

SELECCIÓN DE PROPIEDADES DEL SUELO ESPACIALMENTE RELACIONADAS CON PRODUCCIÓN EN EL CULTIVO DE BANANO

DARÍO ANTONIO CASTAÑEDA SÁNCHEZ^{1,2*}; DANIEL FRANCISCO JARAMILLO JARAMILLO¹
& JOSÉ MIGUEL COTES TORRES¹

Recibido: 25-07-13

Recibido con revisiones: 13-02-14

Aceptado: 15-02-14

RESUMEN

El objetivo del trabajo fue la selección de variables edáficas relacionadas con la producción de banano, así como la caracterización espacial de éstas. Se evaluaron 15 variables químicas, 18 físicas, y dos atributos de la planta que fueron cantidad de raíz funcional y peso del racimo, en 130 sitios, correspondientes a igual número de plantas. Los sitios fueron distribuidos en malla con celdas de 20 x 20 metros de separación. El experimento se realizó en tres lotes de la finca, en los cuales se ubicaron 48, 42 y 40 sitios de muestreo. En la selección de las variables relacionadas con la producción se usaron seis métodos de regresión. Las variables relacionadas se caracterizaron espacialmente. Se encontró que los métodos *Lasso*, *Lar* y *Forward Stagewise* fueron los más precisos, ya que presentaron los menores valores de raíz cuadrada media de los valores de predicción (RMSPD) y los más parsimoniosos al tomar el menor número de variables. Usando el método *Lasso* se obtuvo un modelo con siete variables. El diámetro ponderado medio en seco de agregados presentó una fuerte relación negativa con la producción (DPMS, $\beta = -0,96$). Similar comportamiento presentaron el aluminio, el cobre y el hierro, pero antagonísticos al del calcio, indicando desequilibrios entre iones. Las relaciones Ca/K y (Ca+Mg)/K estuvieron desbalanceadas. El modelo permitió extraer el componente de tendencia espacial de la producción de banano, así como determinar las propiedades con mayor potencial de manejo por sitio específico.

Palabras clave. Geoestadística, variograma, interpolación, *Musa acuminata* (AAA).

SELECTION OF SOILS PROPERTIES ASSOCIATED SPATIALLY WITH BANANA CROP PRODUCTION

ABSTRACT

This research was carried out to find both, a parsimonious and precise methodology for the selection of variables related to banana production and its spatial characterization. Fifteen chemical variables, eighteen physical variables and two attributes of the plant, quantity of functional root and the weight of the banana bunch were evaluated in 130 sites, corresponding to an equal number of plants, distributed on a 20 x 20 m-cell grid. The experiment was performed on three fields in the farm, each one with 48, 42 and 40 sample sites, respectively. Six regression methods were assessed at the field level and at the level of the entire sample. The related edaphic variables were characterized spatially. *Lasso*, *Lar* and *Forward Stagewise* were the most accurate methods, presenting the lowest values in the root mean square prediction difference (RMSPD) and using the lowest number of variables. The *Lasso* method produced a model with seven variables. The weighted average diameter of dry aggregates showed a strong negative relationship with the banana production variables (DPMS, $\beta = -0,96$). A similar trend was obtained with aluminum, copper, and iron, but were antagonistic with calcium, indicating unbalance between ions. The ratios Ca/K and (Ca+Mg)/K were imbalanced. The model also rendered the spatial trend component of banana production, and determined the properties with greater potential for site-specific management.

Key words. Geostatistics, variogram, krigging, *Musa acuminata* (AAA).

¹ Universidad Nacional de Colombia - Sede Medellín. Calle 59ª #63-20. Medellín - Colombia.

* Autor de contacto: dacasta4@unal.edu.co

INTRODUCCIÓN

El clima y el suelo determinan el éxito de la producción en las fincas bananeras. Según Vaquero (2003), la textura, la compactación y el drenaje son las propiedades físicas del suelo que tienen mayor influencia en el crecimiento y desarrollo de la planta de banano, ya que pueden limitar la profundidad efectiva, el suministro de agua y la aireación en la rizósfera. Las características de los suelos bajo alta saturación con agua, permanente o temporal, limitan el crecimiento y causan necrosis y pudrición de la raíz. Según Gauggel *et al.* (2003), también es reconocida la importancia de las relaciones entre las propiedades químicas del suelo y el sistema radical de la planta de banano.

