

Agricultura de precisión: Análisis exploratorio de evolución espacio-temporal de la incidencia de Pudrición del Cogollo (PC) de la palma de aceite (*Elaeis guineensis* Jacq.).

Núcleo productivo de San Andrés de Tumaco.

Período: enero 2007 - febrero 2008

Precision Agriculture: An exploratory time - spatial analysis of the rate of Bud Rot of oil palm (*Elaeis guineensis* Jacq.). San Andres, Tumaco productive nucleus. Period: January 2007 - February 2008

AUTORES

Rafael E. Hurtado Camacho,

Investigador PhD. Titular, líder científico. Cenipalma (líder Corpoica hasta el 31 de julio de 2008)

Víctor Rincón Romero,

Auxiliar de investigación. Cenipalma

Hugo M. Mercado Marino

Auxiliar de investigación, líder desde el 1 de agosto de 2008.

Corpoica, C.I. Tibaitatá

Douglas Gómez Latorre

Profesional, apoyo trabajo de campo y revisión bibliográfica. Corpoica, C.I. Tibaitatá.

Palabras CLAVE

Palma de aceite, Pudrición del Cogollo, Análisis exploratorio de datos, Incidencia, propagación, Análisis temporal, Análisis espacial.

Oil Palm, Bud Rot, Precision Agriculture, Geoestadistic, Data Analysis, Incidence, Spreading, Spatial Analysis, Temporal Analysis.

Recibido: 1 junio 2009

Aceptado: 5 agosto 2009

Resumen

Colombia, primer productor de aceite de palma en Latinoamérica y quinto en el mundo, enfrenta un problema de gran magnitud como lo es la enfermedad de la Pudrición del Cogollo (PC), cuyos efectos se manifiestan en tres de las cuatro zonas productoras del país con graves consecuencias sociales y económicas. El núcleo productivo más afectado es San Andrés de Tumaco (Zona Occidental palmera), donde la mayor parte del área sembrada (alrededor de 36.000 hectáreas) ha sido atacada a tal punto, que empresas, consorcios, pequeños y medianos palmicultores tuvieron que erradicar parcial o totalmente sus plantaciones. Este documento presenta un análisis exploratorio de la evolución espacio-temporal de la PC de la palma de aceite (*Elaeis guineensis* Jacq.), en Tumaco, entre enero de 2007 y febrero de 2008, que pretende convertirse en una herramienta útil de los productores para la toma de decisiones. La fase I de la investigación fue financiada por el Ministerio de Agricultura y Desarrollo Rural, apoyado técnicamente por Corpoica por intermedio del Centro de Investigación Tibaitatá y el Centro Experimental El Mira. La fase II fue financiada por Fedepalma - Fondo de Fomento Palmero (FFP). El investigador Rafael E. Hurtado Camacho lideró ambas fases.

Summary

Colombia is the fifth worldwide producer of oil palm (*Elaeis guineensis* Jacq.) after Malaysia, Indonesia, Nigeria and Thailand. The crop has expanded throughout the country, passing from 11.500 Ha in 1970 to more than 157.000 Ha in 2004; unfortunately, in Central and South America this crop is affected by a wide variety of pests and diseases that could threaten its persistence. Diseases such as "bud rot" (BR) represent the highest threat to the oil palm crop, these have considerable economical importan-



ce and have even caused the disappearance of several crops in South America. Diseases of the BR type can assume several forms: a lethal one that is predominant in the Amazon region of Ecuador, Brazil and certain zones of Colombia and Surinam; and a non-lethal form, with a high recovery rate, which is mostly found in the Colombian plains. Up to date it is not known if these are manifestations of the same disease, or if they are distinct pathological and physiological phenomena. However, no etiological agent has been found, and therefore what is referred to as “BR” in this research, remains a controversial issue. The BR is widely distributed and spread throughout the western producing region of Colombia, which has almost 36,000 Ha, corresponding mainly to the department of Nariño and almost totally to the municipality of Tumaco, where the agricultural industry of this product supports, at a large extent, the economy of the region. According to Corpoica (2007), oil palm generates 7,000 direct jobs and 3,500 indirect jobs; which translates in over 44,000 people depending on this crop, taking into account that the total population of this municipality is around 160,000 inhabitants. This research aims to illustrate through the time evolution model of the BR and the partial exploratory analysis of the obtained data, the space-time evolution of the bud rot over African oil palm in South-East Colombia, producing nucleus of the municipality of San Andrés de Tumaco (Nariño), with primary and secondary information collected from April 2007 until February 2008. The present work does not seek to provide a direct solution to the problem, none withstanding it pretends to give and approach to the study of the behavior of the disease, so that agronomists, economists and governments involved in decision making, can design preventive, adaptative and/or re-convertive measures to manage the disease.



Introducción

La investigación pretende establecer el avance de la incidencia de la Pudrición del Cogollo (PC) en los cultivos de palma de aceite en el suroccidente colombiano, específicamente en el núcleo productivo de San Andrés de Tumaco situado en el departamento de Nariño (Colombia). Los resultados son el producto de la integración de dos fases de estudio: una denominada “Análisis espacial y temporal de la Pudrición del Cogollo (PC) de la palma de aceite *Elaeis guineensis* Jacq.” (PC fase 1) y otra, PC fase 2.

La primera se realizó con miras a predecir la incidencia de la enfermedad en Nariño, y comprende el período de febrero a octubre de 2007 para grandes, medianos y pequeños productores asociados en un área total de 29.000 hectáreas. La PC fase 2 se desarrolló entre febrero y junio de 2008, y corresponde a la georreferenciación de 5.000 hectáreas de pequeños productores no asociados del núcleo productivo de palma de aceite en el municipio de Tumaco.

Las actividades llevadas a cabo en los procesos de recolección de la información en el campo se enmarcaron en la reestructuración de la información

cartográfica y alfanumérica de la primera fase, lo que permitió estandarizar las bases de datos de la información recolectada. Debido a que en esta fase no se contempló a los pequeños productores no asociados, se implementó un nuevo proyecto financiado directamente por el Fondo de Fomento Palmero (FFP) por intermedio de Cenipalma, cuyo objetivo es incorporar al modelo la georreferenciación de 5.000 hectáreas y determinar mediante modelos de análisis espacial y temporal, la incidencia y la propagación de la PC.

Para establecer la dinámica espacio-temporal de la incidencia acumulada de la PC en todos los lotes cultivados con la oleaginosa en el núcleo productivo, se integraron las bases de datos geográficas de las fases 1 y 2, con lo cual se logró consolidar la información en una cobertura de polígonos que representan cada uno de los lotes sembrados con palma de aceite.

En la fase 1 se estudió el comportamiento espacial de la PC, con el fin de realizar la extrapolación de predicciones de incidencia de sitios no muestreados, pero se consideró que el comportamiento de la enfermedad era continuo dentro del núcleo productivo. Ahora bien, con la integración de los datos de la fase 2 se identificaron conductas particulares en el núcleo que



motivaron la realización de un estudio sectorizado, y en esa dirección se procedió a analizar la discontinuidad espacial para identificar factores naturales como el río Mira y factores antrópicos como la carretera en la dirección Pasto-Tumaco. Estos se constituyen en elementos de interrupción del flujo de incidencia de la PC, que permiten dividir el núcleo productivo en tres zonas más homogéneas.

Como se manejaron dos tiempos de captura de información primaria de incidencia de PC (abril de 2007 en la fase 1 y febrero de 2008 en la fase 2), y tomando en cuenta que la gran mayoría de los lotes observados en las dos fases no son los mismos, fue necesario utilizar el modelo de predicción temporal desarrollado en la primera, para realizar el cálculo de la incidencia mensual de PC para el período enero 2007 a febrero 2008, a partir de las incidencias acumuladas observadas (IAO) en abril de 2007 y febrero de 2008 para las fases 1 y 2, respectivamente.

A las incidencias mensuales del período de 14 meses se les aplicó el análisis exploratorio de datos. Vale aclarar que ello se hizo de forma independiente para cada una de las zonas definidas por discontinuidad geográfica, en principio, con el fin de entender la incidencia a lo largo del período. De manera que lo primero fue el análisis estadístico unidimensional, mediante gráficos y estadística descriptiva, que permitió obtener un resultado inicial de la distribución mensualizada de la PC y el desplazamiento gradual de su incidencia en el conjunto de lotes.

Después se realizó el análisis multidimensional usando la técnica de clúster, que permitió agrupar los datos y obtener rangos únicos para clasificar la incidencia de PC en todos los meses.

Para completar el análisis espacio-temporal, por último, los rangos obtenidos se materializaron en cartas de coropleta¹ para cada uno de los meses del período, las cuales, superpuestas de forma gradual, permitieron estudiar el comportamiento de la incidencia de la PC en el espacio y en el tiempo.

Proceso de levantamiento de información en el campo²

Metodología

Para seguir este proceso se organizó la información base de tal forma que las visitas al campo fueran más eficientes. En el momento de la georreferenciación, a cada parcela se le asignó un número único e irrepetible como identificador o llave primaria.

Todos los días se verificaba la ubicación espacial de la información recolectada y la consistencia de los datos alfanuméricos. Una vez hechos los controles, se tabulaban los datos para establecer la relación existente entre los de la encuesta (previamente realizada entre diferentes actores del núcleo productivo de Tumaco) y los recolectados en el campo. La georreferenciación de las parcelas consistió en la toma de las coordenadas geográficas con el dispositivo GPS en la zona central del lote, al cual por lo general se accedía a pie (Figura 1).

Durante el proceso llevado a cabo en la fase de campo para el levantamiento de la información, se surtieron las siguientes etapas:



Figura 1. Proceso de visita de campo para georreferenciación de las parcelas.

¹ La carta coropleta es la representación de variables cualitativas o cuantitativas relativas a áreas geográficas por medio de una escala de tonos graduados.

