNEP/Earthprint.

O

Olivares, B. (2014). Aplicación del Análisis de Componentes Principales (ACP) en el diagnóstico socioambiental. Caso: sector Campo Alegre, municipio Simón Rodríguez de Anzoátegui. *Multiciencias*, 14(4).

Orellana A, J. (2018). Árboles de decisión y Random Forest. (En línea en: <https://bookdown.org/content/2031/>)

Orzack, S. H., & Sober, E. (1993). A critical assessment of Levins's The strategy of model building in population biology (1966). *The Quarterly Review of Biology*, 68(4), 533-546.

P

Palomares R. JA, González S.J D, & Mireles R. SC (2012). Investigación: café orgánico en México. UNAM, México.

Peña, D. (2002). Análisis de Datos Multivariantes. Madrid: McGraw Hills.

Power, A. G. (2010). Ecosystem services and agriculture: tradeoffs and synergies. *Philosophical transactions of the royal society B: biological sciences*, 365(1554), 2959-2971.

Q

Quintero Peralta, M. A. (2017). Políticas públicas, soberanía alimentaria y estrategias campesinas en zonas rurales pobres de México.

R

Rodríguez, J. G. L., & Balmón, M. A. (2011). Evaluación de la sostenibilidad de las explotaciones de olivar en Andalucía.

Rodríguez-Salazar, M E., Álvarez S., Bravo-Núñez E. (2000). Coeficientes de asociación. Universidad Autónoma Metropolitana-Iztapalapa. Plaza y Valdés. México.

Rosenzweig, A. (2005). El debate sobre el sector agropecuario mexicano en el Tratado de Libre Comercio de América del Norte. Cepal.

Ruíz, C (coordinador). (2002). Microfinanzas: mejores prácticas a nivel nacional e internacional. México: Universidad Nacional Autónoma de México–Gobierno del Distrito Federal.

S

Sáenz L, K., & Tamez G.,G (2004). Métodos y técnicas cualitativas y cuantitativas aplicables a la investigación en ciencias sociales. *México: Tirant Lo Blanch*.

Sánchez Almanza, A. (1985). La política agropecuaria contemporánea en México y la coyuntura de los terremotos. *Problemas del Desarrollo*, 16(62/63), 149-171.

- Sánchez Cano, J. E. (2014). La política agrícola en México, impactos y retos. *Revista Mexicana de Agronegocios*, 35(1345-2016-104400), 946.
- Sánchez Z, C. C., Giraldo, M., L.M , Piedrahita, E., C. C., Bonet, I., Lochmuller, C., Tabares, B., M. S., & Peña, A. (2018). Análisis comparativo entre: «el análisis exploratorio de datos» y los modelos de «árboles de decisión» y «k-means» en el diagnóstico de la malignidad en algunos exámenes de cáncer de mama. Un estudio de caso. *Revista ESPACIOS*, 39(28).
- SAGARPA (2015). Convención Internacional del Café México 2015. Carpeta de Difusión, 22.
- SAGARPA (2017). Café Mexicano. Planeación Agrícola Nacional (2017-2030). Ciudad de México, México.
- Scheffer, M., & Carpenter, S. R. (2003). Catastrophic regime shifts in ecosystems: linking theory to observation. *Trends in ecology & evolution*, 18(12), 648-656.
- Shriar, A. J. (2000). Agricultural intensity and its measurement in frontier regions. *Agroforestry Systems*, 49(3), 301–318. <http://doi.org/10.1023/A:1006316131781>
- SIAP-SAGARPA (2016). Atlas Agroalimentario 2016. México siembra éxito. Servicio de Información Agroalimentaria y Pesquera. México. (En línea en: https://nube.siap.gob.mx/gobmx_publicaciones_siap/pag/2016/Atlas-Agroalimentario-2016)
- SIAP-SAGARPA (2017). Atlas Agroalimentario 2017. México siembra éxito. Servicio de Información Agroalimentaria y Pesquera. México. (En línea en: <http://online.pubhtml5.com/clsi/ibhs/#p=4>)
- Singh, B. K., Trivedi, P., Singh, S., Macdonald, C. A., & Verma, J. P. (2018). Emerging microbiome technologies for sustainable increase in farm productivity and environmental security. *Microbiology Australia*.
- Song, Y. Y., & Ying, L. U. (2015). Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry*, 27(2), 130.
- Sosa-Baldivia, A., Ruiz-Ibarra, G., Gordillo-Sobrino, G., West, H., Sharma, M., Liu, X., & Robles-de la Torre, R. R. (2017). Fecha de siembra: un método para evadir el ataque de *Diabrotica speciosa* Germar (Coleoptera: Chrysomelidae) y aumentar el rendimiento de semilla en chíca (*Salvia hispanica* L.). *Entomología Mexicana*, 4, 277-284.
- SRA, 2010. Las transformaciones del cardenismo. Documento en línea en: <http://www.sra.gob.mx/sraweb/conoce-la-sra/historia/las-transformaciones-del-cardenismo/> (consultada en septiembre de 2012).

T

- Tauer, L. W. (1984). Productivity of farmers at various ages. *North Central Journal of Agricultural Economics*, 81-87.
- Taylor, R. (1990). Interpretation of the correlation coefficient: a basic review. *Journal of diagnostic medical sonography*, 6(1), 35-39
- Téllez L, A. (2003). La memoria humana: Revisión de los hallazgos recientes y propuesta de un modelo

neuropsicológico. Tesis de Maestría. Universidad Autónoma de Nuevo León. Facultad de Filosofía y Letras. División de Estudios de Posgrado.