A los productores les interesa conocer la eficiencia productiva de cada uno de los lotes de su finca, así como poder identificar las causas de las diferencias de productividad observadas entre ellos. La relación espacial entre la producción y las propiedades del suelo se ha explorado en cultivos como palma de aceite, mango, maíz, papa, frijol (Barrios y Florentino, 2009; Valbuena *et al.*, 2008; Miao *et al.*, 2006; Persson *et al.*, 2005), lo que ha permitido identificar las variables edáficas que han tenido mayor impacto en la producción de cada lote, y a través del análisis espacial se han podido delimitar zonas de manejo diferente para cada variable, estrategia conocida como manejo por sitio específico, que ha mostrado ser amigable con el ambiente puesto que permite optimizar el uso de los insumos (Ritter *et al.*, 2008). Como son muchas las variables que se relacionan con la producción de los cultivos y no todas tienen la misma relevancia en esa relación, en el presente trabajo se tuvo como propósito encontrar un método que permitiera seleccionar una pequeña cantidad de variables edáficas que representaran adecuadamente la producción de banano, de modo que se pudiera tener como filtro que ayude en la definición de prácticas de manejo por sitio específico en dicho cultivo.

MATERIALES Y MÉTODOS

La investigación se desarrolló en una parcela de 6 ha sembrada en cuadro a 2,5 m con el clon de banano Williams (*Cavendish* AAA) desde el año 2005, perteneciente al Campo Experimental de Augura (36 ha), situado al occidente del municipio de Carepa, departamento de Antioquia, Colombia. El campo se ubica en la llanura aluvial del piedemonte de la serranía de Abibe, tiene un relieve plano (pendiente <3%), una altura sobre el nivel del mar de 40 m y una temperatura promedio anual de 28 °C. Según Gutiérrez (2007), la parcela se asienta en tres consociaciones de suelos: la Fluventic

Eutrudepts fina, la Fluvaquentic Eutrudepts franca fina y la Vertic Endoaquepts franca fina. En la parcela experimental se seleccionaron 130 plantas (unidades productivas) en una malla regular con una cuadrícula de 20 x 20 m, distribuidas en tres lotes de la siguiente manera: 48 en el L3, 42 en el L4 y 40 en el L5 (Fig. 1A). Cada una de las plantas seleccionadas fue georreferenciada empleando un GPS TrimbleGeoXT estandarizado bajo el sistema de proyección UTM WGS84, zona 18 N. La población de plantas de la tercera generación en cada unidad productiva fue homogenizada seleccionando yemas de la primera corona del cormo con un tamaño menor a 5 cm de altura y eliminando las demás. La población de plantas seleccionadas se llevó hasta producción registrándose, para cada una de ellas, el peso total del racimo producido en kg (PR). Cuando floreció el 50% de las plantas, se tomaron muestras de suelo y se determinaron 15 propiedades químicas y 18 físicas. Además, se evaluó la cantidad de raíz funcional en cada planta. El muestreo y las condiciones en que se tomaron las muestras para cada variable, las determinaciones realizadas y los métodos utilizados, se presentan en la Tabla 1. Se evaluaron seis métodos para seleccionar las propiedades del suelo que se relacionan linealmente con el peso de los racimos de banano en cada lote, así como en toda el área de investigación, lo que implicó cuatro diferentes tamaños de muestra. Las propiedades del suelo se estandarizaron, para evitar efectos de magnitud debido a las diferentes unidades de medición. El modelo de regresión lineal usado cuenta con un vector de variables explicativas $X^T = (X_1, X_2, \dots, X_p)$, con las cuales se espera predecir la variable respuesta Y. El modelo de regresión tiene la forma:

$$\hat{Y} = \beta_0 + \sum_{j=1}^p X_j \beta_j$$

donde, \hat{Y} es la variable respuesta Y predicha, el β_0 y los β_j son el intercepto y los coeficientes de las variables explicativas respectivamente, los cuales son desconocidos y deben ser estimados, procedimiento detallado en Montgomery (2001). Los métodos de selección, del intercepto y coeficientes de las propiedades del suelo relacionadas con la producción, fueron (Hastie *et al.*, 2008):