² Se presentan las actividades realizadas durante el período comprendido entre el 17 de febrero (fecha de inicio del proyecto de investigación) y el 17 de marzo de 2008 (fecha en que finalizó la primera fase del trabajo de campo), cuyo objetivo principal fue la georreferenciación de más de 5.000 hectáreas de pequeños y medianos productores no asociados dedicados parcial o totalmente al cultivo de palma africana y que presentan problemas de PC.

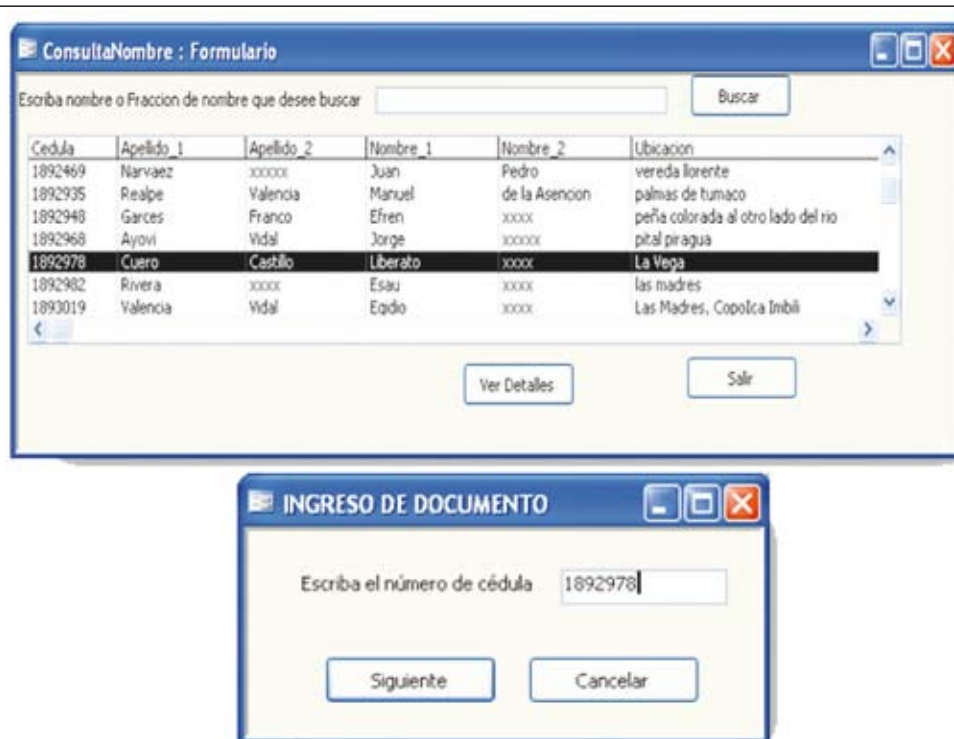


Figura 2. Módulos de consulta por nombres y apellidos, y de consulta por número de cédula.

Encuestas de información básica de pequeños productores no asociados de palma de aceite. La encuesta³ contiene ítems como: nombre del propietario, cédula de ciudadanía, extensión de la plantación, afectación de la plantación y área sembrada, entre otros. De igual manera, la ubicación geográfica (vereda), que constituye el elemento de mayor peso en el proceso de campo.

Generación del módulo de consulta de información. A partir de la tabulación de las encuestas se generó un módulo de consulta básico, que permitía visualizar de manera ágil y rápida la información obtenida (Figura 2).

Generación de la capa de puntos de las parcelas de pequeños y medianos productores de palma de aceite, en proyección Magna Sirgas. A partir de la información recolectada en el campo se generaba la cobertura de puntos de las parcelas georreferenciadas; la capa fue proyectada al sistema geográfico Magna Sirgas (Figura 3).

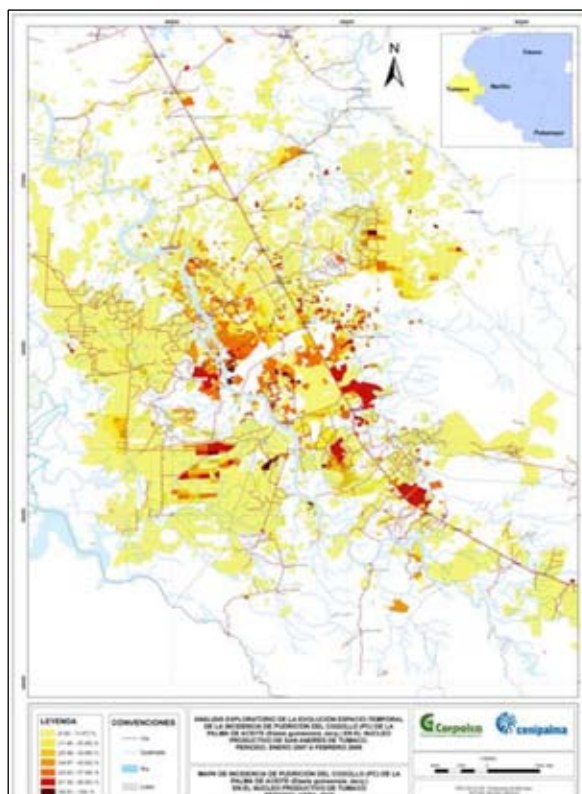


Figura 3. Cobertura e incidencia de la PC de la palma aceitera en Tumaco, abril de 2007.

3 Encuesta concertada con los productores, Alcaldía de Tumaco, Cordeagropaz y Corpoica. Enero de 2008.



Estructuración de la información

Los datos correspondientes al proyecto fase 1 y proyecto fase 2 de la zona palmera se integraron con el fin de contar con información suficiente y consistente para explicar la realidad del fenómeno.

Estructuración de la información geográfica y alfanumérica del proyecto PC fase 2

De la etapa de campo del proyecto PC fase 2 se obtuvieron como productos parciales: una cobertura de puntos de la ubicación de las fincas palmeras de los pequeños y medianos productores encuestados, y una tabla en formato digital con los datos de las encuestas realizadas a dichos palmicultores. Ambos productos se estructuraron usando la metodología de diseño de bases de datos relacionales, que comienza por elaborar un modelo conceptual que abstrae los objetos de la realidad que se analiza, e integra los elementos del modelo Entidad-Relación (ER) determinando las entidades y sus atributos.

Validación de la consistencia de los datos del proyecto PC fase 2

La validación de los datos corresponde a la evaluación y corrección de los posibles errores e inconsistencias que estos puedan presentar. Se hace primero integrando los datos de las dos fases del proyecto para identificar la ubicación de puntos de la fase 2 en polígonos de la fase 1, y luego integrando los de la fase 2 en la base geográfica. Al tener de fondo las coberturas de geografía base, se inspeccionan todos los puntos de identificación de las fincas para detectar errores sistemáticos (debidos al nivel de exactitud de los navegadores GPS) relacionados con la posible ubicación de puntos sobre la cobertura de río o centros poblados.

Resultados de revisión de campo

La revisión de campo arroja como producto la corrección de la ubicación de los puntos con inconsistencias. Así mismo, se realiza la captura de 13 puntos en lotes correspondientes a 12 encuestas. Los siguientes resultados corresponden a los datos finales totales capturados:

Lotes georreferenciados:	1.319
Palmicultores encuestados:	1.066

Análisis de discontinuidad espacial en los datos

Debido a la percepción de los habitantes de la zona y de la misma comisión de profesionales que realiza el trabajo de campo sobre la dinámica de la incidencia de la PC, se establece la opción de analizar los datos de manera independiente, dado que desde un principio la zona con mayor incidencia fue identificada al oeste del río Mira. Por tanto, se procede al análisis tomando en cuenta las discontinuidades geográficas (antrópicas: la vía; naturales: el río) que se convierten en barreras “naturales” que impiden o frenan la propagación del inóculo de la PC, pero que a su vez generan subgrupos o zonas, sobre las cuales se realizan análisis diferenciados y consistentes con la realidad.

Según lo dicho por François (2004):

“...el análisis de discontinuidad espacial parte de la identificación de indicadores de discontinuidad, los cuales de manera independiente se consideran discontinuidades elementales, de igual forma aclara que una de las clasificaciones de discontinuidad es la discontinuidad de barrera, para la cual el indicador más simple de discontinuidad espacial lo constituyen las barreras que perturban el flujo del sistema”.

Para efectos del estudio, se determinan como *barreras* el río Mira y la carretera Tumaco-Pasto, dado que interrumpen el uso agrícola de los suelos. Así, se tienen tres zonas (Figura 4) para el análisis independiente de los datos:

- Subzona Mira: ubicada en el costado oeste del río Mira

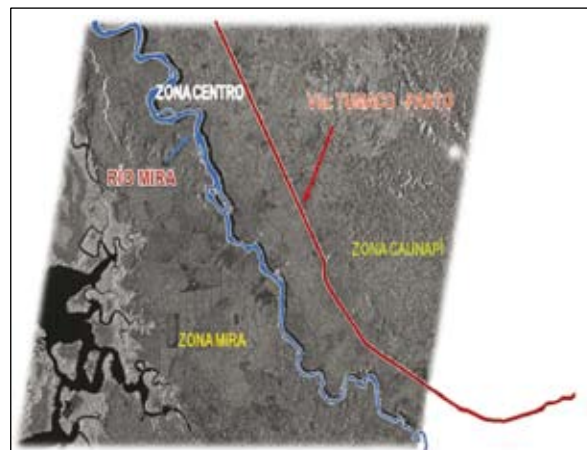
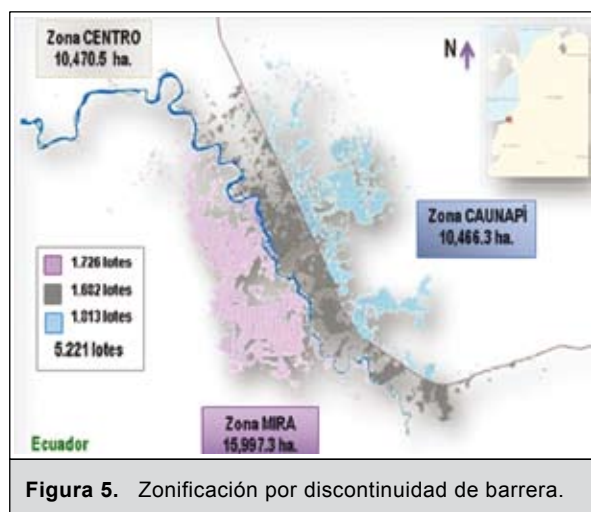


Figura TerraSAR-X, modo Strip Map, resolución 3 metros. Feb. 2008, Infoterra – Geospatial, Colombia

Figura 4. Identificación de las barreras geográficas.