Timofeev R. (2004). Classification and Regression Trees (CART). Theory and Applications. Center of Applied Statistics and Economics Humboldt University, Berlin.

Traver, M. B. (2014). Aplicaciones prácticas del Análisis de Componentes Principales en Gestión de Carteras de Renta Fija (II): Gestión del riesgo de tipo de interés de una cartera mediante PCA. Análisis Financiero, (124), 39-54.

V

Van der Vossen, H. A. M. (2005). A critical analysis of the agronomic and economic sustainability of organic coffee production. Experimental agriculture, 41(4), 449-473.

W

Weihs C, & Ickstadt K (2018). Data Science: the impact of statistics. International Journal of Data Science and Analytics (<https://doi.org/10.1007/s41060-018-0102-5>)

Willians, G. (2011). Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery. Londres: Springer.

Woodsong, C. (1994). Old farmers, invisible farmers: Age and agriculture in Jamaica. Journal of cross-cultural gerontology, 9(3), 277-299.

Y

Ye, N. (2014). Data Mining. Theories, Algorithms, and Examples. New York: CRC press.

Yúñez, A. N., Rivera, R.F., Chávez, A. M., Mora, R. J.J., Taylor J. E (2015). La economía del campo mexicano: Tendencias y retos para su desarrollo. El Colegio de México AC.

Z

Zubcoff J. J. (2017). FAMD. Factor analysis of mixed data. Departamento de Ciencias del Mar y Biología Aplicada. Universidad de Alicante (En línea en: <https://rua.ua.es/dspace/bitstream/10045/72567/3/Analisis-multivariante-FAMD.pdf>).

XIII. ANEXOS

1. Metodología

a. Proceso de agregación



Figura A1. Ejemplo simulado del proceso de agregación.

b. AGEBS rurales de México en dónde se cultiva café.



Figura A2. Mapa de la República mexicana que muestra señaladas en color café las AGEBS rurales en dónde se cultiva café, de acuerdo al Censo Agropecuario 2007. El mapa presenta como fondo la capa "El relieve como atractivo natural" descargada del catálogo de metadatos geográficos de CONABIO.

c. Análisis estadísticos

❖ **Multicolinealidad**

La multicolinealidad se refiere a la relación lineal entre dos o más variables, lo que también significa falta de ortogonalidad¹¹ entre ellas. Esta relación también se llama colinealidad o mal condicionamiento. La multicolinealidad puede considerarse como la dependencia lineal entre las variables independientes. (Alin, 2010).

El análisis conjunto de datos para la existencia de multicolinealidad suele realizarse como un paso inicial en cualquier análisis de regresión múltiple. El principal problema con la multicolinealidad es que los estimadores de mínimos cuadrados de los coeficientes de las variables involucradas en las dependencias lineales tienen grandes varianzas (Mansfield & Helms, 1982).

El *Factor de inflación de varianza* (VIF, por sus siglas en inglés) es un indicador que proporciona al usuario una medida de cuántas veces mayor será la varianza para datos multicolineales que para datos ortogonales (donde cada VIF es 1.0). Si los VIF no son inusualmente mayores que 1.0, entonces la multicolinealidad no es un problema. Una ventaja de conocer el VIF para cada variable es que brinda a los usuarios una sensación tangible de cuánto se ven afectadas las variaciones de los coeficientes estimados por la multicolinealidad (Mansfield & Helms, 1982).

En muchos estudios, la multicolinealidad se ha confundido con la *correlación*. La correlación es la relación lineal entre solo dos variables, mientras que la multicolinealidad puede existir entre dos variables o entre una variable y una combinación lineal de las otras. Por lo tanto, la correlación es un caso especial de multicolinealidad. La alta correlación implica multicolinealidad, pero lo inverso no es verdadero (Alin, 2010).

❖ **Correlación**

El análisis de correlación es uno de los más utilizados para analizar datos de investigación científica y se utiliza para determinar si existe una relación entre dos diferentes variables y si sí existe, cuán significativa o cuán fuerte es la asociación entre esas dos variables (Taylor, 1990).

En sí el concepto de relación o correlación se refiere al grado de variación conjunta que existe entre dos o más variables. Una relación lineal *positiva* entre dos variables indica que los valores de las dos variables

¹¹ En geometría elemental, ortogonal es lo mismo que perpendicular (<http://mathworld.wolfram.com/Orthogonal.html>)

varían de forma parecida, mientras que una relación lineal *negativa* significa que los valores de las dos variables varían justamente de forma contraria o invertida (Sáenz & Tamez, 2004):

Los coeficientes de correlación son índices numéricos que permiten cuantificar el grado de relación lineal existente entre dos variables cuantitativas, al mismo tiempo que sirven para valorar el grado de ajuste de la nube de puntos a una línea recta. Para realizar un análisis de correlación de dos variables, se seleccionan uno o más de los siguientes coeficientes (Sáenz & Tamez, 2004):

