

No. 9

OCTUBRE de 2018

quantil

ISSN 1234-5790 Edición electrónica

**Documentos
de Trabajo**

El Efecto Globo: Identificación de Regiones Propensas a la Producción de Coca

Juan Sebastián Moreno Pabón

Serie Documentos de Trabajo Quantil, 2018-20
Edición electrónica.

OCTUBRE de 2018

Comité editorial:

Álvaro J. Riascos, CoDirector General y Director Modelos Económicos e I&D

Diego Jara Pinzón, CoDirector General y Director Matemáticas Financieras

Andrés F. Galeano Zurbaran, Director Asociado Matemáticas Financieras

Natalia Iregui Parra, Directora Administrativa

Simón Ramírez Amaya, Director de Tecnologías de la Información

Erick M. Translateur Martínez, Investigador

Mateo Dulce Rubio, Investigador

Hamadys L. Benavides Gutiérrez, Investigadora

© 2018, Quantil S.A.S., Modelos Económicos,
Carrera 7 # 77 - 07. Oficina 901, Bogotá, D. C., Colombia
Teléfonos: 3718132 – (310)6791459 – (320)8461236
E-mail: info@quantil.com.co
<http://www.quantil.com.co>

Impreso en Colombia – Printed in Colombia

La serie de Documentos de Trabajo Quantil se circula con propósitos de discusión y divulgación. Los artículos no han sido evaluados por pares ni sujetos a ningún tipo de evaluación formal por parte del equipo de trabajo de Quantil.

Publicado bajo licencia:



Atribución – Compartir igual

Creative Commons: <https://co.creativecommons.org>

El Efecto Globo: Identificación de Regiones Propensas a la Producción de Coca^{*}

Juan Sebastián Moreno Pabón[†]

24 de febrero de 2019

Resumen

La lucha contra las drogas, especialmente contra la cocaína, ha empleado una gran porción de los recursos económicos, ambientales y humanos de Colombia en las últimas cuatro décadas. No obstante, no parece que haya avances significativos en la reducción del tráfico de cocaína ni en la reducción de cultivos de coca. Una de las principales razones por la que las estrategias que atacan la oferta de narcotráfico han fracasado es el llamado Efecto Globo, según el cual represiones en la producción de drogas de una región están asociados a aumentos en otras regiones por efectos de desplazamiento. Con el objetivo de reducir de manera definitiva los cultivos de coca en Colombia durante la implementación de los acuerdos de paz, se busca identificar los municipios susceptibles a las consecuencias del Efecto Globo, es decir, encontrar los municipios que son potencialmente cocaleros a través de un ejercicio de predicción. Este ejercicio permitirá alertar al gobierno para que implemente políticas focalizadas que eviten el desarrollo de producción de hoja de coca. La metodología empírica se centra en técnicas de análisis supervisado de aprendizaje de máquinas (*machine learning*), en particular ensambles de modelos a través de *subbagging*, los cuales permiten el desarrollo de un modelo de predicción agregado que pueda clasificar los municipios potencialmente cocaleros.

Palabras claves: Efecto Globo, cultivos ilícitos, hoja de coca, *machine learning*, *subbagging*. Códigos JEL: B49, K42, P37

^{*}Agradezco de forma especial Hernando Zuleta por asesorar este trabajo, a Paula Jaramillo, Álvaro Riascos y Ricardo Rocha por sus valiosos comentarios y correcciones, a mis colegas del Centro de Estudios sobre Seguridad y Drogas (CESED) de la Universidad de los Andes, Juan F. Godoy y Claudia Rodríguez y mis colegas de Quantil S.A.S. Olga Barrios y Mateo Dulce. También quisiera agradecer a la Facultad de Economía de la Universidad de los Andes por permitir, fomentar y colaborar en el desarrollo de este trabajo.

[†]Investigador Senior, Quantil S.A.S. Correo electrónico: juansebastian.moreno@quantil.com.co

1. Introducción

El Efecto Globo (*Balloon Effect*) es un resultado de las estrategias de la guerra contra las drogas. Este efecto consiste en que la reducción de la producción de drogas en una región se traduce en el desplazamiento de la misma a otra región que puede ser o no vecina, de ahí el nombre ya que se asimila al desplazamiento del aire en un globo al ejercer presión sobre un área. La ocurrencia del Efecto Globo fue registrada por primera vez en la región Andina entre 1993 y 1998 cuando Perú, debido a diversos factores internos y externos, redujo sustancialmente su producción y tráfico de drogas y la producción se trasladó a Colombia y Bolivia, en este período Colombia paso a ser el primer productor de hoja de coca (Antesana y García, 2010). La región Andina representa la mayor parte de la producción de hoja de coca en el mundo. En la Figura 1 se puede evidenciar que el cambio en la producción de hoja de coca en una región se comporta de manera inversa a los cambios en producción de la otra (desplazamiento de producción).

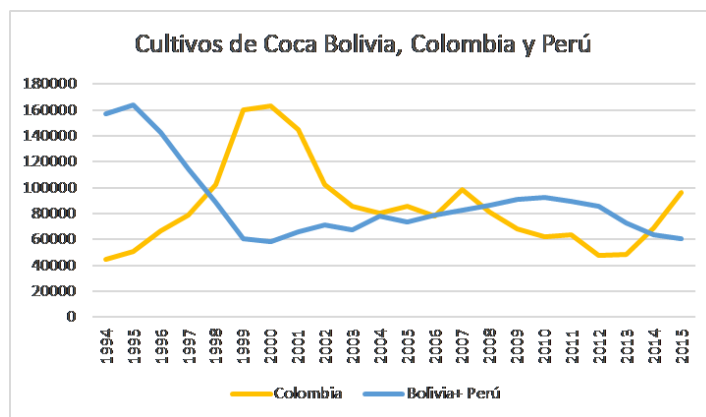


Figura 1: Evolución de hectáreas de cultivos ilegales de coca en la Región Andina. Fuente (UNODC).

Se ha estudiado y estimado estadísticamente la existencia del Efecto Globo. Castro, Raffo y Díaz (2016) encontraron, a través de una metodología SUR generalizada a ecuaciones simultáneas, que efectivamente los cultivos de coca de cualquiera de los tres países en la Región Andina (Colombia, Perú y Bolivia) dependen inversamente de los cultivos de los otros dos países productores. En principio se utilizaba el término Efecto Globo para hacer alusión al desplazamiento de cultivos entre un país y otro, pero estudios (Bagley, 2012) encontraron que la producción de los cultivos ilícitos en los municipios de Colombia y Perú depende de los cultivos de otros municipios en el mismo país (también llamado efecto cucaracha). Moreno-Sanchez, Kraybill y Thompson (2003) utilizan especificaciones econométricas que incluyen los cultivos externos e internos de los países de la región Andina, y encuentran que la producción de cultivos no solo se ve afectada por la oferta agregada de la región Andina, en cambio encuentran que la elasticidad entre la producción de un municipio es negativa respecto a la producción del

resto del país respectivo. Este mismo resultado lo encuentra Vargas (2004) al aplicar modelos de control óptimo para estudiar el problema de los productores, encontrando que erradicaciones forzosas entre 1999 y 2003 en el Putumayo llevaron al desplazamiento de los cultivos a Guaviare y Nariño. Evidencia empírica más reciente indica que el Efecto Globo no tiene que ser necesariamente entre regiones vecinas, se encuentra que éxitos en la reducción de la oferta en Colombia tienen repercusiones sobre la producción de drogas y violencia en México, calculando la oferta de drogas y cantidad de violencia a través del número de carteles (Castillo, Mejía y Restrepo, 2013).

Se han hecho acercamientos desde un punto de vista más teórico, estudios como el de Ortiz (2003) modela el mercado ilegal a través de modelos de equilibrio general, de donde concluye que represiones al narcotráfico, en el largo plazo, amplifican esta actividad al incrementar su rentabilidad por las reducciones en la oferta. La mayoría de los estudios hechos sobre el Efecto Globo concluyen y recomiendan sobre sus hallazgos de manera similar. La existencia (validada estadística y teóricamente) del Efecto Globo lleva a dos grandes conclusiones: primero que las políticas de erradicación del estado colombiano son ineficaces ya que la producción se desplaza entre municipios al ser reprimida (Moreno-Sanchez et al., 2003). Segundo, la intervención hasta el momento de EE.UU. es un esfuerzo en vano, ya que las victorias son parciales, el Efecto Globo se asegura que la oferta agregada de la cocaína en EE.UU. sea constante en el largo plazo (Bagley, 2012 y Granados, 2009).

Estas dos conclusiones llevan a dos recomendaciones claves: para reducir los cultivos es necesario ofrecer otras fuentes de ingresos y oportunidades a los campesinos productores de los cultivos para que estos no reincidan o nuevos ingresen en el mercado ilegal. Finalmente, la existencia del Efecto Globo obliga a que las políticas no sean focalizadas en las áreas donde existen los cultivos ilícitos, en cambio recomendaciones como la anteriormente expuesta deben ser implementadas de manera global, es decir que también se deben aplicar en regiones a donde los cultivos se puedan desplazar, y así lograr un impacto de largo plazo (Castro et al., 2016).

Actualmente, Colombia es, después de cuatro décadas de lucha contra las drogas, el máximo productor de hoja de coca y cocaína a nivel mundial (UNODC, 2016). La lucha contra el narcotráfico ha tenido altos costos ambientales, económicos y sociales sin muchos resultados. Dado que el principal productor de hoja de coca es el grupo de las Fuerzas Armadas Revolucionarias de Colombia (UNODC, 2015), la firma de los acuerdos de paz crea una oportunidad excepcional para reducir de manera significativa y permanente la cantidad de cultivos de coca a lo largo del país. Del anterior contexto nace la pregunta que motiva este trabajo ¿qué impide que se desplacen los cultivos a nuevas regiones en el futuro? Para responder esta pregunta es necesario pensar en las dinámicas del Efecto Globo, es decir, comprender a donde puede desplazar los cultivos el Efecto Globo.