- Zona Centro: limitada por el oeste con el río Mira y por el este con la carretera Tumaco-Pasto
- Zona Caunapí: se encuentra ubicada al este de la carretera Tumaco-Pasto, y es recorrida en su interior por el río Caunapí.

De manera independiente para los datos de la fase 1 y fase 2, se seleccionan los elementos que pertenecen a cada una de las zonas. Para facilidad del manejo de la información se crean coberturas geográficas independientes (Figura 5).



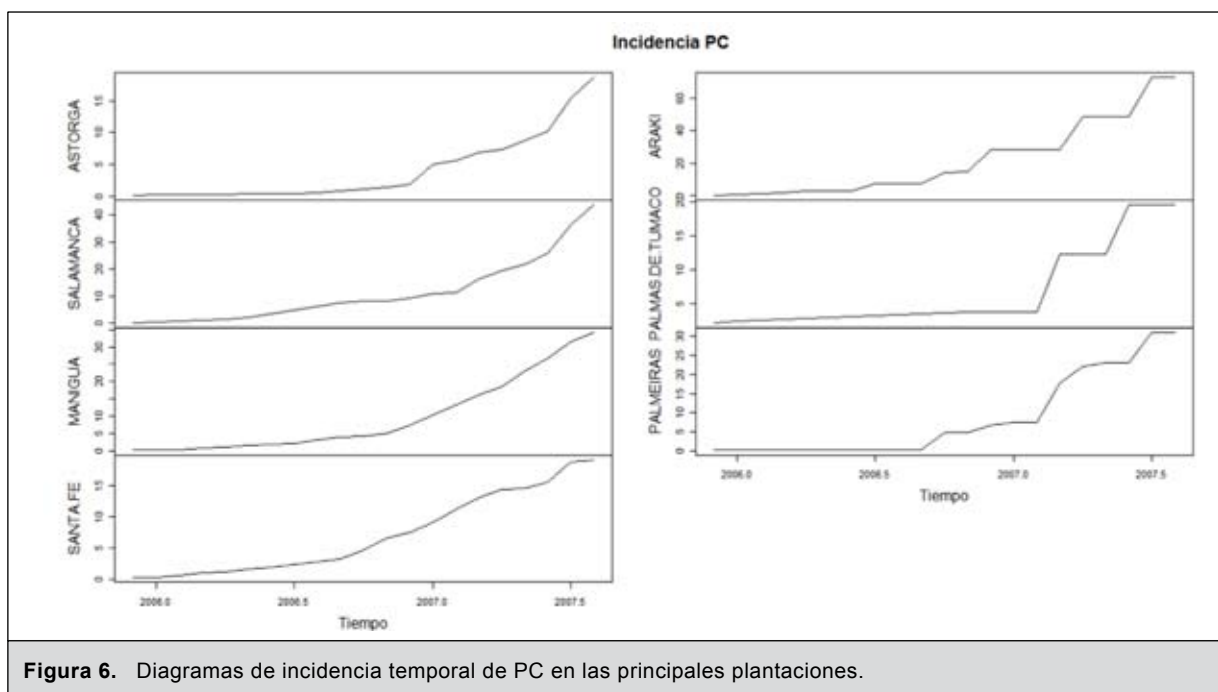
Consolidación de datos e integración de información

Para facilitar el análisis espacio-temporal de la enfermedad, la consolidación de los datos corresponde al cálculo de las incidencias mensuales en el período de enero de 2007 hasta febrero de 2008 para cada uno de los predios de las dos fases del proyecto.

Cálculo de la incidencia mensual acumulada

El cálculo de la incidencia mensual acumulada para el período enero de 2007 a febrero de 2008 se realizó a partir de la aplicación de modelos de suavizamiento exponencial doble de Holt, obtenidos con las series de tiempo de información mensual suministrada por las plantaciones de la región, entre diciembre de 2005 y agosto de 2007 en la fase 1. Dichos modelos se desarrollaron puntualmente sobre los cultivos de las principales empresas del núcleo productivo, y se escogieron debido a que las series temporales de las incidencias mostraban una clara tendencia creciente de la enfermedad (Figura 6).

Las predicciones de la incidencia de PC en cada una de las plantaciones a k períodos, se obtienen de la aplicación de la siguiente expresión:





$$\hat{y}_{T+k} = a_T + b_T k$$

Donde, $\hat{y}_{T+k}$ es el valor pronosticado de incidencia de PC en la respectiva plantación en k periodos. Los términos a_T y b_T se obtienen a partir de las expresiones recursivas siguientes:

$$a_T = \alpha y_T + (1 - \alpha) (a_{T-1} + b_{T-1})$$

$$b_t = \beta (a_t - a_{t-1}) + (1 - \beta) b_{t-1}$$

Los resultados de la estimación de parámetros para los modelos de cada una de las plantaciones, así como las respectivas medidas de los errores de predicción se muestran en la Tabla 1.

La Figura 7 y la Tabla 2 muestran los valores pronosticados de incidencia de PC en las empresas dadas. En ellas se aprecia la tendencia creciente de la enfermedad en la zona.

Como se explicó, los modelos solo aplican consistentemente a los cultivos de las empresas en los que se desarrollaron y ello constituye un problema al momento de extrapolar sus resultados a las demás fincas palmeras. Por tanto, como hipótesis de soporte se asumió la primera ley de la geografía, según la cual los puntos más cercanos espacialmente tienden a tener valores más parecidos que los puntos más alejados entre sí (Tobler, 1970). De manera que en las zonas definidas (Mira, Centro y Caunapí) se tuvieron en cuenta únicamente los modelos de las empresas palmeras con presencia en cada una de ellas y a partir de ahí se obtuvieron los datos de incidencias mensuales acumuladas de las demás fincas productoras.

Análisis exploratorio

Metodología de trabajo

Se realizó un análisis estadístico descriptivo y exploratorio de la incidencia de PC, caracterizando las distribuciones muestrales mensuales que permitieron identificar los valores medios, la dispersión y la existencia de observaciones atípicas. A ello se integró el efecto de la toma de datos en el tiempo y el espacio.

Consistencia de la información y crítica estadística

La evaluación de la consistencia de la información residió en la revisión y estandarización de los datos de incidencia recogidos en el campo, en formatos adecuados para su análisis. La crítica estadística busca evaluar la calidad de la información disponible, revisando posibles errores contenidos en los datos y su atipicidad, así como verificando que esté completa.

Análisis estadístico unidimensional

Con el fin de caracterizar las distribuciones muestrales de las incidencias de PC, los datos se calcularon mediante estadísticas de resumen (promedio, mediana, coeficientes de variación, asimetría y kurtosis), y se construyeron histogramas.

Análisis de tendencia espacial

Para evaluar si el comportamiento de la incidencia de PC presentaba alguna tendencia, es decir, algún tipo de relación funcional de las incidencias en la dirección de las coordenadas geográficas este y norte, se construyeron diagramas de dispersión. En el eje x se ubicó a las coordenadas este o norte del medio de los lotes de palma, y en el eje y el valor de la incidencia en el mes de análisis, con el fin de definir si existe o no una relación causa efecto entre dos variables, y qué tan firme es.

Tabla 1. Tabla de modelos temporales

Plantación	Parámetros del modelo		Medidas del error					
	Alpha	Beta	ME	RMSE	MAE	MPE	MAPE	MASE
Astorga	0,9161	0,5108	0,3	1,124	0,5	6,35	10,7	0,539
Salamanca	0,9999	0,6226	0,57	1,88	0,95	5,39	8,87	0,44
Manigua	0,8531	0,6605	0,24	0,84	0,53	5,46	12,07	0,31
Santa Fe	0,8145	0,2895	0,19	0,78	0,52	2,74	9,73	0,55
Araki	0,1921	0,1921	2,30	5,50	3,77	25,49	31,98	1,04
Palmas de Tumaco	0,5895	0,2118	0,43	2,27	1,00	2,91	7,79	1,15
Palmeiras	0,2669	0,2667	0,52	2,70	1,63	2,21	47,97	1,06