Ajuste por Mínimos Cuadrados Ordinarios (OLS). El método se basa en seleccionar los coeficientes $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_p)^T$, que minimizan la suma de cuadrados de los residuales (RSS).

$$RSS = \sum_{i=1}^N (y_i - \hat{y}_i)^2 = \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p X_{ij} \beta_j)^2$$

donde, y_i es la i-ésima observación de la variable respuesta, \hat{y}_i es la i-ésima estimación de y, β_0 es el intercepto, x_{ij} es la i-ésima observación de la j-ésima variable explicativa, β_j es

Tabla 1. Características del muestreo y métodos de evaluación de propiedades físicas y químicas del suelo y de raíz funcional de las plantas de banano estudiadas en el campo experimental bananero, Carepa, Colombia.

Table 1. Soil characteristics and evaluation methods of physical and chemical soils properties and functional roots of bananas plants, studied on the banana experimental farm, Carepa, Colombia.

Propiedad	Método
Físicas	
Resistencia superficial a la penetración (kg cm^{-2}) determinada a 30, 60 y 100 cm de distancia de la planta: CP30, CP60, CP100, respectivamente.	Penetrómetro de bolsillo
Densidad aparente superficial (Mg m^{-3}), a 30 cm de la planta (Da).	Cilindro biselado
Índices de estabilidad estructural, evaluados sobre una muestra sin disturbar, tomada a 30 cm de la planta	
Diámetro ponderado medio, en mm, determinado en seco y húmedo (DPMS y DPMH).	
Índice de estructuración (%) determinado en seco y húmedo (IES, IEH).	Tamizado en seco y en agua según métodos descritos en JARAMILLO (2011)
Agregados finos ($< 0,5$ mm) en (%) determinados en seco (AFS) y húmedo (AFH).	
Agregados extremos (> 2 mm y $< 0,5$ mm de diámetro) (%) determinados en seco (AES) y en húmedo, (AEH).	
Índices de humedad, evaluados sobre una muestra tomada a 30 cm de la planta	
Retención de humedad gravimétrica a capacidad de campo (33 kPa), y en el punto de marchitez permanente (1500 kPa): H0.3, H15, respectivamente.	Desorción de humedad en platos y ollas de presión con oxígeno.
Textura , evaluada sobre una muestra compuesta de cuatro submuestras tomadas en forma ortogonal a 30 cm de la planta y entre 0 y 20 cm de profundidad. Contenido de arena (A), limo (L) y arcilla (Ar), en %. Arcilla dispersa (ArD) y coeficiente de dispersión (CD) expresadas en % [$CD = \text{ArD}/\text{Ar} * 100$].	La A, el L, y la Ar determinadas por el método de Bouyoucos (IGAC, 2006). La ArD, determinada por el método de la pipeta (JARAMILLO, 2011).
Químicas	
Evaluadas sobre una muestra compuesta de cuatro submuestras tomadas en forma ortogonal a 30 cm de la planta y entre 0 y 20 cm de profundidad.	
pH determinado en agua 1:1.	
Contenido de materia orgánica (MO) expresada en %.	Walkley-Black, (IGAC 2006).
Capacidad de intercambio catiónico efectiva (CICE) y a pH 7 (CIC7), expresadas en $\text{cmol}(+) \text{kg}^{-1}$ suelo.	CICE, método suma de cationes. CIC7, método del Acetato de amonio 1N y neutro (IGAC, 2006).
Contenidos intercambiables de Ca, Mg, K y Al, expresados en $\text{cmol}(+) \text{kg}^{-1}$ suelo.	Ca, Mg, K, extraídos con acetato de amonio 1M y neutro. Al, extraído con KCl, (IGAC, 2006).
Contenidos disponibles de P, S, Fe, Mn, Cu, Zn y B, expresadas en mg kg^{-1} suelo.	P: Bray II. S: fosfato monocalcico 0,008M. Fe, Mn, Cu, Zn: Olsen modificado. B: Agua caliente, (IGAC, 2006).
Raíz funcional (RF), expresada en porcentaje.	Método de la cajuela, descrito en ARAYA Y CALVO (2001).

el coeficiente de la j-esima variable explicativa y p: son las variables explicativas.