El objetivo de este trabajo es identificar las regiones potencialmente coccaleras con el

fin de prevenir la aparición de nuevos cultivos a través de un modelo de predicción. Se utilizan técnicas de *machine learning* y no econométricas, ya que el objetivo final de este trabajo es predecir y no estudiar inferencias causales. En particular para solucionar problemas de desbalance en la variable a predecir se utiliza *subbagging*, una metodología que consiste en la creación de sub-muestras balanceadas a partir de las cuales se entrenan múltiples predictores, finalmente, teniendo en cuenta cada uno de estos en un sistema de votación simple se crea una predicción final (Hirose y Zaman, 2009).

Los resultados anteriormente descritos permitirán al gobierno prevenir el desarrollo de futuros cultivos de coca en nuevas regiones que son susceptibles a las consecuencias del Efecto Globo, mientras reducen los cultivos ilícitos en las regiones actualmente cocaleras y así lograr impactos de largo plazo en la oferta agregada de coca.

Este trabajo está separado en cinco secciones de las cuales esta introducción es la primera. En la segunda sección, se describe la información que se utiliza y algunos hechos estilizados que caracterizan, en una primera instancia, las diferencias entre los municipios cocaleros y no productores. Después, se expone la metodología utilizada para abordar el problema, la estrategia empírica, sus ventajas y desventajas, la literatura que soporta el uso de las variables escogidas en el modelo y la especificación. En la cuarta sección, se presentan los resultados del modelo y se obtienen los municipios potencialmente cocaleros. Por último, está la sección de conclusiones y recomendaciones.

2. Datos

2.1. Estadísticas descriptivas

En esta sección se muestran estadísticas descriptivas con el fin de estudiar la evolución de los cultivos de coca en Colombia y caracterizar y diferenciar los municipios cocaleros de los no productores de cultivos ilícitos. El anterior ejercicio tiene como objetivo estudiar como se ve un municipio que tiene la actividad ilícita en cuestión y como se ve uno que no, esto permitirá ubicar a los municipios que sean clasificados por el modelo como potencialmente cocaleros.

2.1.1. Panorama de Colombia

Se comienza con una primera mirada a la evolución de cultivos en los últimos años. La Figura 1 se ilustra la evolución de los cultivos de coca en Colombia. Actualmente, la producción se encuentra en lo más alto desde el 2007, hubo un crecimiento de los cultivos ilícitos entre el 2013 y 2015 del 100 %. Después del 2007 la producción de hoja de coca se reduce consistentemente hasta el 2012, cuando coincide con el inicio de los acuerdos de paz con las FARC. Una hipótesis plantea que los beneficios ofrecidos en los

acuerdos de paz a aquellos que sustituyan y ayuden a erradicar voluntariamente cultivos de coca incentivan a los campesinos a producir con el fin de verse favorecidos cuando se implementen los acuerdos (UNODC, 2016).

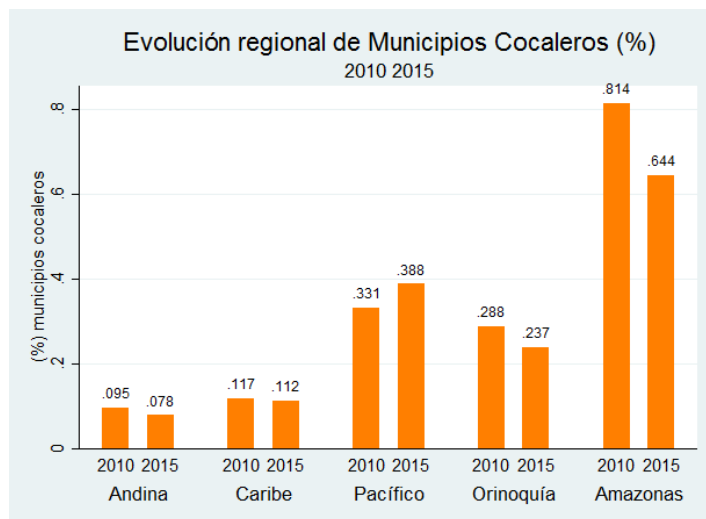


Figura 2: Proporción de municipios cocaleros (2010 y 2015).

En la Figura 2 se evidencia que la proporción de municipios cocaleros se ha reducido en cada región exceptuando el pacífico. En adición, en la Figura 3, se muestra que las hectáreas de cultivos han aumentado de manera significativa para el Pacífico y la Amazonia, se concluye que la producción de hoja de coca se ha densificado ya que la producción agregada ha crecido un 50% del 2010 al 2015 (ver Figura 1) pero la cantidad de municipios se ha reducido en todas las regiones, exceptuando el pacífico donde aumentaron en un 5%.

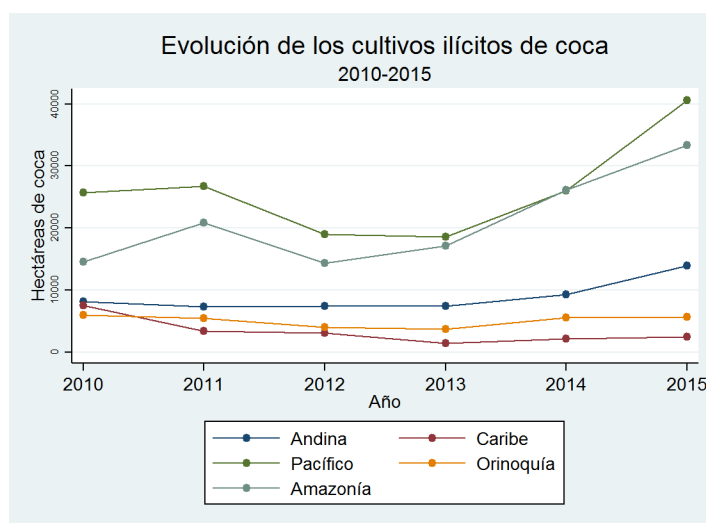


Figura 3: Evolución de los cultivos ilícitos de coca de 2010-2015

2.1.2. Caracterización de municipios cocaleros

El panel de datos mencionado en la sección 2.1 permite dar una primera mirada a sus disimilitudes. Se empieza por estudiar las diferencias en el desarrollo entre los municipios cocaleros y no productores. Se procede a estudiar el NBI (necesidades básicas insatisfechas), las distancias a centros urbanos y mercados, el índice de ruralidad y los ingresos totales e ingresos totales per cápita del municipio. Los datos muestran que el promedio nacional de NBI¹ de los municipios cocaleros es casi un 20 % más alto que el de los no productores (vea la Figura A en el apéndice). Se realizó el mismo ejercicio a nivel regional, debido a que las distintas variables socioeconómicas cambian de manera significativa entre regiones. Se encontró que los municipios cocaleros tienen para todos los casos excepto uno (región Amazónica) un NBI más alto (vea Figura B).

La literatura ha encontrado que uno de los principales determinantes de los cultivos de coca es la marginalización de una región (Rocha 2010), a continuación se pueden observar las distancias² entre los municipios y la capital del departamento respectivo, y el mercado mayorista de alimentos más cercano.

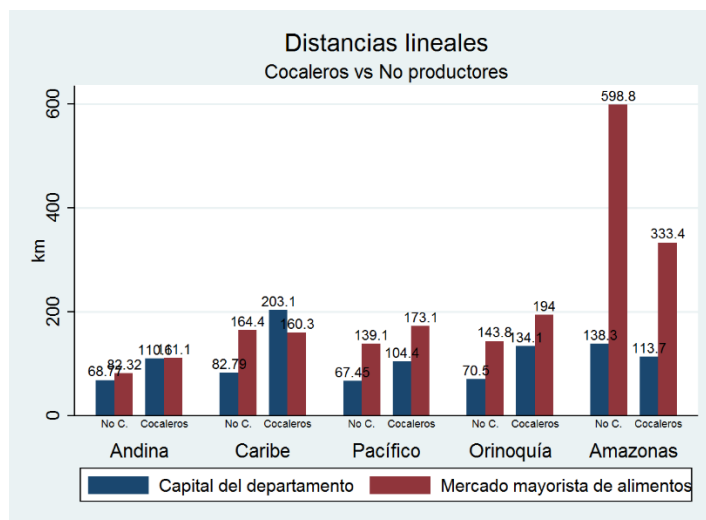


Figura 4: Distancias Lineales (Regional)

De la Figura anterior se puede concluir que los municipios cocaleros están más lejos, en promedio, de las capitales de sus respectivos departamentos que los municipios no productores. Para la distancia al mercado mayorista de alimentos más cercano se observa el mismo patrón en las regiones Andina, Pacífica y Orinoquía. De nuevo, la región Amazónica muestra un comportamiento atípico respecto al resto del país.

Se estudian los índices de ruralidad promedio (proporción de la población que vive

¹La medición del NBI va a ser el porcentaje de personas que incumplen con al menos una de las cinco dimensiones del indicador.

²Son distancias lineales tomadas desde los centroides de los municipios.

en áreas rurales) entre municipios cocaleros y no productores. Se encuentra que, en promedio, los municipios cocaleros son más rurales (vea Figura C). Para la región Andina no hay evidencia estadística suficiente para afirmar que el promedio de ruralidad es distinto y el Amazonas muestra que los municipios cocaleros son menos rurales.