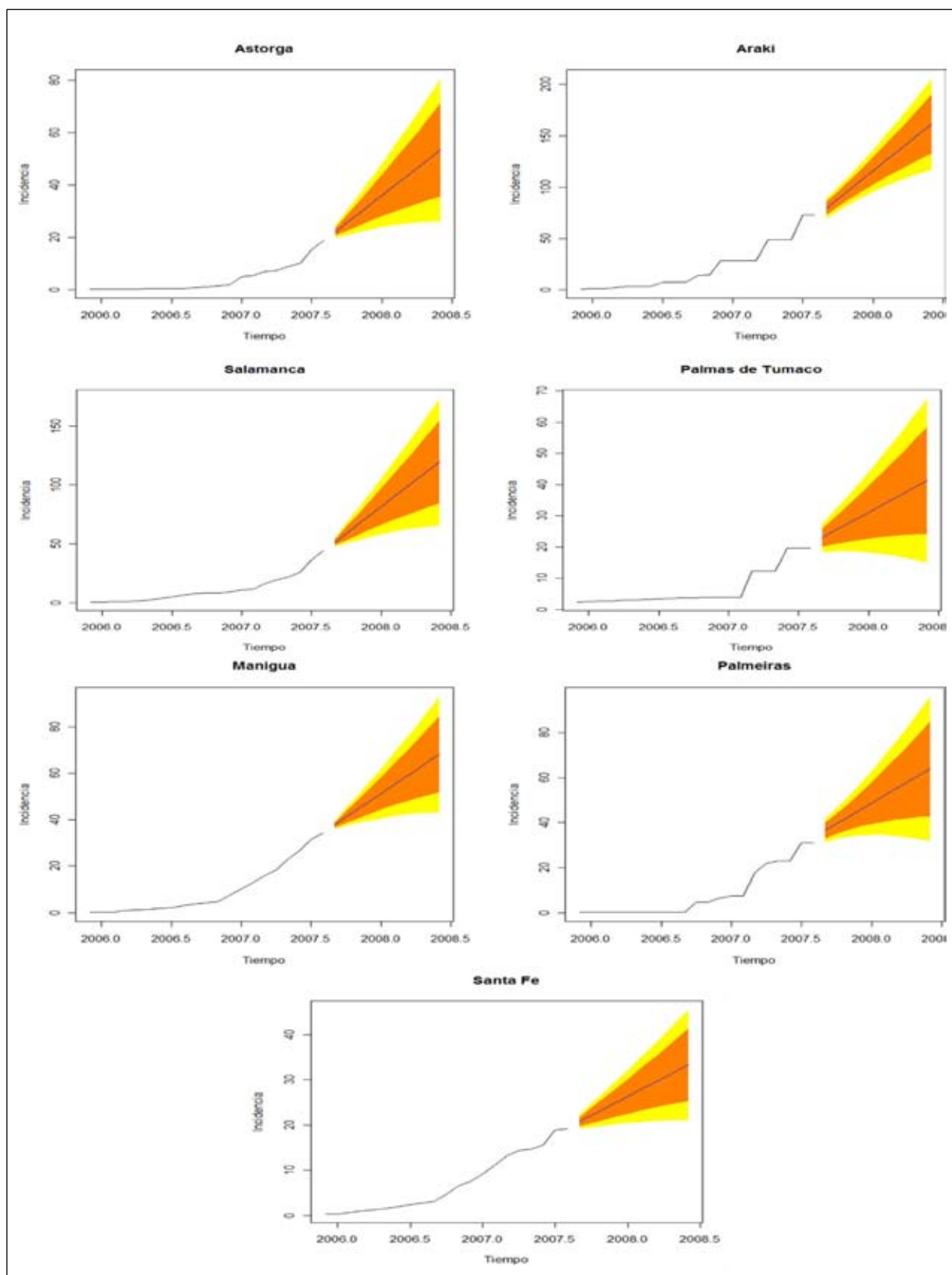


Figura 7. Gráficos de la predicción de incidencia de PC en las principales plantaciones.

**Tabla 2.** Tabla de predicción de incidencia de PC para las palmas de las empresas modeladas

Período	Astorga	Salamanca	Manigua	Santa Fe	Araki	Palmas de Tumaco	Palmeiras
Sep-07	22,16	51,38	37,83	20,77	79,29	22,95	36,59
Oct-07	25,63	58,96	41,21	22,16	88,42	25,00	39,63
Nov-07	29,10	66,54	44,59	23,56	97,56	27,04	42,67
Dic-07	32,57	74,12	47,97	24,95	106,69	29,09	45,71
Ene-08	36,04	81,70	51,35	26,35	115,83	31,13	48,74
Feb-08	39,51	89,28	54,74	27,74	124,96	33,18	51,78

Análisis estadístico multidimensional

Por medio del análisis de conglomerados, es decir, usando algoritmos de clasificación jerárquica, se agruparon los valores de incidencia de PC en meses de mayor variabilidad. El número de grupos que se habría de formar se estableció con base en el análisis del dendograma, tomando en cuenta que como resultado se debe obtener una tipificación que maximice la homogeneidad de las incidencias dentro de cada grupo y la heterogeneidad entre los mismos. Con esta clasificación se generaron representaciones espaciales que permitieron conocer la distribución de la incidencia en el tiempo y el espacio.

Resultados obtenidos

Descripción general del conjunto de datos

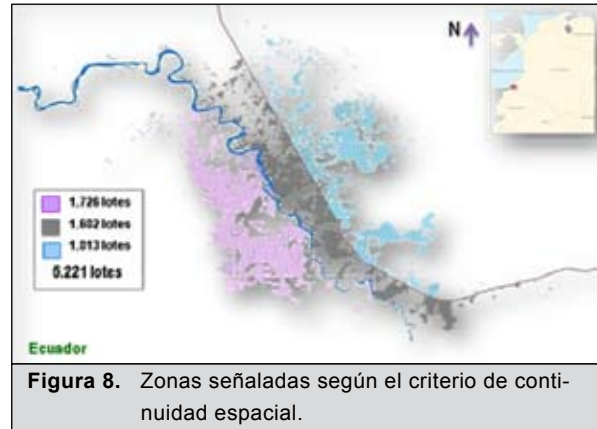
Los datos corresponden a la información de palmas de aceite afectadas por PC en los lotes de los grandes y medianos productores del municipio de San Andrés de Tumaco. Para el análisis se definió la variable *incidencia* como la razón del número de palmas infectadas por PC sobre el total de palmas sembradas en el lote:

$$\text{INCIDENCIA} = \frac{\text{No. Palmas con PC en lote}}{\text{No. Total de palmas sembradas en el lote}}$$

Debido al criterio de continuidad espacial, los datos fueron fraccionados en tres subzonas denominadas Mira, Centro y Caunapí, de acuerdo con las divisiones espaciales dadas por el río Mira y la vía Tumaco-Pasto (Figura 8).

Los datos se estructuraron bajo los siguientes criterios:

- Solo se conservan el identificador de la finca, las coordenadas y las incidencias, que son los valores necesarios para realizar el análisis exploratorio de los datos.

**Figura 8.** Zonas señaladas según el criterio de continuidad espacial.

- En la primera fila de datos se ubican datos no nulos.
- El tipo de datos que se ha de trabajar está constituido por variables cuantitativas continuas.

Análisis exploratorio de los bloques de datos

Los datos contemplados en el análisis exploratorio se describen en la Tabla 3.

Análisis estadístico unidimensional

Las Tablas 4, 5 y 6 muestran los estadísticos descriptivos de las variables de incidencia mensual de PC para cada una de las Subzonas: Mira, Caunapí y Centro.

La Tabla 4 muestra que la incidencia de PC cambia en el tiempo para la Subzona Mira. Mientras que

Tabla 3. Descripción general de los datos tomados en cuenta en el análisis exploratorio

SUBZONAS		
Mira	Caunapí	CENTRO
1727 lotes	1681 lotes	1812 lotes
14 columnas que representan el porcentaje de incidencia en el mismo número de meses (de enero de 2007 a febrero de 2008).		

Tabla 4. Estadísticos descriptivos de las variables de incidencia mensual de PC para el conjunto de datos Mira

	Inc. Ene 07	Inc. Feb 07	Inc. Mar 07	Inc. Abr 07	Inc. May 07	Inc. Jun 07	Inc. Jul 07	Inc. Ago 07	Inc. Sep 07	Inc. Oct 07	Inc. Nov 07	Inc. Dic 07	Inc. Ene 08	Inc. Feb 08
Mínimo	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Máximo	0.921	0.941	0.945	0.972	1	1	1	1	1	1	1	1	1	1
Promedio	0.067	0.074	0.081	0.095	0.125	0.161	0.200	0.247	0.298	0.355	0.419	0.490	0.570	0.659
Mediana	0.000	0.000	0.000	0.008	0.042	0.082	0.127	0.180	0.238	0.306	0.389	0.480	0.563	0.655
Desviación estándar	0.145	0.152	0.160	0.168	0.172	0.177	0.180	0.184	0.189	0.195	0.202	0.210	0.217	0.220
Desviación mediana	0.000	0.000	0.000	0.008	0.018	0.018	0.021	0.028	0.035	0.042	0.040	0.042	0.050	0.065
Kurtosis	9.603	8.822	8.139	7.466	7.192	6.918	6.395	5.884	5.200	4.597	4.147	4.001	4.151	4.468
Coefficiente de asimetría	2.507	2.389	2.277	2.149	2.093	2.011	1.846	1.650	1.333	0.933	0.447	-0.05	-0.50	-0.88
Coefficiente de variación	2.148	2.061	1.961	1.769	1.371	1.101	0.901	0.747	0.634	0.548	0.482	0.428	0.380	0.334
Coefficiente mediano de variación	NaN	NaN	NaN	1	0.441	0.223	0.165	0.156	0.149	0.139	0.103	0.088	0.089	0.099

Tabla 5. Estadísticos descriptivos de la incidencia mensual de PC para el conjunto de datos Caunapí

	Inc Ene 07	Inc Feb 07	Inc Mar 07	Inc Abr 07	Inc May 07	Inc Jun 07	Inc Jul 07	Inc Ago 07	Inc Sep 07	Inc Oct 07	Inc Nov 07	Inc Dic 07	Inc Ene 08	Inc Feb 08
Mínimo	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Máximo	0,9652	0,9753	0,9869	1,0000	1,0000	1,0000	1,000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Promedio	0,0554	0,0609	0,0684	0,0801	0,1049	0,1314	0,159	0,1891	0,2204	0,2535	0,2882	0,3246	0,3629	0,4030
Mediana	0,0000	0,0000	0,0000	0,0127	0,0417	0,0674	0,096	0,1274	0,1607	0,1956	0,2317	0,2689	0,3073	0,3473
Desviación estándar	0,1452	0,1514	0,1579	0,1638	0,1649	0,1667	0,169	0,1725	0,1761	0,1800	0,1845	0,1891	0,1936	0,1978
Desviación mediana	0,0000	0,0000	0,0000	0,0127	0,0135	0,0157	0,016	0,0217	0,0242	0,0288	0,0311	0,0342	0,0347	0,0390
Kurtosis	11,185	10,704	10,211	9,7453	9,6425	9,3923	9,005	8,4691	7,8744	7,2699	6,7031	6,2120	5,8138	5,4985
Coefficiente de asimetría	2,9848	2,9059	2,8184	2,7269	2,7083	2,6462	2,542	2,3990	2,2339	2,0609	1,8865	1,7290	1,6087	1,5316
Coefficiente de variación	2,6203	2,4857	2,3068	2,0450	1,5718	1,2692	1,061	0,9119	0,7988	0,7101	0,6401	0,5826	0,5335	0,4907
Coefficiente mediano de variación	NaN	NaN	NaN	1,0000	0,3237	0,2329	0,172	0,1703	0,1506	0,1472	0,1342	0,1272	0,1129	0,1123

de enero a octubre de 2007 su distribución es claramente asimétrica con tendencia a concentrarse hacia valores bajos (es decir, incipiente propagación de la enfermedad) y se presentan fincas atípicas con altos niveles de PC, de noviembre a enero aumentan los valores medios de incidencia, cuyos promedios y medianas además son muy similares. Lo que muestra que el promedio es un buen indicador del centro del conjunto de datos.