Selección por Pasos hacia Adelante y hacia Atrás (Forward and Back Ward Step-Wise), usando el Criterio de Información de Akaike (STEP AIC) para la selección del modelo. El método se fundamenta en el criterio de información de Akaike (AIC

por sus siglas en inglés), para modelos paramétricos ajustados y es definido por el R Core Team (2013), como:

$$AIC = -2 \log(L) + k \cdot e \cdot d \cdot f$$

donde, log: es el logaritmo, L: es la verosimilitud, k: es un

parámetro de ponderación y edf: grados de libertad del modelo. El procedimiento por pasos hacia adelante y hacia atrás consiste en estimar el criterio *AIC*, tanto para el modelo sin variables explicativas, es decir para el intercepto, como con todas ellas. En cada ciclo y de manera repetida se adiciona una variable y se quita una de los dos modelos anteriores respectivamente, estimando para cada nuevo modelo el criterio *AIC* y así sucesivamente en cada ciclo hasta encontrar el modelo con el menor *AIC*. Para mayor detalle acerca de la fundamentación estadística del criterio *AIC* consultar Miller (2002).

Método de regresión de Avance por Etapas (Forward Stagewise). El procedimiento inicia de manera similar al del método de selección por pasos hacia adelante, elaborando un modelo general con un intercepto ($\bar{y} = \beta_0$), y centrando los coeficientes de todas las variables explicativas en cero ($\beta_1, \beta_2, \dots, \beta_p = 0$). En cada ciclo, se construye un modelo de regresión lineal simple parcial entre los residuales y la variable explicativa (x_i) más correlacionada con éstos y se obtiene el coeficiente (β_j), el cual es sumado al respectivo coeficiente de esa variable en el modelo general. El proceso continúa hasta que ninguna de las variables se correlacione con los residuales. El procedimiento permite un número de ciclos mayor al número de variables explicativas, puesto que una variable explicativa puede presentar la mayor correlación con los residuales en más de un ciclo.

Tres métodos de encogimiento: Arista de regresión (Ridge Regresion), LASSO y Mínimo Ángulo de Regresión (LAR). El procedimiento en el método de encogimiento *Arista de regresión*, selecciona aquellos β s que minimizan la suma de cuadrados de los residuales, como en el método OLS, pero además, la suma de los cuadrados de los coeficientes (β), excluyendo el intercepto (β_0), debe ser menor que un parámetro de penalización (t). El procedimiento en cada ciclo hace que la magnitud de los coeficientes sea menor y tienda a cero. El método puede ser representado como:

$$\hat{\beta}_{\text{arista}} = \underset{\beta}{\text{Armin}} \left\{ \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p X_{ij} \beta_j)^2 \right\}$$

$$\text{Sujeto a que, } \sum_{j=1}^p \beta_j^2 \leq t$$

El método de encogimiento *LASSO*, es similar al de *Arista de regresión*. La solución de β_0 es \bar{y} . Luego, se ajusta un modelo sin el intercepto que simultáneamente minimice la suma de cuadrados de los residuales y haga que la suma de sus valores absolutos sea menor que un parámetro de penalización (t). El método se puede representar como:

$$\hat{\beta}_{\text{lasso}} = \underset{\beta}{\text{Armin}} \left\{ \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p X_{ij} \beta_j)^2 \right\}$$

$$\text{Sujeto a que, } \sum_{j=1}^p |\beta_j| \leq t$$

El método de encogimiento mínimo ángulo de regresión, inicia de manera similar al de avance por pasos hacia adelante, estimando el intercepto e igualando los coeficientes de las demás variables a cero. En el primer paso, encuentra la variable explicativa (x_i) más correlacionada con los residuales y se construye un modelo lineal simple. Se mueve el coeficiente entre 0 y el coeficiente (β_j) obtenido, hasta que alguna otra variable respuesta (x_k) presente la mayor correlación con los respectivos residuales y se construye un modelo lineal entre los residuales y las dos variables encontradas. En el siguiente paso se mueven los coeficientes desde cero hasta los respectivos β_j y β_k encontrados, hasta que alguna otra variable (x_l), presente la mayor correlación con los respectivos residuales. El proceso continúa hasta que todas las variables explicativas hayan entrado.