Asimismo, se estudian los ingresos municipales totales para estudiar las diferencias económicas, además de una aproximación a la calidad de las instituciones económicas estatales, entre los cocaleros y los no productores. Como es esperado, se ven menores ingresos municipales en las regiones cocaleras (vea la Figura D en el apéndice). Para hacer un estudio que permita mayor comparación entre los municipios se analiza la variable per cápita (vea el apéndice, Figura E). Se halla que, en el Pacífico, la región que posee más hectáreas de coca (como se puede observar en la Figura 3), los municipios productores de hoja de coca tienen mejores niveles de ingresos per cápita, contrario al resto del país.

En la siguiente tabla se pueden observar los promedios y diferencias estadísticas de las medias de los determinantes socioeconómicos, naturales y de conflicto de cultivos entre los municipios cocaleros y no cocaleros.

Tabla 1: Diferencia de medias, Cocaleros y no productores de coca

	No .C	Cocalero	Diferencia de Medias
Índice de ruralidad	0.554	0.635	0.0808***
	[0.247]	[0.223]	[9.66]
Distancia a Capital del Departamento (km)	72.35	121.9	49.56***
	[51.03]	[79.89]	[25.04]
Distancia Mercados Mayoristas (km)	116.1	191.4	75.28***
	[104.2]	[121.9]	[20.29]
Distancia otros Mercados (km)	54.82	130.6	75.77***
	[89.51]	[129.8]	[22.39]
Ingresos per cápita reales (millones COP)	0.996	0.923	-0.0730**
	[0.757]	[0.992]	[-2.57]
Área (km ²)	522.6	3216.1	2693.5***
	[1293.4]	[6511.8]	[25.83]
Altura (msnm)	1254.1	635.7	-618.4***
	[1201.2]	[735.0]	[-15.87]
Precipitaciones (mm)	1847.3	3253.3	1405.9***
	[909.9]	[1670.4]	[37.41]
Presencia de grupos armados al margen de la ley	0.0369	0.146	0.109***
	[0.188]	[0.354]	[12.46]

Errores estándar entre paréntesis

*** p<0.01, ** p<0.05, * p<0.1

En la tabla anterior es importante resaltar que hay una mayor presencia de grupos armados al margen de la ley (FARC, AUC, ELN) en los municipios donde hay cultivos

ilícitos de coca que en aquellos que no, alrededor de 11 %. Esto se alinea con la literatura, la cual atribuye la mayoría de la producción en Colombia a estos grupos (UNODC 2015).

En la tabla 1 se observa que para todos los factores hay evidencia estadística a cualquier nivel de significancia (excepto ingresos municipales per cápita, los cuales son significativos al 5 %) para concluir que la diferencia de medias entre los municipios cocaleros y no productores de coca es distinta de cero. Lo anterior confirma a través de evidencia empírica y robusta que los municipios productores de hoja de coca se diferencian de manera significativa de aquellos que no producen.

2.2. Base de datos

Con el objetivo de encontrar regiones potencialmente cocaleras se utiliza un panel de datos a nivel municipal para los años 2010 a 2014, años para los cuales hay información disponible para todas las variables consideradas en este trabajo (en la sección 3.2 son explicadas). Los datos de hectáreas de cultivos ilícitos de coca provienen de la Oficina de Naciones Unidas contra la Droga y el Delito (UNODC por sus siglas en inglés), información que proviene del Sistema integrado de monitoreo de cultivos ilícitos (SIMCI). Para construir la base de datos con las características sociales, económicas, geoFiguras y de conflicto de cada región se utilizan las fuentes estadísticas que provee el panel municipal CEDE. Datos sobre erradicación de cultivos e incautaciones de coca, pasta y cocaína provienen del Ministerio de Justicia. Finalmente, los datos climáticos provienen del IDEAM.

3. Metodología

3.1. Abordando el problema

Esta investigación se aproxima al problema planteado a través de un análisis supervisado³ con el fin de hallar patrones y pronósticos sobre las regiones susceptibles a las consecuencias del Efecto Globo. La variable dependiente de este modelo es una variable dicótoma que toma el valor de 1 si el municipio hizo la transición a convertirse en cocalero, 0 en caso contrario, esto significa que es un problema de clasificación. De manera general, la metodología mencionada utiliza algoritmos que se encargan de enseñarle a la “máquina” como se caracteriza una observación que toma el valor de 1 y una observación con valor de 0 para después aplicar lo aprendido por el modelo creado a nuevas observaciones y poder clasificarlas (predecirlas) de manera correcta.

³Análisis supervisado: predicen un dato (o un conjunto de ellos) desconocido a priori, a partir de otros conocidos.

Los problemas de clasificación en aprendizaje de máquinas requieren partir la base de datos en dos grupos, una parte de entrenar el modelo, que llamaremos la base de entrenamiento (*training set*), y otra parte para medir el desempeño del modelo (*test set*), se va a referir a esta parte como base de prueba.

Un municipio potencialmente cocalero se define en este trabajo como aquel que no produce cultivos ilícitos de coca pero que tiene la combinación de características y factores que permiten la proliferación de los mismos. Para encontrar estos municipios se aglomeró el panel de datos a un corte transversal, es decir que las observaciones se convirtieron a nivel de municipio y no de año-municipio. Para lograr lo anterior se promediaron los valores de las variables entre el 2010 y el 2013 por municipio (se utilizan los datos del 2014 como base de prueba al ser la información más actualizada y completa posible). Posteriormente, se marcaron con un uno a los que cambiaron de estado, es decir los que empezaron a producir de un año a otro. Ya que el objetivo es encontrar los municipios potencialmente cocaleros se eliminan de la base los municipios que fueron productores de coca entre el 2010 y 2013, en particular esto reduce el ruido en los modelos de aprendizaje ya que los municipios potencialmente cocaleros solo pueden ser elegidos del conjunto de municipios que no producen coca. El proceso anterior reduce 185 municipios, los cuales produjeron constantemente durante los 4 años, de un total de 1,122. De esta nueva base de 937 municipios, 69 son los que cambiaron a ser productores de coca en algún momento del periodo de estudio. Esta base es usada para entrenar el modelo de clasificación.

La creación de la anterior base nos permite buscar cuales son los municipios que se convertirán en productores de hoja de coca. Finalmente, cuando se prediga sobre una nueva base de datos (base de prueba) aquellos que el modelo pronostique como 1 serán los municipios que el modelo clasifica como futuros productores de coca.

Esta misma base también tiene un problema que genera sesgo en los clasificadores, la variable dependiente está fuertemente desbalanceada, esto significa que hay muy pocos positivos con relación a los negativos o viceversa. Para este caso de estudio hay muy pocos positivos, es decir que el número de observaciones que pasaron por la transición de transformarse en productores de coca en el lapso del estudio son muy pocas en relación con el número de municipios que no hicieron la transición. Lo anterior hace que los modelos se sesguen a pronosticar dentro del grupo mayormente representado. Para corregir este problema se utiliza *subbagging* (Antonini, Elisseef y Paleologo, 2010), metodología que será explicada en la subsección 3.3.

El problema de clasificación se aborda a través de modelos de regresión y aprendizaje basado en árboles de decisión. Los modelos de regresión pueden ser no lineales (*Logit*, *Probit*) o a través de un modelo de probabilidad lineal (MPL). Se desecha la posibilidad de usar un MPL debido a posibles problemas de heterocedasticidad y clasificaciones por fuera del intervalo de cero a uno. Se va a utilizar un *Logit* ya que tiene un poder predictivo más fuerte para las observaciones que tienen valores atípicos al tener una

mayor área bajo las colas de su distribución, relativa al *Probit*. Adicionalmente, los modelos *Logit* tienen una ventaja computacional sobre los *Probit* al estar definidos sobre un espacio cerrado, lo que asegura que el modelo siempre converja y en menor tiempo. Las regresiones logísticas presentan únicamente una desventaja frente al *Probit* para modelos binomiales, este último supone que los errores se distribuyen como variables normales. De igual forma las estimaciones de ambos modelos son muy similares luego no es preocupante este último factor.

Los árboles de decisión son árboles que predicen (*Classification Trees* o CTs por sus siglas en inglés) a través de particiones de los posibles valores que pueden tomar las variables predictoras. Cada nodo del árbol representa la variable a través de la cual se hace una partición y cada rama representa el valor que el nodo asume para la partición. Las observaciones son clasificadas desde el primer nodo (raíz) y recorre las ramas del árbol basado en los valores que tomen sus predictores (Kotsiantis, 2007).

3.2. Especificación del modelo y descripción de variables

A continuación, se presenta la especificación del modelo:

$$Pr(y_i = 1 | x_i) = G(Y_i) \quad (1)$$

Donde

$$Y_i = \beta_0 + Desarrollo_i \alpha_j + Naturales_i \gamma_m + Conflicto_i \delta_n + Regiones_i \rho_d \quad (2)$$

Donde $G(\cdot)$ es una función predictiva, ya sea un *Logit* o árbol de clasificación. $Desarrollo_i$, $Naturales_i$, y $Conflicto_i$ representan los grupos de variables que son explicados más adelante. $Regiones_i$ son los efectos fijos por regiones (a través de *dummies*). Esta investigación no tiene variables de interés porque no se estudia causalidad ya que esta ha sido demostrada en la literatura predecesora a este trabajo, en cambio el objetivo de la especificación es la predicción.

3.3. Predictores

Las variables escogidas representan los determinantes de los cultivos ilícitos de coca que la literatura ha hallado hasta el momento. Los regresores de este modelo estarán divididos en cuatro grupos:

- Desarrollo: Se ha evidenciado que entre los determinantes de los cultivos de coca se encuentra el aislamiento geográfico, la marginación del desarrollo, la ruralidad de la región y la pobreza (Cárdenas, 2011). Los incentivos de los agricultores se

ven orientados hacia los cultivos ilícitos debido a las distancias a los mercados y centros de comercios más cercanos y las presiones ejercidas por grupos armados. Por esto se escogen las variables independientes socioeconómicas que caracterizan a los municipios tales como los ingresos municipales (en niveles y per cápita), las distancias a la capital del departamento y mercados mayoristas de alimentos y el índice de ruralidad.