Los valores de los coeficientes de asimetría exhiben una alta concentración hacia los valores centrales de

incidencia, y esto indica que la gran mayoría de las fincas presenta niveles de PC cercanos al promedio. En febrero la incidencia media es muy alta, y empieza a concentrarse en valores altos; en efecto, son pocas las zonas con baja incidencia de la enfermedad.

Los histogramas mensuales de incidencia (Figura 9) permiten apreciar el cambio en la forma de la distribución muestral de PC, que es asimétrica hasta octubre de 2007 y luego aumenta de manera considerable. En general hay un avance progresivo de la enfermedad.

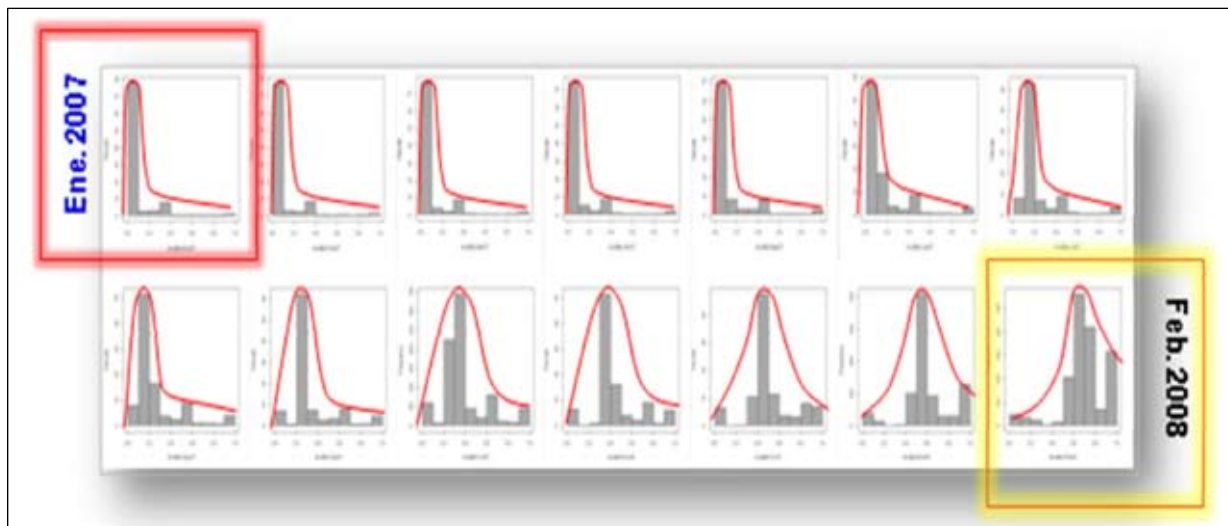
**Tabla 6.** Estadísticos descriptivos de la incidencia mensual de PC para el conjunto de datos Centro

	Inc Ene07	Inc Feb07	Inc Mar07	Inc Abr07	Inc May07	Inc Jun07	Inc Jul07	Inc Ago07	Inc Sep07	Inc Oct07	Inc Nov07	Inc Dic07	Inc Ene08	Inc Feb08
Mínimo	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Máximo	0,9266	0,9482	0,9726	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,000	1,0000	1,000	1,0000	1,0000
Promedio	0,1506	0,1612	0,1742	0,1913	0,2181	0,2478	0,280	0,3165	0,3560	0,398	0,4461	0,497	0,5531	0,6136
Mediana	0,0000	0,0052	0,0296	0,0571	0,0877	0,1205	0,157	0,1990	0,2502	0,295	0,3455	0,402	0,4642	0,5303
Desviación estándar	0,2149	0,2227	0,2312	0,2392	0,2415	0,2443	0,247	0,2514	0,2555	0,260	0,2642	0,267	0,2704	0,2705
Desviación mediana	0,0000	0,0052	0,0296	0,0571	0,0877	0,0934	0,097	0,0963	0,0895	0,094	0,0969	0,097	0,1198	0,1282
Kurtosis	3,8258	3,6229	3,4263	3,2587	3,1022	2,9385	2,766	2,5821	2,4208	2,309	2,2462	2,246	2,3159	2,4108
Coefficiente de asimetría	1,2852	1,2265	1,1641	1,1041	1,0583	0,9962	0,913	0,8056	0,6774	0,533	0,3765	0,218	0,0600	-0,0638
Coefficiente de variación	1,4267	1,3818	1,3270	1,2500	1,1072	0,9861	0,883	0,7944	0,7177	0,652	0,5923	0,538	0,4889	0,4409
Coefficiente mediano de variación	NaN	1,0000	1,0000	1,0000	1,0000	0,7747	0,616	0,4837	0,3576	0,320	0,2805	0,243	0,2581	0,2417

De esta información, se aprecia que la incidencia de la PC ha presentado valores relativamente bajos para el intervalo comprendido entre enero y septiembre de 2007. Se ve un claro comportamiento unimodal asimétrico a la derecha, con algunas pocas fincas para las cuales la afectación del inóculo es relativamente alta. A partir de octubre de 2007, la tendencia creciente de afectación de la enfermedad en los predios es leve, aunque con algunos datos atípicos, donde es alta. En otras palabras, la incidencia de la PC en esta zona es relativamente moderada; sin embargo, la tendencia de infección hace prever que a medida que haya un avance en el tiempo, la infestación puede llegar a ser alta.

Como se puede observar en la progresión de histogramas de las Figuras 10 y 11, los niveles de PC en las fincas de la zona empiezan a aumentar y se concentran en los valores medios de la enfermedad. Son pocas las que presentan muy bajos o muy altos niveles de infestación.

Entre enero y mayo de 2007 los valores se concentran principalmente en niveles bajos de incidencia, aunque se observa un pico de alta frecuencia que sugiere una posible bimodalidad. Tal comportamiento es más evidente de junio a septiembre, meses entre los cuales se diferencian dos picos de frecuencias, uno en el que la afectación de la PC es moderadamente

**Figura 9.** Histogramas de las incidencias mensuales de PC para el conjunto de datos Mira.

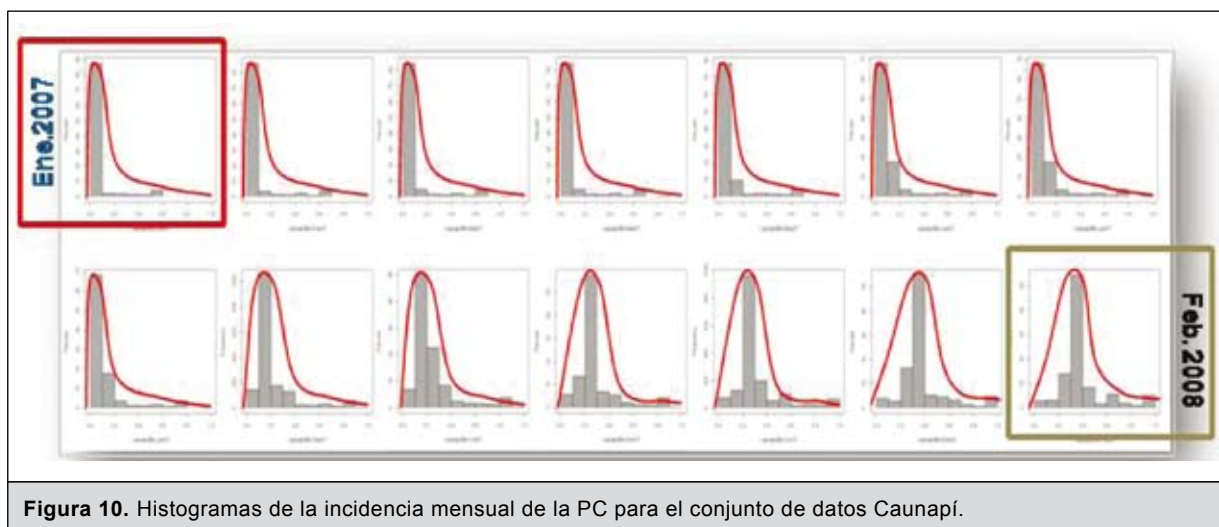


Figura 10. Histogramas de la incidencia mensual de la PC para el conjunto de datos Caunapí.