- *Pearson*: es un método paramétrico, que asume una distribución normal de los datos y una asociación lineal entre las variables X y Y . Se representa por r .
El coeficiente de correlación de Pearson toma valores entre -1 y 1, un valor de 1 indica relación lineal perfecta positiva, mientras que un valor de -1 indica relación lineal perfecta negativa, y un valor de 0 indica relación lineal nula.
- *Tau-b de Kendall*: Este coeficiente de correlación es apropiado para estudiar la relación entre variables ordinales. Se basa en el número de inversiones y no inversiones entre casos. Toma valores entre -1 y 1, y se interpreta igual que el coeficiente de correlación de Pearson. La utilización de este coeficiente tiene sentido si las variables no alcanzan el nivel de medida de intervalo y/o no podemos suponer que la distribución poblacional conjunta de las variables sea normal.
- *Spearman*: es una medida de asociación no paramétrica, que puede evaluar la relación no lineal entre dos variables sin hacer suposiciones sobre su distribución de probabilidad. El coeficiente de correlación ρ de Spearman es el coeficiente de correlación de Pearson, pero aplicado después de transformar las puntuaciones originales en rangos. Toma valores entre -1 y 1, y se interpreta igual que el coeficiente de correlación de Pearson. Al igual que ocurre con el coeficiente *tau-b* de Kendall, el de Spearman puede utilizarse como una alternativa al de Pearson cuando las variables estudiadas son ordinales y/o se incumple el supuesto de normalidad. También el coeficiente ρ de Spearman resulta ser menos influenciado por la presencia de datos sesgados.

❖ **Análisis Factorial de Datos Mixtos (FAMD)**

Método multivariado de análisis de datos que reduce la dimensión de los datos, el término mixto se debe a que permite el tratamiento simultáneo de variables de distinta naturaleza (cualitativa o cuantitativa) para describir un mismo conjunto de individuos. En el FAMD se puede estudiar la similaridad entre los individuos teniendo en cuenta todas las variables y la relación entre las variables (Zubcoff, 2017).

En términos generales, podemos decir que FAMD funciona como *un Análisis de Componentes Principales* para variables cuantitativas y como un *Análisis de Correspondencia Múltiple (MCA)* para variables cualitativas (a continuación, se hablan de estos dos análisis a más detalle).

➤ *Análisis de Componentes Principales*

El Análisis de Componentes Principales (PCA, por sus siglas en inglés) es un método estadístico de análisis multivariado que se utiliza para extraer información relevante de un conjunto inicial de variables correlacionadas transformándolas en variables no correlacionadas, con el objetivo de identificar patrones y estructuras, y cuantificar la importancia de cada una de ellas. Al conjunto de variables iniciales, se les denomina *originales*, y a las nuevas variables se les denomina *componentes principales*, los cuales se caracterizan por ser combinaciones lineales de las variables originales (de la Fuente, 2011).

Los componentes principales se pueden ordenar de acuerdo con la información que llevan incorporada, y como medida de la cantidad de información incorporada en un componente se utiliza su **varianza**. Por esta razón se selecciona como primer componente aquel que tenga mayor varianza, mientras que el último componente es el de menor varianza (de la Fuente, 2011).

El PCA es considerado un *método de reducción*, ya que permite reducir la dimensión del número de variables originales que se han considerado en el análisis. (de la Fuente, 2011; Ye, 2014).

Técnicas como el PCA fueron desarrolladas y se aplicaron por primera vez en áreas como las ciencias sociales y la econometría (Ye, 2014). Actualmente el PCA ha sido utilizado por varias ciencias, siendo de gran utilidad en tareas de diagnóstico o predicción. Este método ha sido aplicado por ejemplo en la caracterización de sistemas de producción agrícola, tecnología de alimentos y desarrollo de productos, agronomía, uso de información climática, y evaluación de suelos productivos (Olivares, 2014).

➤ *Análisis de correspondencia múltiple (MCA)*

El análisis de correspondencia múltiple (MCA) es una extensión del análisis de correspondencia (CA) que permite analizar el patrón de relaciones de varias variables dependientes categóricas. Como tal, también puede verse como una generalización del análisis de componentes principales cuando las variables a analizar son categóricas en lugar de cuantitativas. Técnicamente, el MCA se obtiene utilizando un análisis de correspondencia estándar en una matriz de indicadores (es decir, una matriz

cuyas entradas son 0 o 1). Los porcentajes de varianza explicada deben corregirse, y la interpretación del análisis de correspondencia de las distancias entre puntos debe adaptarse (Abdi & Valentin, 2007).

❖ **Árbol de decisión**

El árbol de decisión es uno de los métodos más efectivos de minería de datos y es utilizado extensamente en numerosas disciplinas porque es fácil de usar, es libre de ambigüedad, y es robusto inclusive en presencia de valores faltantes (Song y Ying, 2015). El árbol de decisión se trata de un modelo de predicción o clasificación que fue introducido en 1960 y cuyo objetivo principal es el aprendizaje inductivo a partir de observaciones y construcciones lógicas (Ye, 2014).

A través de la forma de un árbol, captura la relación existente entre variables atributos x ($x_1, x_2 \dots x_p$) con una variable objetivo y ($y = F(x)$). Los árboles de decisión pueden ser de dos tipos según la naturaleza de la variable objetivo (Ye, 2014):

-*Árbol de Clasificación*. Si la variable objetivo es categórica, se dice que el árbol de decisión que se forma a partir de ella es un árbol de clasificación.

-*Árbol de Regresión*. Si predice los valores de una variable objetivo de tipo numérica.