- Naturales y climáticos: en este grupo están las variables que representan los factores necesarios para la producción de la hoja de coca como la temperatura, precipitaciones, altura y el tamaño del municipio. (Rocha, 2010 y Santoni y Torres, 2007).
- Conflicto y violencia: Se ha estudiado que las regiones donde se encuentran los cultivos hay una mayor presencia de grupos armados, combates y en general mayores niveles de violencia hay una mayor probabilidad de encontrar cultivos ya que son vigilados y son pretendidos por los mismos grupos (Castillo, Mejía y Restrepo, 2013). Por esto se escoge la variable que capturen la presencia del conflicto como la presencia de grupos armados. Adicionalmente, en este grupo están variables como la incautación de cocaína, hoja de coca y base de coca con el fin de capturar los efectos de las estrategias contranarcóticas actuales sobre la producción. Se ha hallado que los métodos de erradicación como la aspersión aérea no son eficaces, a diferencia de las incautaciones en laboratorios y otros eslabones de la cadena de producción, ya que al reducir en una porción los cultivos hacen que los precios de la cocaína suban y por ende se creen incentivos para producir más cultivos de coca, pero en nuevas regiones con el fin de evitar las aspersiones (Sandoval, López y Cárdenas, 2009 y Mejía, Restrepo y Roza, 2012).

Note que pueden existir varios problemas de endogeneidad entre las variables anteriormente expuestas y la variable dicotómica que representa si el municipio se transformó en cocalero, pero este no es un problema, ya que la metodología utilizada no es para obtener los mejores estimadores posibles, en cambio es para obtener la mejor clasificación. Se da un *trade-off* entre el sesgo y la consistencia de los estimadores y la precisión del modelo. Aunque no es posible estudiar los estimadores resultantes del modelo, si es posible estudiar las correlaciones entre las variables, además ya han sido demostradas las relaciones causales entre las mismas y los cultivos de coca y es la razón por la cual son incluidas en el modelo.

3.4. *SubBagging*

Como se mencionó anteriormente, un obstáculo importante de este ejercicio de clasificación es que la variable a predecir está desbalanceada, esto significa que hay muy pocos positivos con relación a los negativos o viceversa. Para este caso de estudio hay muy pocos positivos, es decir que el número de observaciones que pasaron por la transición de transformarse en productores de coca en el lapso del estudio (2010-2013) es mucho

menor que el número de municipios que no cambiaron de estado.

Para corregir este problema se utiliza el método *SubBagging* el cual tiene como objetivo la reducción de la varianza del modelo y mejorar la precisión en la predicción. *SubBagging* es una arquitectura de ensamble de modelos, que consiste en un sistema de votación sobre las estimaciones generadas a partir de los modelos agrupados. Cada modelo de aprendizaje se obtiene entrenando el mismo algoritmo de aprendizaje, por ejemplo una regresión logística, sobre una base de entrenamiento distinta que es creada a partir de una muestra aleatoria de la base de entrenamiento completa, posteriormente se agregan las predicciones de cada uno de los modelos. La predicción final del ensamble será la clasificación que la mayoría de los modelos entrenados prediga (Hirose y Zaman, 2009).

Al ser el problema la poca cantidad de observaciones que cumplen con el criterio de haberse transformado en cocaleras entonces todas las muestras creadas van a tener estas observaciones. Después se hace un muestreo aleatorio de las observaciones que no se transformaron en productores de hoja de coca, de tal forma que cada muestra está compuesta por las 69 observaciones positivas y N observaciones elegidas aleatoriamente negativas. Entonces se debe estudiar cual es el tamaño de N que mejores predicciones logra.

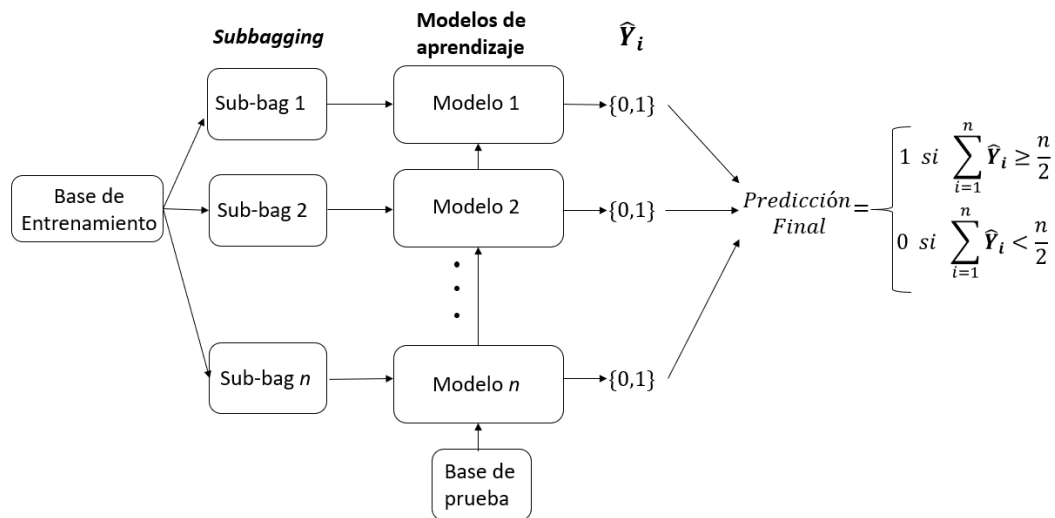


Figura 5: Diagrama de ensamble de modelos a través de subbagging. Fuente: Elaborado por el Autor

El diagrama en la Figura 5 muestra el proceso metodológico. En resumen, se puede representar en cuatro pasos el algoritmo para crear el ensamble:

1. Se construyen n submuestras, todas de tamaño N. Cada una de estas contiene a todas las observaciones que son positivas y una selección aleatoria de las observaciones negativas. El tamaño de la selección aleatoria determinara el tamaño total de cada submuestra y el desbalance de las mismas (este parámetro se escoge a

- través de métricas que comparen y permitan elegir el mejor ensamble).
2. Con cada submuestra se entrena el mismo modelo de predicción (un ensamble homogéneo), en este caso de estudio se usan *Logit* y árboles de decisión. Lo anterior implica que si existen n submuestras entonces el ensamble va estar compuesto por n modelos de predicción.
 3. Se procede a introducir una o las observaciones que se desean predecir en cada uno de los n modelos ya entrenados, cada uno de estos van a dar su veredicto sobre si una observación es 1 o 0.
 4. Con estas n predicciones se hace un sistema de votación simple para dar el veredicto del ensamble. Si más de la mitad de los ensambles consideran que la observación debe ser un positivo, entonces el ensamble concluirá lo mismo.

4. Resultados

4.1. Desempeño de los modelos

Para comparar y encontrar el mejor modelo se utiliza el área bajo la curva (AUC por sus siglas en inglés) de la curva Característica Operativa del Receptor (ROC por sus siglas en inglés). La curva ROC indica cómo cambia la sensibilidad del modelo, es decir la razón de verdaderos positivos, frente a la especificidad, la razón de verdaderos negativos, cuando varía el umbral de discriminación. El umbral de discriminación es el valor a partir del cual se decide que una predicción es un positivo o negativo. Entonces el desempeño general del clasificador, resumido para todos los umbrales de discriminación posibles, va a estar representado por el AUC de la curva ROC. Un AUC es mejor mientras más cercano esté a 1, es peor mientras más cercano esté a 0.5 (James et. Al, 2015). Cuando el AUC es 0.5, intuitivamente, indica que las predicciones correctas del modelo son aleatorias, mientras que un AUC más cercano a 1 muestra que las predicciones correctas son atribuidas a la calidad del modelo.

Evaluando el desempeño de los modelos se encontró que los resultados eran muy volátiles, es decir que el AUC respecto al desbalance cambiaba de manera significativa cada vez que se ejecutaban los resultados. Esto se debe a que las muestras son muy pequeñas, lo cual hace que cada vez que se hace el muestreo aleatorio los modelos *Logit* y CTs se enfrenten a bases de entrenamiento que difieren de gran manera. Ante esto se decidió testear el modelo sobre la base de datos del 2014 treinta veces, promediar sus AUC, aquel que tuviera el mejor promedio de AUC sería el modelo que se ejecutaría sobre la base del 2014 y se concluye a partir de esta, al ser la información más actualizada, cuáles son los municipios potencialmente cocaleros.