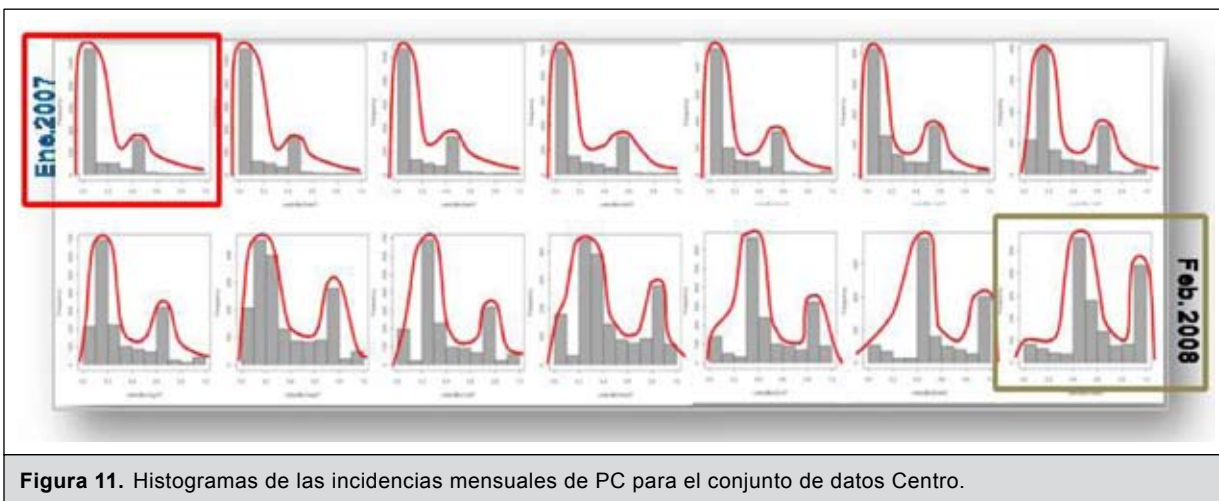


Figura 11. Histogramas de las incidencias mensuales de PC para el conjunto de datos Centro.

baja, es decir, un porcentaje significativo de fincas estuvo afectado incipientemente; y otro, con menor frecuencia, en el que es moderadamente alta en una cantidad considerable de fincas.

Desde octubre de 2007 hasta febrero de 2008, los valores de kurtosis cercanos a 2 y de coeficiente de asimetría ubicados en el intervalo $(-0,5, 0,5)$, señalan que existió una distribución de frecuencia bimodal simétrica, y que la alta variabilidad de los datos es indicativa de dos tipos de tendencias en las fincas: de incidencia media y de incidencia moderadamente alta.

Análisis de tendencias espaciales

Este análisis buscó encontrar la correlación existente entre la ubicación espacial del lote y el grado de

incidencia de la PC que se presenta en cada uno de los meses del período estudiado; en otras palabras, se intentó aislar la ubicación espacial de cada lote en los ejes coordenados (x, y) para compararlos con las incidencias de PC, y observar si algún tipo de desplazamiento de ésta se relaciona con ellos.

A continuación se presenta la subzona Mira con los diagramas de dispersión de los meses enero de 2007 y febrero de 2008, y su correspondiente análisis. Vale aclarar que en las otras dos zonas se aplicó la misma metodología.

Subzona Mira

En enero de 2007 los diagramas de dispersión (incidencia vs. coordenada norte; incidencia vs. coordenada este), indicaban que la PC se encontraba en valores bajos de incidencia; sin embargo, había una

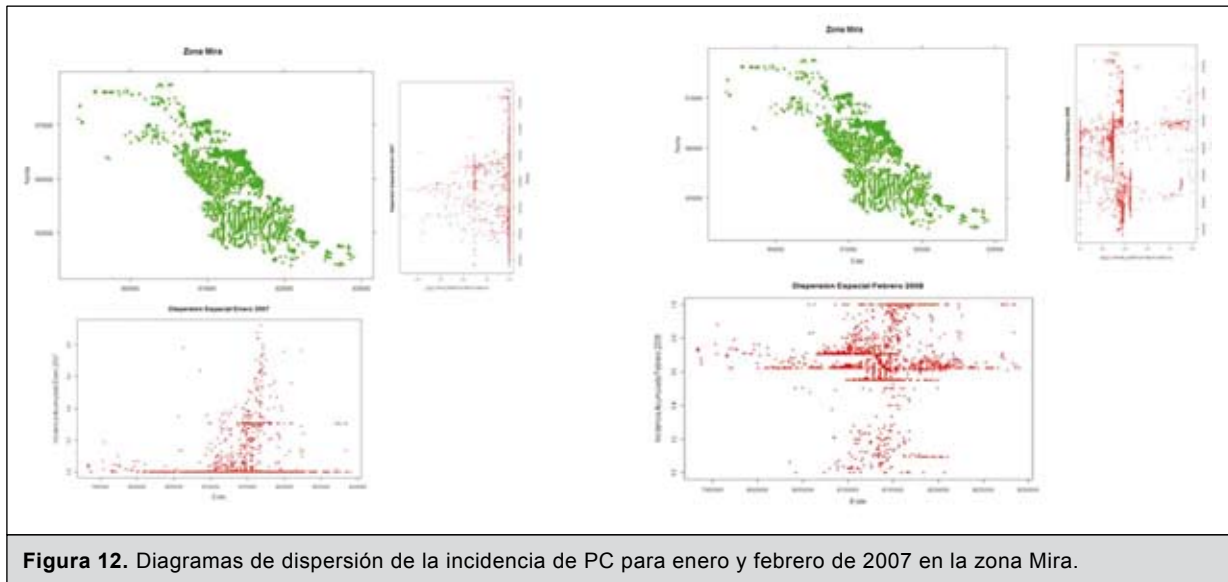


Figura 12. Diagramas de dispersión de la incidencia de PC para enero y febrero de 2007 en la zona Mira.

concentración con valores altos en el sector central de la zona (Figura 12).

Para febrero de 2008 las gráficas de dispersión mostraron un nivel alto de incidencia; la enfermedad se propagó de forma generalizada a lo largo de la zona, y solo quedó una pequeña proporción de lotes con valores bajos de incidencia en el sector central de la misma.

Análisis estadístico multidimensional

Definición de número de clases

Para la selección del número de grupos o clases en las cuales se tipifica el conjunto de datos, se tomó el mes con mayor coeficiente de variación dado que a su vez presenta la mayor dispersión de datos.

Subzona Mira

El mes con mayor coeficiente de variación fue enero de 2007, razón por la cual se le aplicó el método gráfico de dendrograma para establecer el número de grupos óptimos de tipificación (Figura 13).

Se realizó el corte del dendrograma a un nivel medio, de altura 0,2, con el fin de lograr un número de grupos representativos para describir el problema. Se trazó una línea vertical y se optó por definir seis grupos para el conjunto de datos.

El mismo procedimiento se aplicó para las otras dos zonas del estudio.

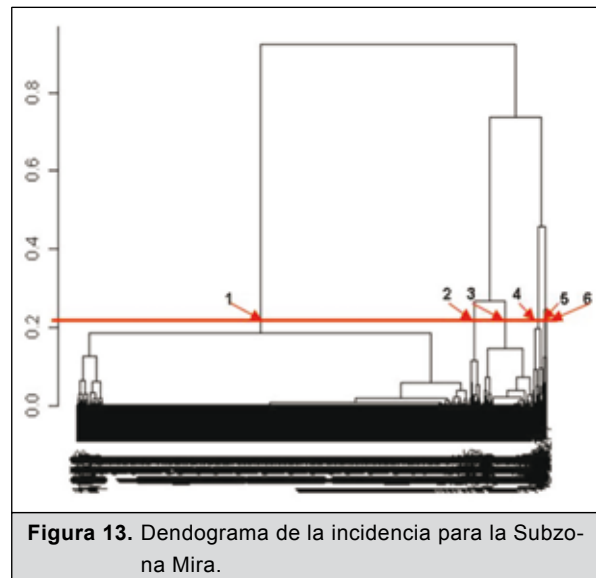


Figura 13. Dendrograma de la incidencia para la Subzona Mira.

Definición de la longitud de las clases

Con la ayuda de la función “clara” del software estadístico R se generó la distribución del conjunto de datos en el número de grupos definido con anterioridad, a los meses con los coeficientes de variación máximo y mínimo. Tal distribución proporciona las cotas mínimas y máximas de los grupos. El valor de las cotas de la clasificación es el resultado del promedio de las cotas resultantes del procedimiento anterior, con lo cual se garantizó una equidad de tipificación en el conjunto de datos.

Los resultados de las cotas de tipificación para las Subzonas Mira, Caunapí y Centro son presentados en las tablas 7, 8 y 9, respectivamente.

Evolución espacio-temporal de la incidencia de PC

En las Figuras 14 a 16 se presenta la integración de los datos de incidencia mensual acumulada y la representación espacial de los datos en los lotes por rangos de incidencia, mostrando colores amarillos claros para lotes con baja afectación y colores rojizos oscuros para predios con alta afectación.

En enero de 2007 (Figura 14) en la subzona Mira se calculaba una incidencia de PC de 87.812 palmas, en la Centro de 156.978 y en la Caunapí de alrededor de

97.248 palmas; esto es, un total de 342.038 palmas para todo el núcleo productivo.

Para febrero de 2007, el cálculo de palmas afectadas fue de 97.817 en la subzona Mira, 168.685 en la subzona Centro y 105.792 en la subzona Caunapí; tales cifras reflejan un aumento aproximado a 10.000 palmas con incidencia de PC, en cada una de las tres subzonas, esto es, de 372.294 palmas para todo el núcleo productivo.

En marzo de 2007 la cifra de incidencia de PC para las tres subzonas llegó a 403.459.