Los principales usos de un árbol decisión son: (a) la selección de variables, (b) acceder a la importancia relativa de las variables, (c) manejo de los datos faltantes, (d) predicción, y (e) manipulación de datos.

d. Indicador de productividad

❑ **Ley de Weber-Fechner**

La ley de Weber-Fechner proviene del campo de la psicofísica, y lo que hace es describir la relación entre las magnitudes físicas de los estímulos y la intensidad percibida de los estímulos. Esta ley establece que la percepción es proporcional a los aumentos de un determinado estímulo, que se puede notar solo después de que aumenta en un porcentaje constante, conocido como la "diferencia notable". Esta diferencia notable puede interpretarse como el incremento más pequeño necesario para poder discriminar entre los grados de un estímulo. La relación entre la diferencia notable y la percepción es logarítmica; mientras que la diferencia notable aumenta después de una progresión geométrica (es decir, multiplicada por un factor de progresión, en este caso: 2), la percepción correspondiente varía como una progresión aritmética (es decir, en cantidades constantes aditivas). De esta manera, el nivel de productividad debe duplicarse en valor (es decir, 2^1) para ser percibido como el doble de fuerte

(es decir, $1 + 1$), y debe cuadruplicarse en valor (es decir, 2^2) para ser percibido como tres veces más fuerte (es decir, $1 + 1 + 1$) (Bojórquez-Tapia et al., 2009).

❑ Numero de Miller

Se consideró lo escrito por George Miller en 1956 en su trabajo “*El mágico número siete, más o menos dos: Algunos límites en nuestra capacidad de procesar la información*”. Este documento es uno de los textos más citados en la psicología, ya que Miller menciona la existencia de limitaciones severas en la cantidad de información que podemos recibir, procesar y recordar. El número de objetos que un humano promedio puede tener en la memoria de trabajo¹², de acuerdo con la evidencia que Miller propone en su documento, es entre 5 y 9, es decir, 7 ± 2 (Miller, 1956; Téllez, 2003).

2. Resultados

a. Variables obtenidas

Tabla A1. **Grupo 1**, variables de “características sociodemográficas del productor”.

| Nombre de la variable | Abreviatura | Nombre de la variable | Abreviatura |
|--|-------------|---|----------------|
| obtuvo algún crédito/préstamo | crédito | habla español | habla_esp |
| contaba con seguro | seguro | habla lengua indígena | lengua_ind |
| apoyo por parte de los programas de gobierno | apoyo_gob | asistió a la escuela | escuela |
| ahorro de ingresos | ahorro | ningún nivel aprobado en escuela | ning_gdo_aprob |
| edad del productor | edad_prod | otro nivel estudiado | otro_nvl_est |
| dependientes económicos | dep_econ | vivienda con agua entubada | agua_ent |
| ingreso por actividad agrícola y/o forestal | i_act_agr_f | drenaje conectado a red pública | drenaj_pub |
| ingreso por envío de dinero | i_env_din | drenaje conectado a fosa séptica | drenaj_fosa_s |
| ingreso por apoyo del gobierno | i_apoy_gob | energía eléctrica | energía |
| ingreso por otro actividad | i_otra_act | gas para cocinar | gas |
| productor hombre | prod_hom | sanitario/letrina/excusado o pozo ciego | sanitario |
| productora mujer | prod_muj | piso de cemento/madera/otro | piso |
| | | paredes de mampostería/tabique/otro | paredes |

¹² **Memoria de trabajo:** es un tipo de memoria a corto plazo que se compone de dos sistemas, un sistema de memoria de corta duración y un sistema ejecutivo de que usa esa memoria para la ejecución de operaciones cognitivas. La memoria de trabajo dura en promedio de 20 a 30 segundos (Téllez, 2003).

Tabla A2. **Grupo 2**, variables de “manejo del cultivo”.