A continuación, se muestran los distintos desempeños obtenidos por el *subbagging* de *Logits* y CTs cambiando el tamaño de las muestras. El número de muestras está fijo

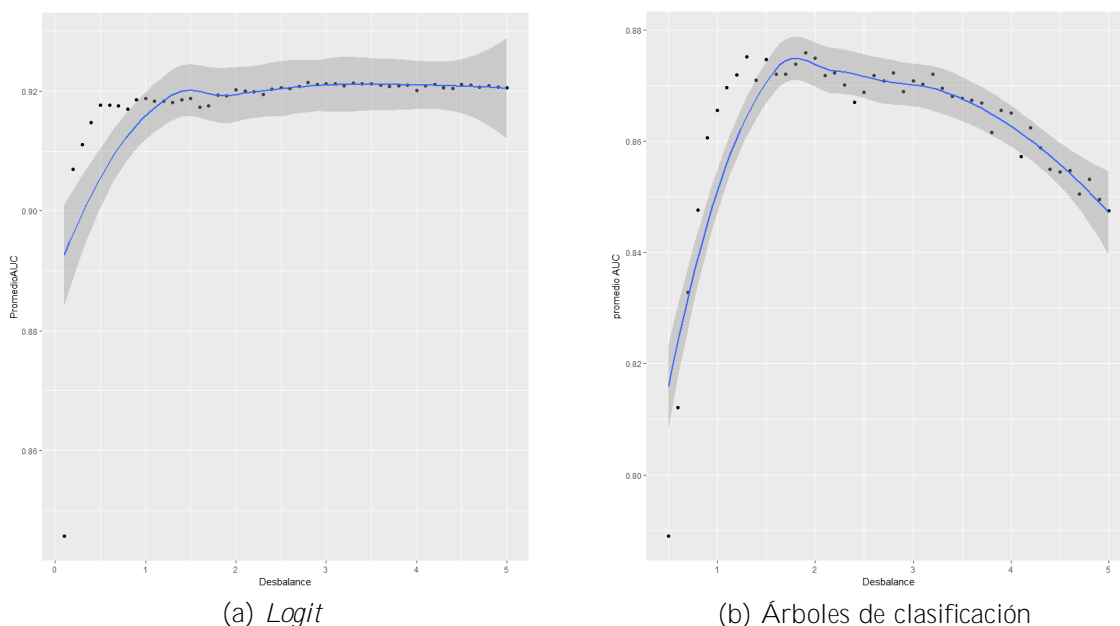


Figura 6: AUCROC vs Proporción de desbalance

en 150, esto significa que cada modelo es el resultado de la votación de 150 modelos de predicción. Se observa que el comportamiento del ensamble de *Logits* se estanca alrededor de un AUC después de cierto desbalance, mientras que el ensamble de árboles tiene un máximo claro.

En la Figura 6, el eje vertical se observa el AUC promedio, mientras que en el eje horizontal se muestra el número de veces que representan los negativos en la base de entrenamiento respecto a los positivos. El valor de proporción de desbalance 1 significa que en cada muestra de entrenamiento hay tantos municipios que no hicieron la transición a coccaleros como los que si la hicieron, es decir 69, para un total de 138 observaciones para entrenar el modelo en cada muestra. En el apéndice de tablas (A y B) se pueden ver los valores del promedio de AUC para cada nivel de desbalance.

Se encontró que, en promedio, el mejor valor de AUC para el ensamble de *Logits* se presenta cuando la proporción de desbalance es 2.8, esto significa que en cada muestra de entrenamiento hay 2.8 veces el número de municipios que no hicieron la transición a coccaleros como los que si la hicieron. Lo anterior implica que cada una de las 150 bases de entrenamiento de los modelos tiene 193 municipios que no hicieron la transición y los 69 que si la hicieron, para un total de 262 observaciones por base de entrenamiento.

Para el caso del ensamble de árboles de clasificación se encontró que, en promedio, el mejor valor de AUC se presenta cuando la proporción de desbalance es de 1.9, esto significa que cada muestra es de 200 observaciones.

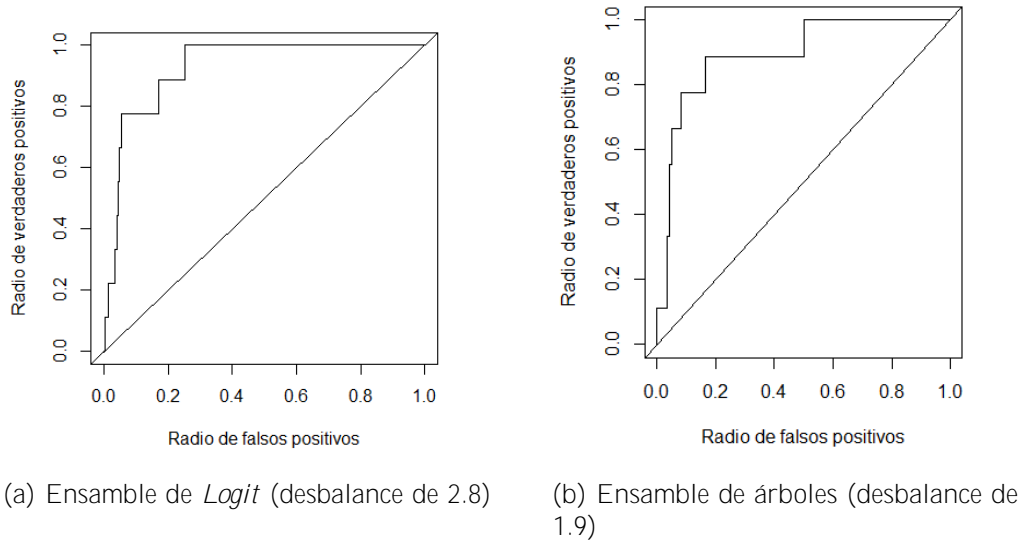


Figura 7: Curvas ROC. 150 muestras

A continuación, tomando el parámetro de desbalance óptimo para el ensemble de *Logit* y CTs (2.8 y 1.9 respectivamente) se construyen las curvas ROC.

La AUC del ensemble de *Logit* fue de 0.9240 y su precisión (porcentaje de observaciones correctamente clasificados) de 96.06 %. Mientras que el AUC del ensemble de árboles de clasificación fue de 0.8890 y su precisión fue de 95.31 %. Lo anterior implica que el mejor clasificador fue el ensemble de *Logit* por la métrica de AUC de la curva ROC. Note que el poder de predicción del ensemble *Logit* también fue superior⁴.

4.2. Estimadores del ensemble de *Logit*

Ya elegido el mejor modelo, se analizan los coeficientes y resultados del ensemble de *Logit* dados por el parámetro de desbalance 2.8 y 150 bases de entrenamiento. Este ensemble está compuesto de modelos paramétricos, entonces procedemos a estudiar las características de estos parámetros.

Al ser un ensemble de 150 *Logit* se tiene que para cada predictor existen 150 coeficientes. Se muestra a continuación cual fue la frecuencia de la significancia, a través del p-valor de los coeficientes.

⁴En principio se estimó el modelo con dos variables más, la aptitud del suelo para cultivar y la variable dicotoma que representa si el municipio tenía un vecino cocalero. Se descartaron ya que el mejor desempeño del modelo fue peor respecto al presentado, el ensemble de *Logit* logró un AUCROC de 0.87 para una proporción de desbalance de 1,9.

Tabla 2: Frecuencia de rangos por P-valores de los coeficientes

Variables	Rangos de p valores	
	P-valor <0.05	P-valor >0.05
Intercepto	94	56
Área	11	139
Altura	37	113
Altura2	101	49
Precipitaciones	150	0
Índice de ruralidad	117	33
Distancia a la capital	149	1
Distancia al mercado mayorista	146	4
Distancia a otros mercados	130	20
Ingresos reales per cápita	0	150
Incautaciones de Cocaína	0	150
Incautaciones de Base de Coca	20	130
Incautaciones de Hoja de Coca	77	73
Presencia de grupos armados	137	13
Andina	0	150
Caribe	1	149
Orinoquía	8	142
Pacífico	0	150

La tabla anterior muestra el número de modelos para el que la variable clasificó en cierto rango de p-valor. En el apéndice de tablas es posible encontrar de manera más desagregada los rangos de p-valores para cada variable. Es necesario tener en cuenta que estas significancias están a nivel local, es decir, que corresponden a cada una de las submuestras creadas en el ensamble. En este trabajo se define la importancia de las variables en el ensamble por la constancia que la significancia tenga a lo largo del ensamble.

Se encuentra que las variables más importantes en el ensamble de regresiones logísticas fueron las precipitaciones, la distancia a la capital del departamento, el índice de ruralidad, la presencia de grupos armados y la distancia al mercado mayorista más cercano. Se resalta el caso de los ingresos totales per cápita reales, esta variable no fue significativa al 5% para ninguno de los 150 *Logit*.

Una vez encontradas las variables más importantes del modelo se procede a estudiar la media y desviación estándar de los coeficientes y efectos marginales. En el apéndice de Figuras, la Figura F contiene las distribuciones de los mismos.

Tabla 3: Coeficientes y Efectos Marginales promedio

VARIABLES	Coeficientes Promedio	Efectos marginales Promedio
Intercepto	-4.76 (3.35)	
Precipitaciones	0.00077 (0.00014)	0.000079 (0.000014)
Índice de ruralidad	2.53 (0.59)	0.26 (0.065)
Distancia de la capital del departamento	0.014 (0.0028)	0.0015 (0.00027)
Distancia al mercado mayorista	0.012 (0.003)	0.0013 (0.00027)
Presencia de grupos armados	1.26 (0.26)	0.13 (0.028)
Observaciones por muestra	268	
Número de muestras	150	

De la tabla 3 se observa que los estimadores y efectos marginales de las variables más significativas del modelo son consistentes con la literatura. Se reitera que la presencia de grupos armados y la marginalidad de los municipios (las distancias a centros económicos y urbanos) hacen más probable que un municipio se convierta en productor de hoja coca. Además, un municipio que sea más húmedo (precipitaciones), condición indispensable para el crecimiento de la hoja de coca, y más rural, característica que permite al productor cultivar y de manera discreta, incrementa la probabilidad de que se empiece a producir.

4.3. Predicciones del modelo

Tras encontrar el mejor modelo de predicción posible y probarlo sobre la base de datos del 2014 se encontró que el modelo predijo que 36 municipios harían la transición de no productores a cocaleros. De estos 36 municipios 14 pertenecen a la región Pacífica, en particular Cauca (8) y Chocó (4). Los departamentos de Antioquia y Amazonas tuvieron una gran cantidad de municipios predichos como potencialmente cocaleros, 7 y 6 respectivamente. En las tablas E,F y G del apéndice es posible observar los municipios clasificados como potencialmente cocaleros y sus distribuciones a lo largo de las regiones y departamentos.