En julio de 2007 el cálculo de la incidencia de palmas afectadas fue de 382.356 para la subzona Mira,

Tabla 7. Cotas de tipificación para la subzona Mira

Grupo A	Mínimo de incidencia Ene/07	Máximo de incidencia Ene/07	Grupo B	Mínimo de incidencia Feb/08	Máximo de incidencia Feb/08	Promedio mínimo	Promedio máximo
1	0	0,0217	6	0	0,2861908	0	0,1539454
2	0,023	0,0979	7	0,3333	0,5826	0,17815	0,34025
3	0,1013	0,2237	5	0,5861	0,6604	0,3437	0,44205
4	0,2304	0,3662	2	0,6622	0,7661	0,4463	0,56615
5	0,3733	0,4966	1	0,7677	0,9125	0,5705	0,70455
6	0,5026	0,9218	4	0,9163	1	0,70945	0,9609

Tabla 8. Cotas de tipificación para la subzona Caunapí

Grupo A	Mínimo de incidencia Ene/07	Máximo de incidencia Ene/07	Grupo B	Mínimo de incidencia Feb/08	Máximo de incidencia Feb/08	Promedio mínimo	Promedio máximo
1	0	0,0067	7	0	0,1761	0	0,0914
7	0,0083	0,0298	1	0,1865	0,3016	0,0974	0,1657
3	0,0302	0,0625	3	0,3069	0,3937	0,16855	0,2281
4	0,063	0,1421	4	0,3948	0,5594	0,2289	0,35075
5	0,1462	0,2915	6	0,5615	0,7245	0,35385	0,508
2	0,2959	0,4805	5	0,7294	0,8741	0,51265	0,6773
6	0,4828	0,9652	2	0,8909	1	0,68685	0,9826

Tabla 9. Cotas de tipificación para la subzona Centro

Grupo A	Mínimo de incidencia Ene/07	Máximo de incidencia Ene/07	Grupo B	Mínimo de incidencia Feb/08	Máximo de incidencia Feb/08	Promedio mínimo	Promedio máximo
1	0	0,0328	2	0	0,1865	0	0,10965
2	0,0338	0,1086	1	0,1958	0,3959	0,1148	0,25225
3	0,1095	0,184	3	0,3996	0,5105	0,25455	0,34725
4	0,1857	0,2745	5	0,5117	0,5989	0,3487	0,4367
5	0,2766	0,3854	4	0,6	0,7541	0,4383	0,56975
6	0,394	0,6586	6	0,756	0,9266	0,575	0,7926
7	0,679	0,9266	7	0,9331	1	0,80605	0,9633

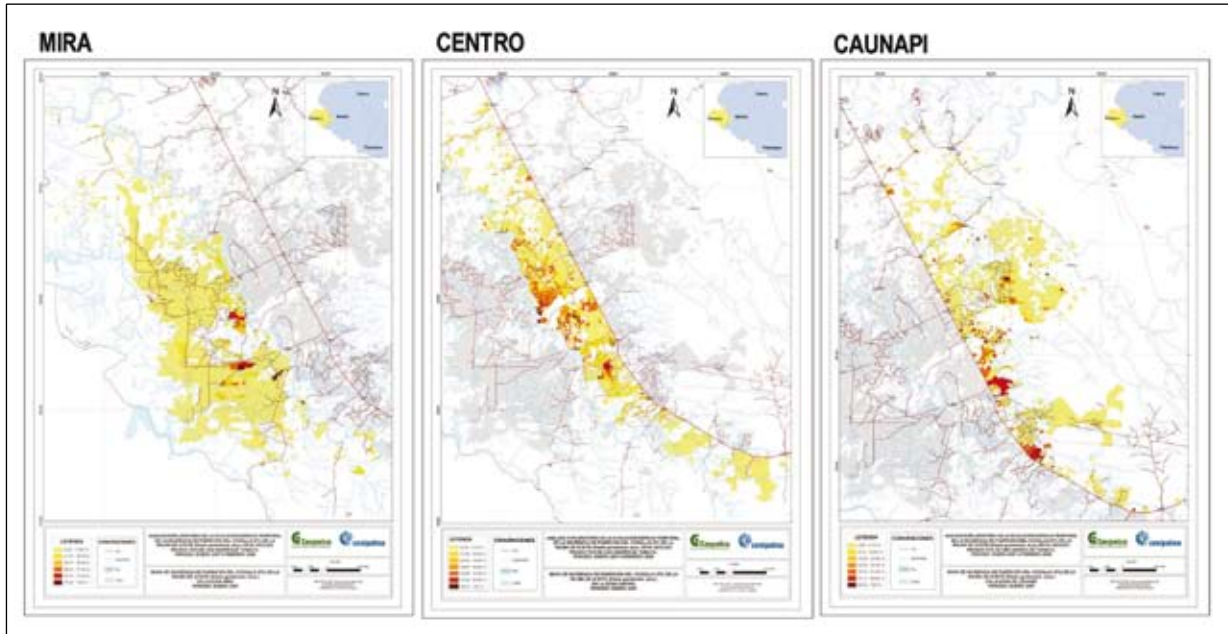


Figura 14. Representación espacial rangos de incidencia de la enfermedad de las tres subzonas en enero de 2007.

332.796 para la subzona Centro y 264.956 para Caunapí. Lo cual significó un aumento aproximado desde enero a julio de más de 650.000 palmas con PC, esto es, 980.106 palmas para todo el núcleo productivo.

Para enero de 2008, el cálculo era de 1.259.983 palmas para Mira, de 738.557 para Centro y de 599.648 para Caunapí. Ello reflejó un aumento con respecto al mes

anterior de más de 250.000 palmas con afectación para las tres subzonas, de lo cual se obtuvo un valor de 2.598.183 palmas para todo el núcleo productivo.

Las imágenes muestran que existen valores altos de incidencia de PC en la subzona Mira, donde los rangos se localizan en 6 y 7 (57,05-70,93%; 70,94-100%, respectivamente).

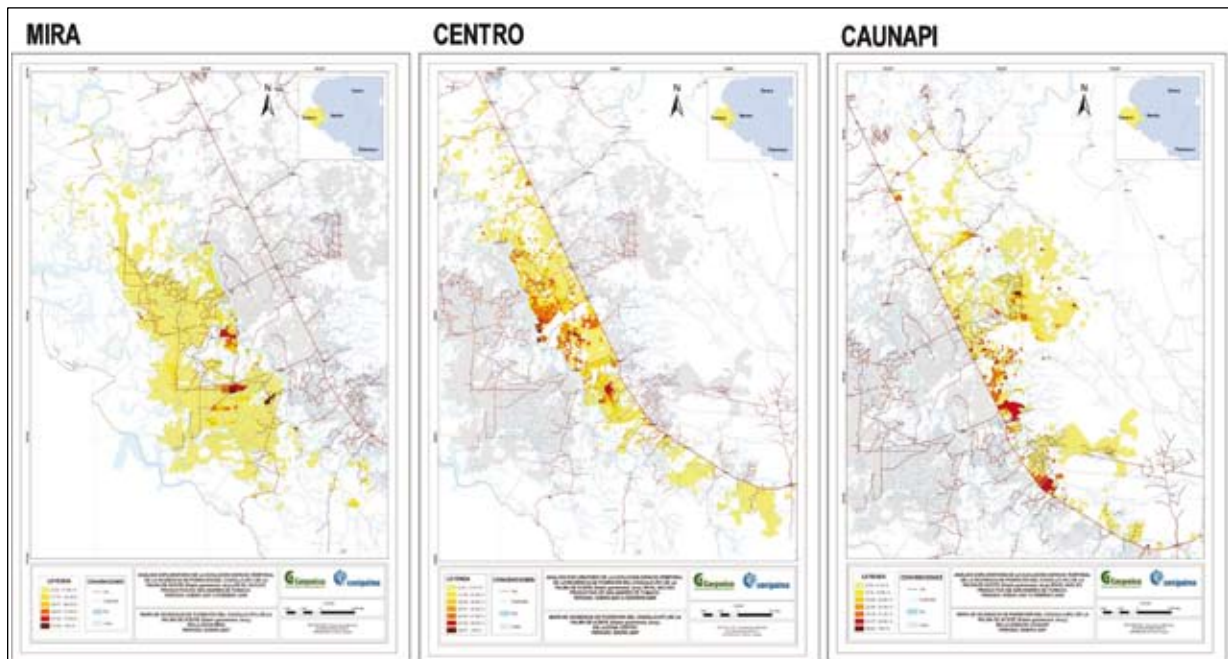


Figura 15. Representación espacial de rangos de incidencia de la enfermedad de las tres subzonas en julio de 2007.

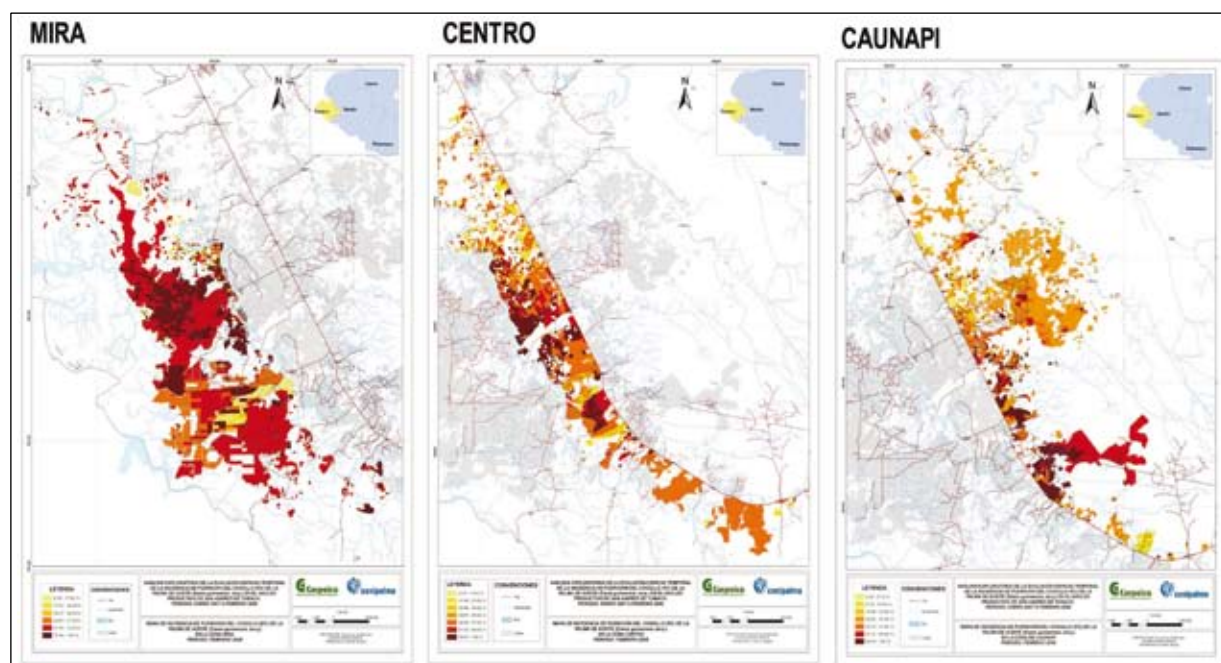


Figura 16. Representación espacial rangos de incidencia de la enfermedad de las tres subzonas en febrero de 2008.