| Nombre de la variable | Abreviatura | Nombre de la variable | Abreviatura |
|--|-----------------|---|-----------------|
| productividad | prod | superficie ensalitrada | sup_ensalitrada |
| superficie de cultivo orgánico | sup_org | tiene superficie erosionada | erosionada |
| uso de fertilizantes químicos | fert | superficie erosionada | sup_erosionada |
| uso de semilla mejorada | sem_mej | tiene superficie cubierta con agua | cub_agua |
| uso de abonos naturales | abono | superficie cubierta de agua | sup_cub_agua |
| uso de herbicidas químicos | h_q | tiene superficie contaminada | contaminada |
| uso de herbicidas orgánicos | h_o | superficie contaminada | sup_contaminada |
| uso de insecticidas químicos | i_q | tiene otra superficie sin vegetación | o_sn_vg |
| uso de insecticidas orgánicos | i_o | superficie sin vegetación (otra causa) | sup_o_sn_veg |
| realiza quemas controladas | quema | tiene superficie cubierta con selva | selva |
| superficie con cultivos intercalados | sup_intercal | superficie cubierta con selva | sup_selva |
| cultivos intercalados | intercal | tiene superficie cubierta con bosque | bosque |
| dejó una parte sin sembrar | sin_semb | superficie cubierta con bosque | sup_bosque |
| tiene vivero | vivero | corte de árboles en la superficie selva/bosque | c_a_sb |
| superficie que ocupa el vivero | sup_vivero | pastoreo/ramoneo de ganado en superficie arbolada | past_sb |
| tiene invernadero | invern | reforestación en superficie arbolada | refor_sb |
| superficie que ocupa el invernadero | sup_invernadero | se plantaron árboles forestales en sus terrenos | arb_terr_prod |
| realiza control biológico de plagas | Ctrl_bio | plantación de especies de la región | p_spp |
| realiza injerto de árboles | injerto_arb | desmonte de superficie selva/bosque/monte | desm_sb |
| realiza rotación de cultivos | rotacion_cult | superficie desmontada que pasó a uso agrícola | sup_desm_agr |
| realiza podas | podas | superficie desmontada que pasó a uso ganadero | sup_desm_gd |
| selecciona semillas para la siembra | sel_sem | superficie desmontada que pasó a otro uso | sup_desm_o |
| tiene superficie de temporal | temporal | realizó aclareo en selva/bosque | aclareo |
| tiene superficie de riego | riego | superficie de aclareo en selva/bosque | sup_aclareo |
| recibió capacitación | capacit | uso de herbicidas químicos en selva/bosque | h_q_sb |
| recibió asistencia técnica | asist_tec | superficie herbicidas químicos en selva/bosque | sup_h_q_sb |
| profundidad a la que extrae agua | prof_a | uso de herbicidas orgánicos en selva/bosque | h_o_sb |
| tiene superficie con pastos no cultivados/ de agostadero/enmontada | enmontada | superficie herbicidas orgánicos en selva/bosque | sup_h_o_sb |
| superficie con pastos no cultivados/de agostadero/enmontada | sup_enm | uso de insecticidas químicos en selva/bosque | i_q_sb |
| corte de árboles en la superficie enmontada | c_a_enm | superficie insecticidas químicos en selva/bosque | sup_i_q_sb |
| superficie con pastos no cultivados principalmente | sup_pastos | uso de insecticidas orgánicos en selva/bosque | i_o_sb |
| superficie enmontada sembrada | sup_enm_semb | superficie insecticidas orgánicos en selva/bosque | sup_i_o_sb |
| tiene superficie sin vegetación | sin_veg | contaba con plan para control de incendios | ctrl_i_sb |
| tiene superficie con arenales o pedregales | arenal_pedregal | realizó control biológico de plagas en selva/bosque | ctrl_b_sb |
| superficie con arenales o pedregales | sup_arenales | realizó quema controlada en selva/bosque | quema_sb |
| tiene superficie ensalitrada | ensalitrada | recibió asistencia técnica en selva/bosque | asist_sb |
| contaba con aserradero | aserra_sb | uso de otra tecnología en selva/bosque | o_tec_sb |
| contaba con vivero forestal | vivero_sb | superficie con otra tecnología en selva/bosque | sup_o_tec_sb |

Tabla A3. **Grupo 3**, variables en la categoría de “procesamiento, transporte, venta y consumo”.

| Nombre de la variable | Abreviatura | Nombre de la variable | Abreviatura |
|--|-------------|--|-------------|
| uso de camiones para comercializar producción | uso_camion | parte de la producción para consumo animal | con_anim_pt |
| uso de ferrocarril para comercializar la producción | uso_ferro | nada de la producción para consumo animal | con_anim_nd |
| uso de transporte aéreo para comercializar la producción | uso_avion | venta en el país para mercado nacional, regional o local | venta |
| uso de embarcaciones para comercializar la producción | uso_embarc | venta a un intermediario (coyote) | v_interm |
| uso de otro medio para comercializar la producción | uso_o_med | venta a un mayorista | v_may |
| no usa ningún medio para comercializar la producción | no_medio | venta a una cadena comercial | v_comer |
| cuanto vendió o espera vender | vender | venta a una empacadora o agroindustria | v_agroi |
| toda la producción a consumo familiar | con_fam_td | venta a otros compradores | v_otro |
| parte de la producción para consumo familiar | con_fam_pt | venta directamente al extranjero | v_ext |
| nada de la producción para consumo familiar | con_fam_nd | contaba con beneficiadora de café o cacao | b_cafe |
| toda la producción para consumo animal | con_anim_td | procesa o transforma parte de la producción | transf |

Tabla A4. **Grupo 4**, variables “relacionadas a la organización del productor”.

| Nombre de la variable | Abreviatura |
|---|------------------|
| organización con otros productores para obtener algún apoyo o servicio | org_p_apoy (105) |
| organización individual para realizar actividades agropecuarias/forestales | org_indiv |
| organización familiar (esposa e hijos) para realizar actividades agropecuarias/forestales | org_fam |
| organización como un grupo o cooperativa para realizar actividades agropecuarias/forestales | org_gpo_coop |
| organización como una empresa para realizar actividades agropecuarias/forestales | org_emp |
| participación de familiares del productor para realizar labores agropecuarias/forestales | part_fam |
| se tuvieron personas contratadas para labores agropecuarias/forestales | contratados |

Tabla A5. **Grupo 5**, variables “relacionadas con la actividad y problemática principal”.