Tabla 4: Estadísticas descriptivas, cocaleros, no productores y potencialmente cocaleros (2014)

	(1)	(2)	(3)	(3)-(2)
	No .C	Cocalero	Potencialmente cocaleros	Diferencia de medias entre potencialmente y cocaleros
Índice de ruralidad	0.54 [0.246]	0.633 [0.223]	0.773 [0.229]	0.14*** [0.0006]
Distancia a la capital del departamento	69.83 [47.07]	121.3 [80.15]	140.2 [100.4]	18.84 [0.21]
Área	434.5 [966.9]	3063.2 [6486.6]	3713 [5061.4]	649.84 [0.57]
Altura	1272.3 [1211.2]	671.1 [744.7]	569.9 [697.8]	-101.2 [0.45]
Precipitaciones	1820.4 [871.2]	3148.1 [1658.1]	3189.4 [1776.2]	41.28 [0.89]
Ingresos municipales per cápita reales	1.299 [0.904]	1.145 [1.291]	1.175 [1.202]	0.03 [0.89]

***p<0.01. En los paréntesis de la última columna se encuentran los p-valores de la diferencia de medias

En la tabla anterior se contemplan las medias y diferencias de medias de los municipios no productores, cocaleros y potencialmente cocaleros. Se observa que los municipios potencialmente cocaleros se asimilan más a los productores de hoja de coca que a aquellos que no. Los potencialmente cocaleros parecen ser más grandes, están a una altura más baja y se encuentran a una mayor distancia de la capital departamental respectiva que los municipios cocaleros, pero solo existe una diferencia significativa, la variable de índice de ruralidad del municipio, los potenciales son más rurales que los productores de hoja de coca.

De los municipios 36 municipios, 21 de ellos tuvieron coca en algún momento desde el 2006. Además, de estos 36, 7 de ellos se convirtieron en productores en el 2015 (en este año 9 municipios hicieron la transformación).

5. Conclusiones y recomendaciones

El efecto globo ha sido una de las limitaciones principales para las políticas de reducción, erradicación y sustitución de cultivos ilícitos. En efecto, en la medida en que no se quiebren los incentivos de los individuos para cultivar coca, los esfuerzos de erradicación o sustitución en una zona hacen que los cultivadores busquen otra zona para sembrar, cultivar y cosechar coca. Este hecho tiene dos implicaciones: en primer lugar, cualquier programa para reducir el área cultivada debe incluir la creación de oportunidades de trabajo y creación de negocios lícitos lo cual está estrechamente ligado a la conectividad (marginalidad) y, en segundo lugar, es necesario identificar municipios vulnerables al cultivo de coca con el fin de adelantar políticas preventivas al desplazamiento de la producción.

En este orden de ideas, la identificación de regiones potencialmente cocaleras es esencial para el diseño e implementación de políticas de reducción de cultivos que sean efectivas y tengan efectos duraderos.

La metodología propuesta en este trabajo permite predecir que municipios tienen la combinación de características socioeconómicas, naturales y de conflicto necesarias para convertirse en productores de hoja de coca. Los municipios identificados como potencialmente cocaleros, es decir, los que el ensamble de modelos predice como productores de coca, tienen características similares a los municipios cocaleros. Lo anterior implica que los municipios potencialmente cocaleros son aquellos que se asemejan a los municipios productores, pero aún no poseen los cultivos. Esto puede sonar obvio en una primera instancia, pero implica que los cultivos ilícitos no generan estas condiciones, por el contrario, las condiciones como la marginalidad, ruralidad, presencia de grupos armados al margen de la ley y condiciones naturales dan paso al desarrollo de la hoja de coca.

La implicación de política que se deriva de este trabajo es clara: para prevenir el Efecto Globo es necesario cambiar las variables sociales y económicas asociadas a la aparición de cultivos ilícitos en los municipios vulnerables que aún están libres de coca.

El método de ensamble de modelos a través de *subbagging* permite evaluar el mejor clasificador entre ensambles de *Logit* y árboles de clasificación para el problema en cuestión. Se evidencia que las regresiones logísticas tienen un desempeño superior al de los árboles bajo la métrica de la AUC de la curva ROC. Por otra parte, la calibración de la proporción de desbalance mostró que para el problema en cuestión la distribución de valores en la base de datos debe ser de 2.8 (74% municipios que no hicieron la transición-26% que se transformaron en cocaleros).

Bibliografía

- Antesana, J. y García, J. (2010). “Estudio Comparativo de la lucha antidroga en Perú y Colombia: La situación de la coca y la cocaína”. Documento 10, Instituto de estudios Internacionales, Pontificia Universidad Católica del Perú.
- Antonini, G. Elisseef, A. y Paleologo, G (2010). “Subagging for credit scoring models”. *European Journal of Operational Research*. ELSEVIER.
- Bagley, B. (2012). “Drug Trafficking and Organized Crime in the Americas: Major Trends in the Twenty First Century”. Woodrow Wilson Center for Scholar. Update on the Americas, August 2012.
- Cárdenas, M. (2011). “Principales determinantes de los cultivos de coca en los municipios de Colombia: Impacto del programa familia guarda bosques (PFGB) 1999-2009”. Seminario de Economía del Banco de la República Sucursal Medellín.
- Castillo, J.C., Mejía, D. y Restrepo, P. (2013). “Illegal Drug Markets and violence in Mexico: The causes beyond Calderón”.
- Castro, J.A., Raffo, L. y Díaz, A. (2016). “Los efectos globo en los cultivos de coca en la Región Andina (1990-2009)”. *Apuntes del Cenes*, vol.35, 207-236
- Granados, A. (2009). “The Failure of US policies in Latin America: From the War on Drugs to the War on Terror”. Thesis, MA War Conflict in the Modern World. Department of Politics and History. Brunel University, London, UK
- James, G. Witten, D. Hastie, T. y Tibshirani, R. (2015). “An introduction to statistical learning”. Sexta Edición. Springer.
- Kotsiantis, S.B. (2007). “Supervised Machine Learning: A Review of Classification Techniques”. *Emerging Artificial Intelligence Applications in Computer Science*. Parte I, 3-22.
- Laffiteau, C. (2011). *The failure of supply side strategies in the war on drugs*. Academia Edu.
- Mejía, D. Restrepo, P. y Rozo, S. (2013). “Efectividad de las campañas de aspersión aérea sobre los cultivos de hoja de coca: Evidencia de un cuasi-experimento en Colombia”. *The World Bank Economic Review*, Volume 31, 570-594
- Mejía, D. y Rico, D. (2010). “La microeconomía de la producción y tráfico de cocaína en Colombia”. Documentos CEDE. Facultad de Economía. Universidad de los Andes.
- Moreno-Sanchez, R. Kraybill, D.S. y Thompson, S. R. (2003). “An Economic Analysis of Coca Erradication Policy in Colombia”. Department of Agricultural, Environmental and Development Economics. Ohio State University.

- Ortiz, C. H. (2003). La guerra contra las drogas es contraproducente: un análisis económico de equilibrio general. *Revista Lecturas de Economía*, 58, 47-68.
- Rico, D. (2013). “Las dimensiones internacionales del crimen organizado en Colombia: Las Bacrim, sus rutas y refugios”. *La diáspora criminal: la difusión transnacional del crimen organizado y como contener su expansión*. Capítulo 2, 27-58.
- Rocha, R. (2010). “Las Nuevas Dimensiones del Narcotráfico en Colombia”. Primera edición, Oficina de las Naciones Unidas contra la Droga y el Delito - UNODC Ministerio de Justicia y del Derecho
- Sandoval, L. López, A y Cárdenas, C. (2009). “Determinantes y Características de la Oferta de Cocaína en Colombia (1989-2006)”. *Revista de la Facultad de Ciencias Económicas de la Universidad Militar Nueva Granada*, volumen 17, 119-208
- Santoni, M.E. y Torres, G. (2007). “LA COCA (*Erythroxylum coca*): Masticando su Historia”. Instituto de Investigaciones en Antropología Médica y nutricional.
- UNODC. (2015). Monitoreo de territorios afectados por cultivos ilícitos 2014.
- UNODC. (2016). Monitoreo de territorios afectados por cultivos ilícitos 2015.
- Vargas, C.E. (2004). “Cultivos ilícitos y erradicación forzosa en Colombia”. Cuadernos de Economía, Universidad Nacional de Colombia.
- Zaman, F. y Hirose, H. (2009). “Effect of subsampling rate on subbagging and related ensembles of stable classifiers”. *Pattern Recognition and Machine Intelligence*, tercera edición, capítulo 8, 44-49.