En febrero de 2008 (Figura 16) se finalizó el intervalo de análisis con una incidencia de PC de 2.291.903 palmas.

Recomendaciones

- Ajustar el estudio incorporando los cambios en los datos causados por las actividades de erradicación o de renovación de las palmas enfermas realizadas mientras éste se encontraba en ejecución. Así podrán disminuirse las anomalías que generan tales actividades en la dinámica del fenómeno estudiado. Valdría la pena involucrar variables (vectores) externas al fenómeno propiamente de dinámica espacial y temporal –como las edáficas y las climáticas–, para analizar la forma como afectan la propagación y la incidencia de la PC.
- Realizar estudios que contemplen variables medioambientales (climáticas, edáficas e inundaciones) que permitan modelar algún tipo de correlación que influya en el comportamiento del fenómeno, no obstante que los resultados obtenidos reflejan en alto grado de confiabilidad su realidad.
- Continuar la retroalimentación de la información del estudio con la colaboración interinstitucional para mantenerlo vigente.

- Estudiar con mayor profundidad en la subzona Centro el comportamiento de la incidencia dada su bimodalidad, con el fin de identificar zonas de escape de la PC. La misma subzona Centro presenta una particularidad y es que, comparada con las otras, posee el mayor número de lotes (628), mientras Mira posee 292 y Caunapí 332. Al mismo tiempo se caracteriza por tener la mayor distribución de pequeños y medianos palmicultores.
- Socializar y divulgar el estudio.

Reflexiones

- Por medio del análisis exploratorio de los datos y la integración con la información espacial se logró determinar la distribución espacial de la PC en el núcleo productivo para cada una de las zonas, así como lograr la representación visual por medio de cartas de coropleta que permitieron presentar los resultados de manera más consistente y comprensible
- El cálculo de la incidencia mensual acumulada para el período de enero de 2007 a febrero de 2008 se realizó a partir de la aplicación de modelos de suavizamiento exponencial doble de Holt, obtenidos con las series de tiempo de información



mensual entre los meses de diciembre de 2005 hasta agosto de 2007 (fase 1 del proyecto)

- Con la utilización de modelos polinómicos simples (p. ej. $y = 0.0014x^3 + 0.0120x^2 - 0.2255x + 0.6378$) generados a partir de los datos suministrados por las diferentes empresas y agremiaciones del núcleo productivo de la zona, el trabajo de campo y el apoyo de entidades como la Alcaldía de Tumaco, el ICA y Cordeagropaz, se logró acceder a información que contribuyó a determinar las incidencias de la enfermedad para cada uno de los predios que integran las tres subzonas descritas (Mira, Centro y Caunapí). Con ello se pudo hacer una comparación temporal entre el número de palmas afectadas al principio y al final del intervalo analizado.
- Los modelos de dispersión permiten observar la tendencia espacial que está teniendo la PC en cada una de las subzonas de estudio: Mira, Centro y Caunapí.
- La crítica estadística buscó evaluar la calidad de la información disponible para el análisis, revisando posibles errores contenidos en los datos así como su atipicidad, y también la verificación de información faltante.
- Con el fin de caracterizar las distribuciones muestrales de las incidencias de PC, los datos se calcularon mediante estadísticas de resumen (promedio, mediana, coeficientes de variación, asimetría y kurtosis) y se construyeron los respectivos histogramas por subzona.
- Por medio del análisis de conglomerados, es decir, usando algoritmos de clasificación jerárquica, se agruparon los valores de incidencia de PC en meses de mayor variabilidad. El número de grupos formados se estableció con base en el análisis del dendograma, teniendo en cuenta que en principio se desconoce el número y la composición de los mismos y como resultado se debe tener una tipificación que maximice la homogeneidad dentro de cada grupo y la heterogeneidad entre ellos. Con esta clasificación se generaron representaciones espaciales que permitieron conocer la distribución de la incidencia en el tiempo y el espacio.
- El empleo de herramientas de tecnología de punta en materia de hardware y con software geográficos licenciado y libre, tanto para análisis vectorial como *raster*, sin duda contribuyó a que los resultados se ajusten a la realidad del fenómeno, y a que los diferentes actores involucrados en la problemática conozcan su estado actual de manera visual más fácil de interpretar.
- Los resultados obtenidos se constituyen en un importante instrumento de ayuda a la orientación de una política pública de renovación en palma de aceite, tanto para el gremio palmero, como para el Ministerio de Agricultura y Desarrollo Rural.
- En el caso de la subzona Mira, la cantidad de palmas sembradas para enero de 2007 fue de 2.249.999 con un total de 87.212 palmas con PC (el 3,87% de toda la subzona); para febrero de 2008 se presentó un incremento de palmas con PC en un porcentaje de 61,32% (1.379.052 palmas). Como resultado se obtuvo una incidencia de 1.466.864 palmas equivalente al 65,19% del total de la subzona.
- En enero de 2007 para la subzona Centro existían 1.497.045 palmas sembradas, de las cuales 156.978 tenían PC (el 10,48% de toda la subzona); para febrero de 2008 se presentó un incremento de palmas con PC en un porcentaje de 44,92%, equivalente a 672.525 palmas. Como resultado se obtuvo una incidencia de PC de 829.503 (55,40% del total de la subzona).
- En la subzona Caunapí, la cantidad de palmas sembradas en el mes de enero de 2007 ascendía a 1.493.428, con 6,51% (97.248) con PC; hacia el mes de febrero de 2008, el incremento fue de 38,05%, que equivale a 568.288 palmas. Como resultado se obtuvo una incidencia de 665.536 palmas con PC (el 44,56% del total de la subzona).
- Se determinó que la subzona con mayor incidencia fue Mira, la cual pese a tener valores bajos de incidencia al inicio del período de estudio, en los meses sucesivos llegó a constituirse en la de mayor crecimiento de la incidencia, con el 65,19% de palmas afectadas.
- El análisis de discontinuidad geográfica (río Mira y carretera Tumaco-Pasto) permitió disminuir la

brecha entre la percepción de la población palmera y los resultados del estudio; este hecho se refleja principalmente en la subzona Caunapí, que contempla el área costado oriental de la carretera Tumaco-Pasto, que se ha caracterizado por poseer menor grado de incidencia; así mismo, los medianos y pequeños palmicultores que en mayor proporción se han visto afectados por la PC son aquellos que se ubican alrededor de Imbilí, en la isla El Guavo, La Loma y Miraspalmas.

- El empleo de las imágenes TerrSAR-X modo Strip Map con resolución a 3 metros, de fecha febrero de 2008 (Infoterra-Geospacial), permitió precisar muchos de los lotes de pequeños productores

no asociados, al igual que actualizar y precisar información topográfica y de orografía de la zona de estudio.

Agradecimientos

Es pertinente agradecer la financiación del Ministerio de Agricultura y Desarrollo Rural (MADR), Fedepalma - Fondo de Fomento Palmero (FFP), Cenipalma, Cordeagropaz, ICA, Estación Experimental El Mira, alcaldía del municipio de Tumaco, así como la colaboración de empresas y agremiaciones palmeras de la zona, y a todos aquellos productores que hicieron posible el desarrollo de esta investigación.



Bibliografía

Angelfire.com. 2008. (en línea) <http://www.angelfire.com/biz2/palma-aceitera/infotecnica.html> (Consulta: 8 de julio) .

Fedepalma. 2007. Enfermedades en palma de aceite: un reto a la sostenibilidad de la agroindustria. *Palmas (Colombia)* 28 (1): Editorial.

Fedepalma. 2008. Historia de la palma de aceite (en línea) <http://www.fedepalma.org/palma.htm> (Consulta: 8 de julio)

Figuera, S. 2000. Introducción al análisis multivariante (en línea) <http://www.5campus.com/leccion/anamul> (Consulta: 27 de mayo de 2008) .

Figuera, S.; Gargallo, P. 2003. Análisis exploratorio de datos (en línea) <http://www.5campus.com/leccion/aed> (Consulta: 23 de mayo de 2008).

François, Jean-Christophe. 2004. Discontinuidad. En: Hypergeo, Análisis Espacial http://www.hypergeo.eu/article.php3?id_article=137 (consulta: 29 de mayo de 2008) .

Franqueville, H. 2001. La pudrición del cogollo de la palma aceitera en América Latina. Revisión preliminar de hechos y logros alcanza-

dos. Cirad-Departamento de Cultivos Perennes. En: <http://bio-nica.info/Biblioteca/Franqueville%202001%20pudricion%20cogollo%20palma%20aceitera.PDF>. (Consulta: 29 de mayo de 2008) .

Gómez, P.1995. Estado actual de la investigación sobre pudrición de cogollo. *Palmas (Colombia)* 16 (1): 76-82.

Ideam. 2008. Carta climatológica Estación Aeropuerto La Florida (Tumaco) (en línea) <http://bart.ideam.gov.co/cliciu/tumaco/temperatura.htm> (Consulta: 10 de julio)

Mingorance, F.; Minelli, F.; Le Du, H. 2004. Cultivo de la palma africana en el Chocó, legalidad ambiental, territorial y derechos humanos. *Human Rights Everywhere*. Diócesis de Quibdó.

Ospina, M.; Ochoa, D. 1996. La palma africana en Colombia. Fedepalma (Colombia). p. 13.

Rocha, P. 2007. Sanidad de la palma de aceite en Colombia: diagnóstico e investigación integral. Ponencia en: XXXV Congreso Nacional de Cultivadores de Palma de Aceite. Cenipalma.

The R Foundation for Statistical Computing. 2007. R - versión 2.7.0.

Universidad de Zaragoza. 2008. Etapas del análisis exploratorio de datos (en línea) <http://www.unizar.es/ice/rec-info/spss/Introduccion-AED.pdf> (Consulta: 21 de julio) .