| Nombre de la variable | Abreviatura | Nombre de la variable | Abreviatura |
|--|------------------|---|--------------|
| agricultura es la principal actividad | p_act_agr | no sembró para dejarla en descanso | en_descanso |
| cría y explotación de animales es la principal actividad | p_act_anim | superficie en descanso | sup_descanso |
| corte de árboles es la principal actividad | p_act_arb | no sembró por otra causa | o_causa_ns |
| recolección de productos silvestres es la principal actividad | p_act_recol | inconveniente para el acceso al crédito problema principal | pro_cred |
| otra actividad principal | p_act_otra | pérdida de fertilidad del suelo principal problema | pro_p_fertil |
| realiza alguna actividad diferente a la agrícola/ganadera/forestal | act_dif | siniestros naturales principal problema | pro_sin |
| superficie sin sembrar | sin_semb | dificultades para la comercialización problema principal | pro_d_com |
| no sembró porque no le interesó | no_intereso | organización poco apropiada para la producción problema principal | pro_org |
| no sembró por falta de dinero o apoyos | falta_dinero | infraestructura insuficiente para la producción problema principal | pro_infrae |
| no sembró por mal temporal | mal_temporal | altos costos de insumos y servicios problema principal | pro_insumos |
| no sembró porque no hubo quien la sembrara | no_hubo_quien | falta de capacitación y asistencia técnica problema principal | pro_asist |
| no sembró porque estaba invadida | estaba_invadida | litigios por la tierra problema principal | pro_litigio |
| no sembró porque son suelos pocos fértiles | suelo_p_fertil | falta de documentación para acreditar la posesión de la tierra problema principal | pro_acred |
| no sembró porque son suelos erosionados | suelo_erosionado | otro problema principal | pro_otro |

Tabla A6. **Grupo 6**, variables “características ambientales”.

| Nombre de las variables | Abreviatura |
|---------------------------|---------------|
| Temperatura media anual | temperatura |
| tipo de clima | clima |
| precipitación media anual | precipitación |
| textura del suelo | textura_s |
| tipo de suelo | suelo |
| pendiente del terreno | pendiente |
| uso potencial del suelo | uso_suelo |

b. Gráficos de contribución del FAMD

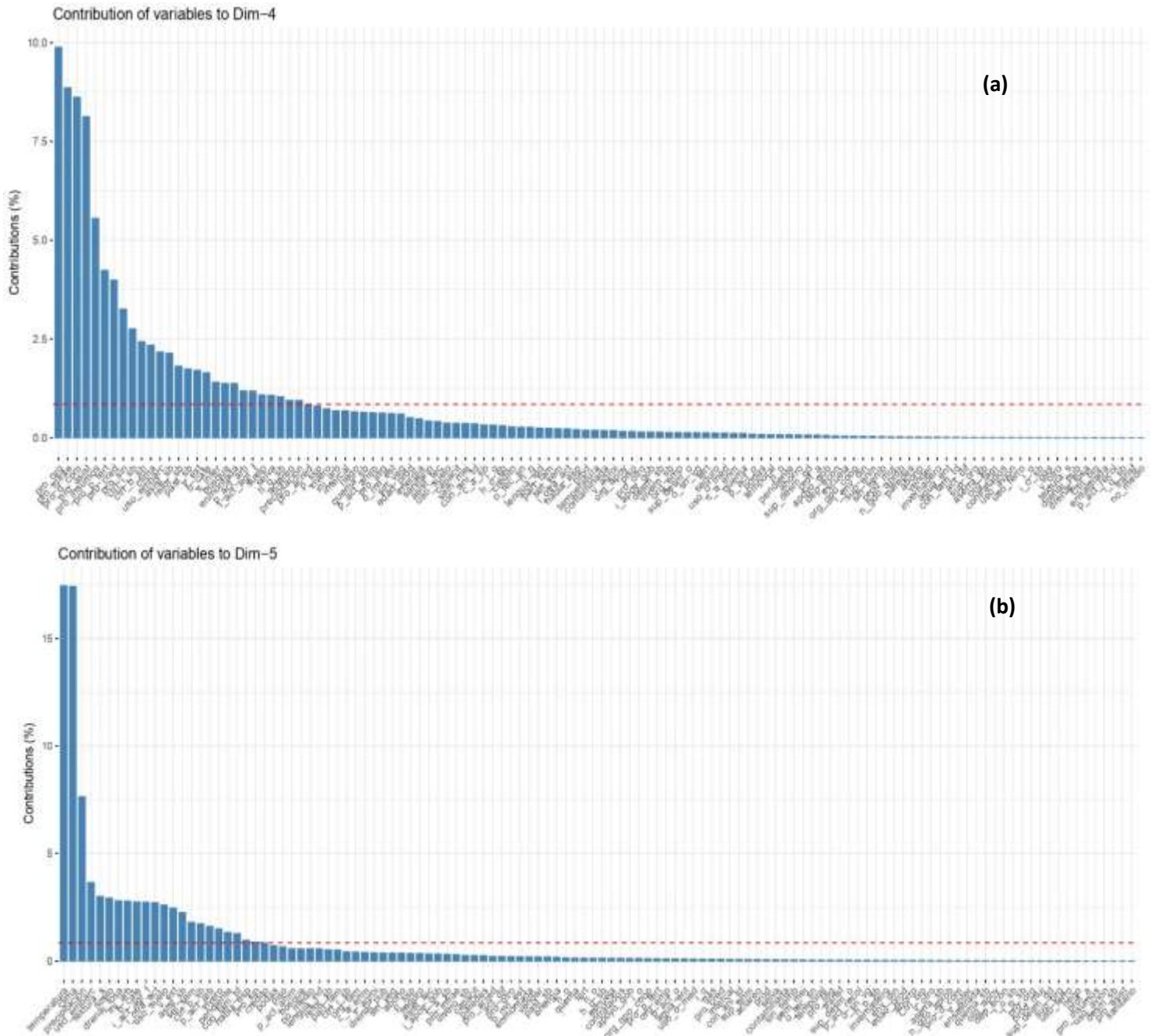


Figura A3. Gráficos del porcentaje de contribución de las variables de las dimensiones 4 y 5, generadas a partir del FAMD. (a) Porcentaje de contribución de la dimensión 4, y (b) Porcentaje de contribución de la dimensión 5. La línea roja punteada corresponde a la media esperada del porcentaje de contribución de las variables si las contribuciones fueran uniformes