Apéndice

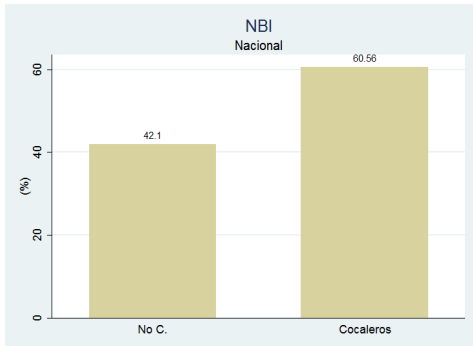


Figura A. NBI Nacional (2005)

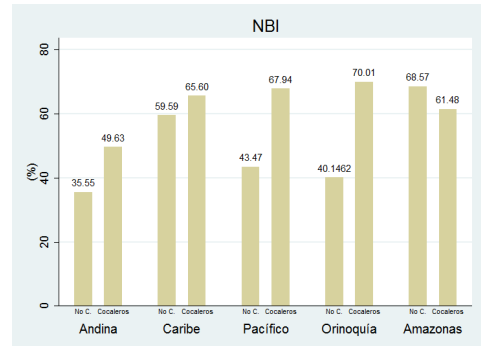


Figura B. NBI regional (2005)

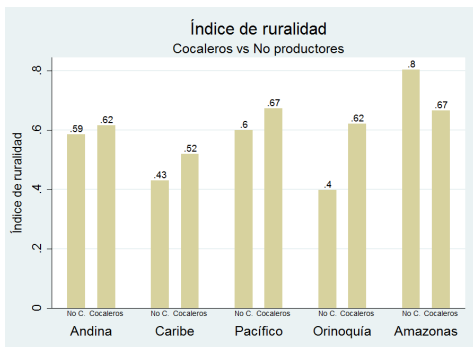


Figura C. Índice de ruralidad

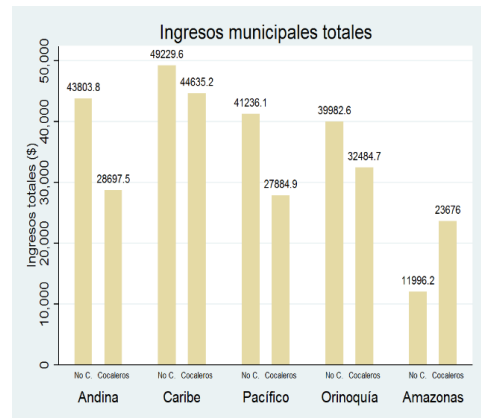


Figura D. Ingresos Totales Municipales

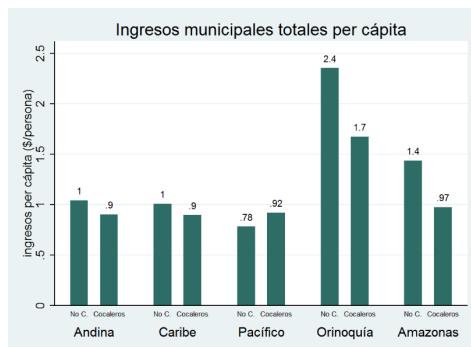
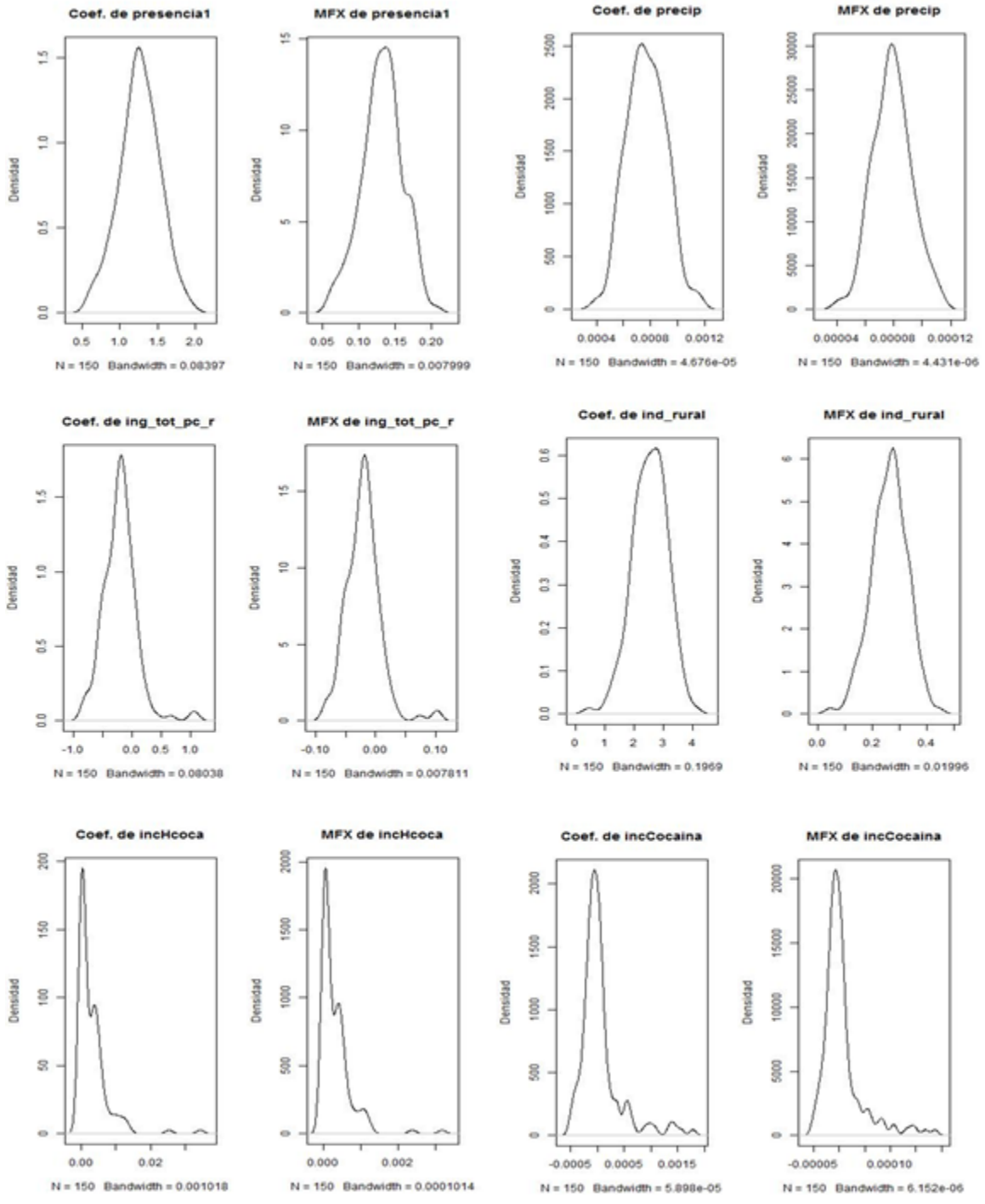
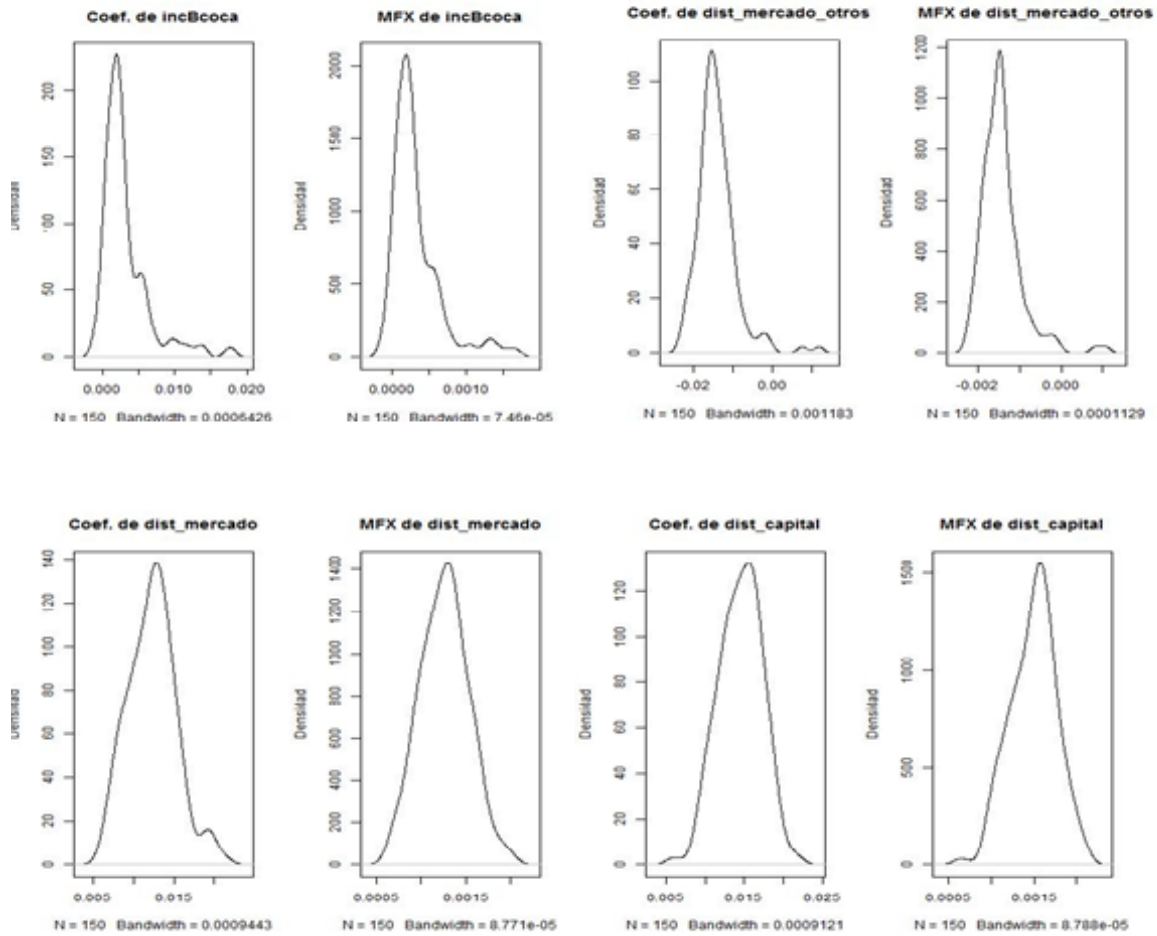