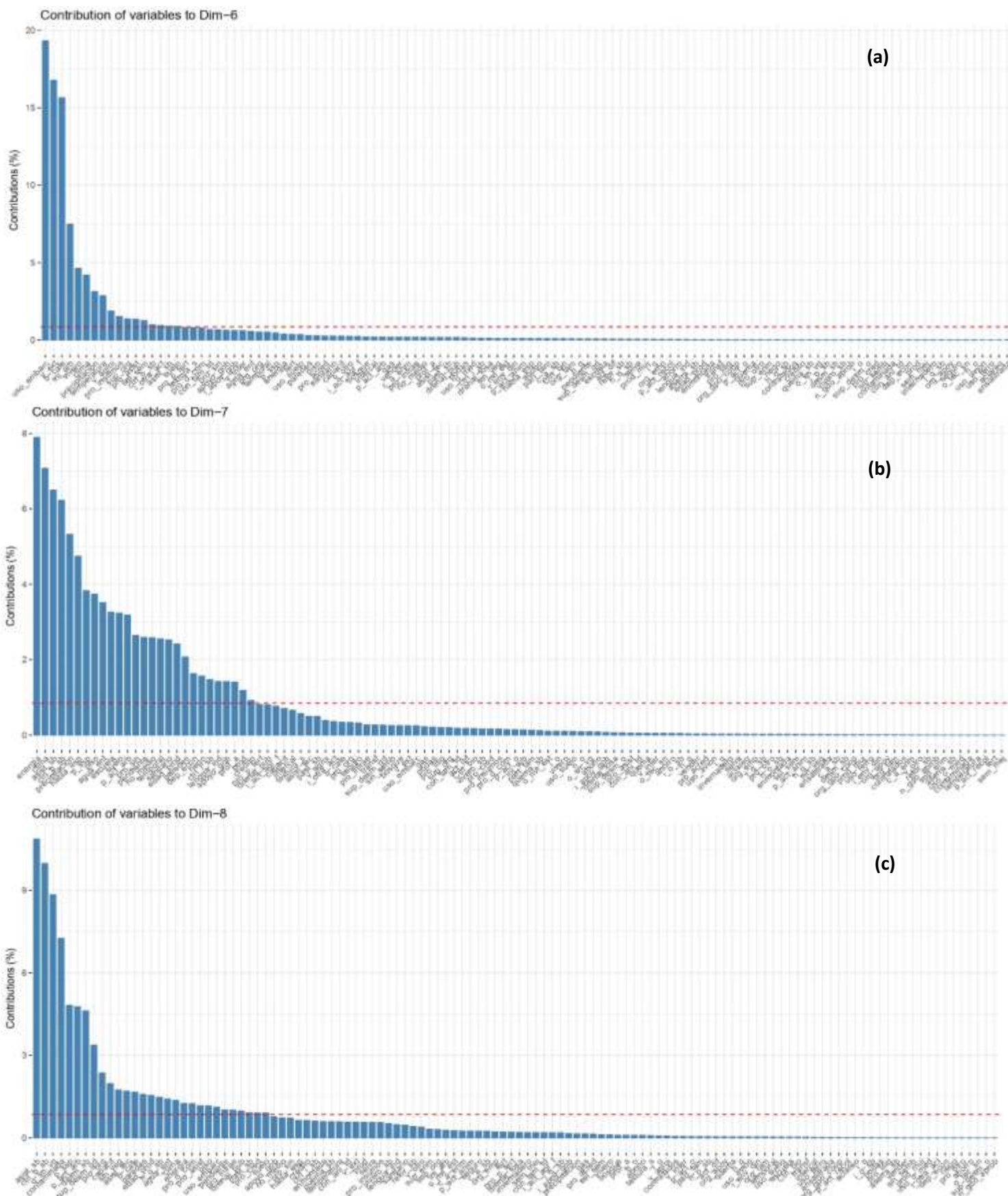
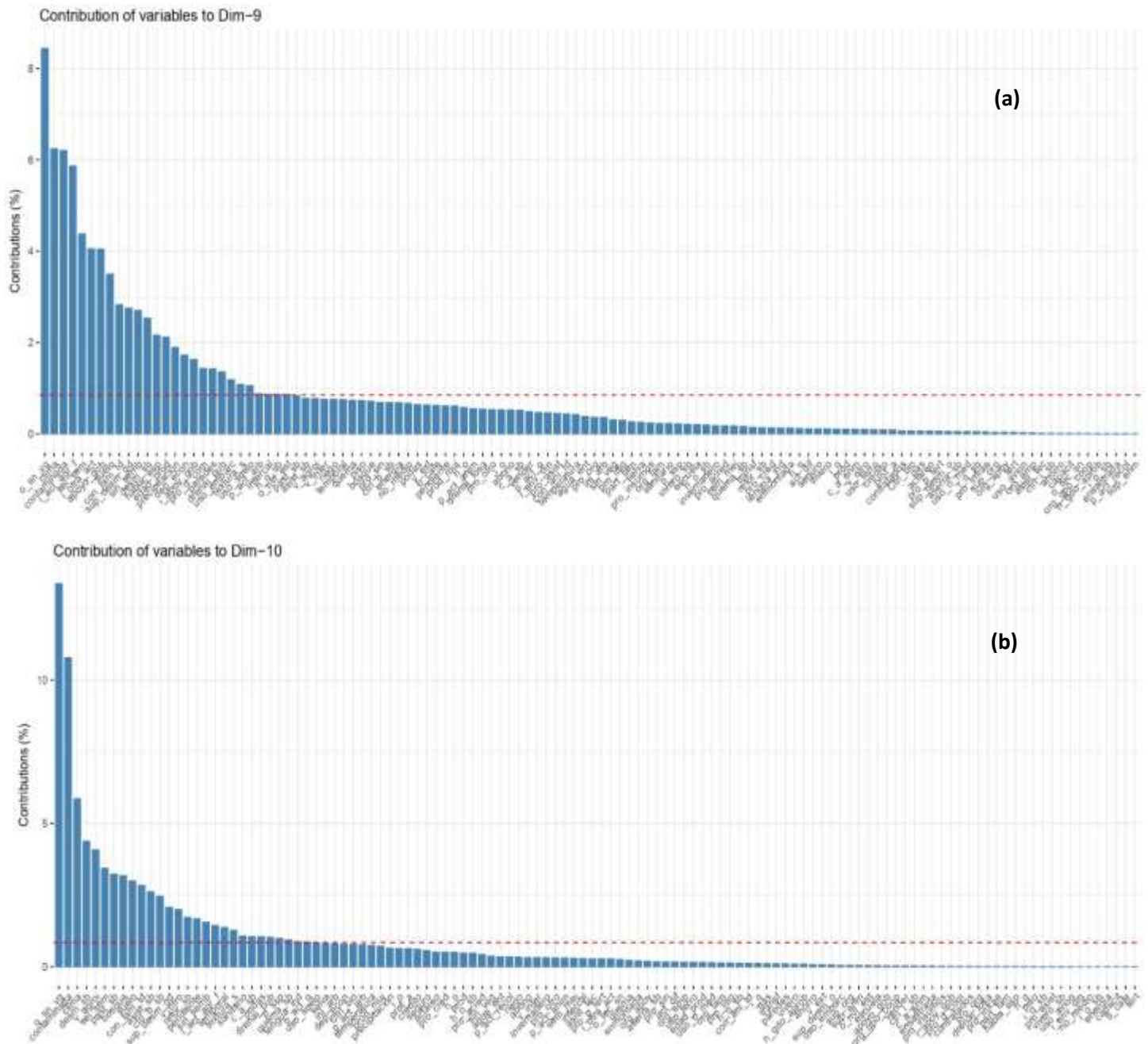


Figura A4. Gráficos del porcentaje de contribución de las variables de las dimensiones 6 (a), 7 (b) y 8(c), generados a partir del FAMD. La línea roja punteada corresponde a la media esperada del porcentaje de contribución de las variables si las contribuciones fueran uniformes



.Figura A5. Gráficos del porcentaje de contribución de las variables de las dimensiones 9 y 10, generadas a partir del FAMD. (a) Porcentaje de contribución de la dimensión 9, y (b) Porcentaje de contribución de la dimensión 10. La línea roja punteada corresponde a la media esperada del porcentaje de contribución de las variables si las contribuciones fueran uniformes.

c. Diagnóstico del modelo

Para evaluar la eficiencia del modelo de árbol de regresión se realizó un gráfico de residuos estandarizados, esto es, se graficaron las observaciones contra los residuos, en dónde a cada valor se le resta la media y se divide por una única desviación estándar del conjunto de residuos (estandarización z) (Menéndez, 2002). Los residuos se calcularon:

$$\text{Productividad observada}_{(n)} - \text{productividad predicha}_{(n)}$$

n= AGEB rural

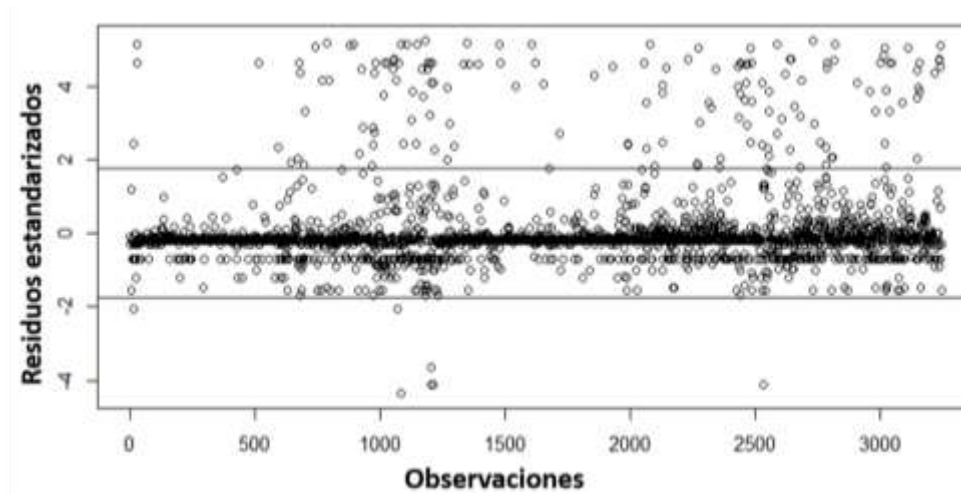


Figura A6. Gráfico de residuos estandarizados. El límite positivo (superior) y el negativo (inferior) corresponden a la desviación estándar. Las observaciones que se salen de los límites corresponden a valores atípicos.