Figura E. Ingresos municipales per cápita





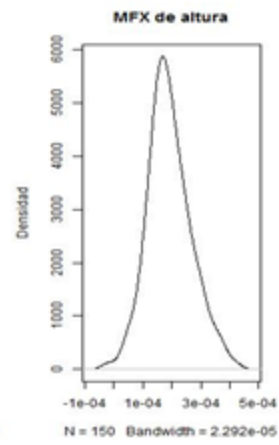
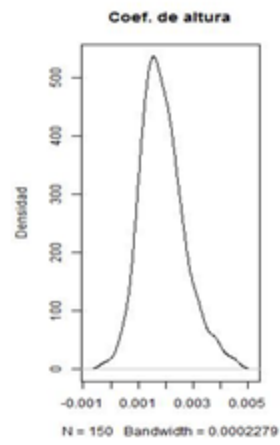
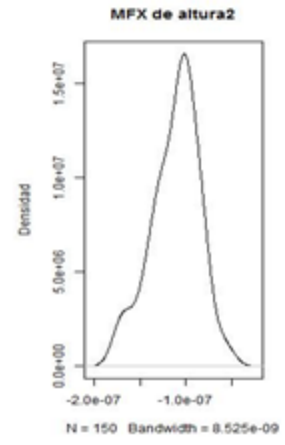
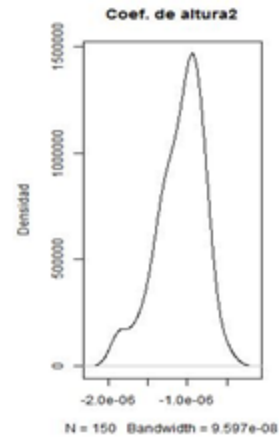
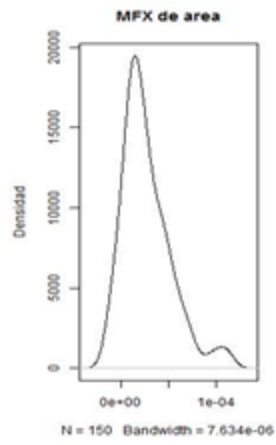
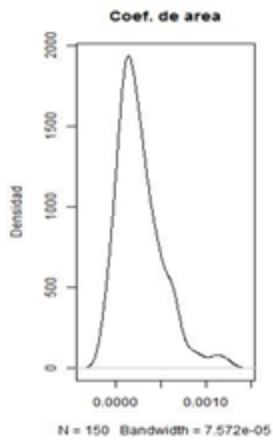


Tabla A. Regresiones logísticas: Promedio de AUC y ACC (precisión) para cada proporción de desbalance a lo largo de 30 iteraciones.

Desbalance	Promedio AUC	Promedio ACC
0.1	0.8457	0.3526
0.2	0.9069	0.524
0.3	0.911	0.6092
0.4	0.9147	0.6589
0.5	0.9176	0.7182
0.6	0.9177	0.7748
0.7	0.9176	0.8108
0.8	0.917	0.8233
0.9	0.9185	0.8506
1	0.9188	0.8817
1.1	0.9184	0.8783
1.2	0.9184	0.8885
1.3	0.9181	0.9137
1.4	0.9185	0.9186
1.5	0.9188	0.9218
1.6	0.9173	0.9279
1.7	0.9176	0.9342
1.8	0.9194	0.9395
1.9	0.9192	0.9493
2	0.9203	0.9399
2.1	0.92	0.9477
2.2	0.9199	0.95
2.3	0.9195	0.9493
2.4	0.9203	0.9528
2.5	0.9205	0.9568
2.6	0.9204	0.9562
2.7	0.9208	0.9601
2.8	0.9214	0.961
2.9	0.9211	0.9626
3	0.9212	0.9624
3.1	0.9212	0.9658
3.2	0.921	0.9677
3.3	0.9213	0.9694
3.4	0.9213	0.9678
3.5	0.9213	0.9719
3.6	0.921	0.9725
3.7	0.9207	0.9725
3.8	0.9209	0.9742
3.9	0.921	0.9731
4	0.9201	0.9751
4.1	0.9209	0.9739
4.2	0.9211	0.9756
4.3	0.9205	0.9757
4.4	0.9205	0.9757
4.5	0.9211	0.9758
4.6	0.921	0.9758
4.7	0.9206	0.9759
4.8	0.9208	0.9758
4.9	0.9207	0.9757
5	0.9205	0.9754

Tabla B. Árboles de clasificación: Promedio de AUC y ACC (precisión) para cada proporción de desbalance a lo largo de 30 iteraciones

Desbalance	PromedioAUC	PromedioACC
0.5	0.789	0.5918
0.6	0.8121	0.6644
0.7	0.8328	0.7132
0.8	0.8476	0.7529
0.9	0.8606	0.784
1	0.8655	0.8097
1.1	0.8697	0.8265
1.2	0.8719	0.8445
1.3	0.8752	0.8567
1.4	0.8709	0.8673
1.5	0.8747	0.8768
1.6	0.872	0.887
1.7	0.8721	0.8967
1.8	0.8739	0.904
1.9	0.8759	0.9107
2	0.8749	0.9158
2.1	0.8718	0.919
2.2	0.8723	0.9243
2.3	0.8701	0.9293
2.4	0.867	0.9351
2.5	0.8687	0.9354
2.6	0.8718	0.9389
2.7	0.8708	0.941
2.8	0.8723	0.9428
2.9	0.8688	0.9433
3	0.8708	0.9438
3.1	0.8702	0.9448
3.2	0.8721	0.9445
3.3	0.8695	0.9458
3.4	0.868	0.9463
3.5	0.8677	0.9482
3.6	0.8674	0.9493
3.7	0.8668	0.9496
3.8	0.8615	0.9505
3.9	0.8655	0.9516
4	0.865	0.9525
4.1	0.8572	0.9538
4.2	0.8624	0.9533
4.3	0.8588	0.9548
4.4	0.8549	0.9549
4.5	0.8545	0.9558
4.6	0.8547	0.9557
4.7	0.8505	0.9568
4.8	0.8531	0.9567
4.9	0.8495	0.9578
5	0.8474	0.9584

Tabla C. Frecuencias de rangos por P-valores de los coeficientes del ensamble de 150 *Logits*, desbalance de 2.8

Variables	Rangos de p valores			
	(0, 0.01]	(0.01, 0.05]	(0.05, 0.1]	(0.1, 1]
Intercepto	72	22	18	38
Área	0	11	17	122
Altura	8	29	23	90
Altura al cuadrado	28	73	35	14
Precipitaciones	148	2	0	0
Índice de ruralidad	55	62	21	12
Distancia a la capital del departamento	145	4	1	0
Distancia al mercado mayorista más cercano	116	30	4	0
Distancia a otros mercados	114	16	7	13
Ingresos reales per cápita	0	0	0	150
Incautaciones de cocaína	0	0	2	148
Incautaciones de base de coca	8	12	16	114
Incautaciones de hoja de coca	6	71	18	55
Presencia de grupos armados al margen de la ley	91	46	7	6
Región Andina	0	0	2	148
Región Caribe	0	1	8	141
Región Orinoquía	0	8	21	121
Región Pacífico	0	0	0	150

Tabla D. Coeficientes y Efectos Marginales promedio del ensamble de 150 *Logits*, desbalance 2.8

VARIABLES	Coeficientes Promedio	Efectos marginales Promedio
Intercepto	-4.76 (3.35)	
Área	0.00027 (0.00025)	0.000027 (0.000026)
Altura	0.0019 (0.0008)	0.00019 (0.000075)
Altura al cuadrado	-0.000001 (2.99E-07)	-1.10E-09 (0.00000003)
Precipitaciones	0.00077 (0.00014)	0.000079 (0.000014)
Índice de ruralidad	2.53 (0.59)	0.26 (0.065)
Distancia de la capital	0.014 (0.0028)	0.0015 (0.00027)
Distancia al mercado mayorista	0.012 (0.003)	0.0013 (0.00027)
Distancia a otros mercados	-0.014 (0.0049)	-0.0015 (0.00049)
Ingreso reales pre cápita	-0.209 (0.28)	-0.021 (0.029)
Incautaciones de Cocaína	0.000055 (0.00038)	0.000006 (0.000039)
Incautaciones de Base de coca	0.0032 (0.0033)	0.00033 (0.00033)
Incautaciones de Hoja de coca	0.0031 (0.0045)	0.0003 (0.00042)
Presencia de grupos armados	1.26 (0.26)	0.13 (0.028)
Andina	-2.59 (3.30)	-0.26 (0.18)
Caribe	-2.92 (3.45)	-0.29 (0.19)
Orinoquía	-3.31 (3.53)	-0.32 (0.19)
Pacífico	-2.36 (3.34)	-0.23 (0.17)
Observaciones por muestra	262	
Número de muestras	150	

Tabla E. Municipios potencialmente coccaleros

Departamento	Municipio	Departamento	Municipio
Antioquia	Arboletes	Chocó	Medio Atrato
Antioquia	Dabeiba	Chocó	Unión Panamericana
Antioquia	Mutatá	La Guajira	Uribia
Antioquia	Necoclí	Meta	San Martín
Antioquia	Sabanalarga	Nariño	Mallama
Antioquia	San Juan de Urabá	Valle del Cauca	Jamundí
Antioquia	San Pedro de Uraba	Putumayo	San Francisco
Caldas	Samaná	Amazonas	La Pedreza
Cauca	Caloto	Amazonas	La Victoria
Cauca	Corinto	Amazonas	Miriti – Paraná
Cauca	La Vega	Amazonas	Puerto Arica
Cauca	Paez	Amazonas	Puerto Santander
Cauca	Rosas	Amazonas	Tarapacá
Cauca	Santander de Quilichao	Guainía	San Felipe
Cauca	Santa Rosa	Guainía	La Guadalupe
Cauca	Toribio	Vaupés	Pacoa
Choco	Jurado	Vaupés	Taraira
Chocó	Lloró	Vaupés	Yavaraté

Tabla F. Distribución departamental de los municipios potencialmente coccaleros

Departamento	Número de Municipios
Antioquia	7
Caldas	1
Cauca	8
Chocó	4
La Guajira	1
Meta	1
Nariño	1
Valle del Cauca	1
Putumayo	1
Amazonas	6
Guainía	2
Vaupés	3
Total	36

Tabla G. Distribución regional de los municipios potencialmente coccaleros

Región	Número de Municipios
Andina	8
Caribe	1
Amazonía	12
Pacífico	14
Orinoquía	1
Total	36