Como mejor método se seleccionó aquel que presentó la menor magnitud en la raíz cuadrada media de los errores de predicción (*RMSPD*) del peso de los racimos de banano, obtenidos por el procedimiento de validación cruzada dejando una observación por fuera (*leave-one-out*) y mediante el número promedio de variables explicativas (*NV*) implicadas en el modelo de regresión final (Dias & Krzanowski, 2003). El método elegido se usó para seleccionar las variables explicativas del suelo con mayor nivel de predicción del peso de los racimos.

Se realizó un análisis espacial a cada una de las variables explicativas seleccionadas. El método de mínimos cuadrados ponderados se utilizó para estimar los parámetros del variograma teórico con mejor ajuste al variograma experimental (Diggle & Ribeiro, 2007). Para cada propiedad física y química del suelo, seleccionada por su relación con el peso de los racimos de banano, se construyó, por interpolación, un mapa de requerimientos, teniendo en cuenta para su elaboración los variogramas ajustado y experimental y el punto crítico de decisión por encima del cual se encuentra el rango óptimo. En el caso de las propiedades químicas se empleó el límite inferior del rango óptimo propuesto por el Ministerio de Agricultura y Ganadería de Costa Rica, como el nivel crítico de decisión (Bertsch, 1986). En el caso de los nutrientes: calcio, magnesio, potasio y fósforo, para su análisis se tuvieron en cuenta los niveles propuestos por Espinosa y Mite (2002), para plátano, puesto que en la propuesta de Bertsch (1986) se usaron soluciones extractoras diferentes para estos nutrientes a las usadas en este trabajo. Para las propiedades físicas no se usó ningún valor como punto crítico de decisión, dado que éstos aún no se han establecido para el cultivo de banano.

Los análisis estadísticos se realizaron en el lenguaje y ambiente para cálculos estadísticos R (R Core Team, 2013). Se utilizaron dos paquetes adicionales instalados sobre el ambiente R, así: con el paquete *Monomvn* (Gramacy, 2009), se ajustaron los modelos de regresión según los diferentes métodos de selección y con el paquete *geoR* (Ribeiro & Diggle, 2001) se realizó todo el análisis espacial.

RESULTADOS Y DISCUSIÓN

Los gráficos de burbujas muestran la distribución espacial, en el campo experimental, del peso de los racimos, delimitados por los lotes (Fig. 1 A). Nótese que sobre la diagonal sur-norte del campo tienden a concentrarse las plantas que produjeron racimos con mayor peso y hacia los vértices laterales los de menor peso sugiriendo la presencia de tendencia espacial. Con respecto al peso promedio de los racimos, el lote L4 presentó el mayor promedio (32,71 kg) aunque no fue significativamente diferente de los lotes L3 (28,13 kg) y L5 (30,85 kg) (Fig. 1 B). Aunque en este caso no fue significativa, es común la existencia de sectores dentro de la finca con diferentes productividades (Serrano *et al.*, 2006). En la Tabla 2, se presentan los principales estadísticos descriptivos de las propiedades físicas y químicas del suelo y se define si existe diferencia estadísticamente significativa entre sus promedios, por lote.

En la Tabla 3, se presentan los resultados de la validación cruzada para los seis métodos usados en la selección de las variables relacionadas de manera lineal con el peso de los racimos de banano, con diferentes tamaños de muestra, así como el número de variables implicadas en el modelo final. El método de ajuste por mínimos cuadrados ordinarios tuvo una baja precisión en la predicción (*RMSPD*

entre 15,85 y 49,74 kg) para tamaños de muestra pequeños ($n \leq 47$). La selección de variables por pasos hacia adelante y hacia atrás, usando el Criterio de Información de Akaike como criterio de selección del modelo, tampoco aumentó la precisión de la predicción (*RMSPD* entre 16,45 y 47,38 kg), cuando la selección de variables se efectuó con tamaños de muestra pequeños ($n < 48$) e involucró entre 22 y 28 variables explicativas del peso del racimo de banano. El comportamiento en los dos casos anteriores puede deberse al alto número de variables explicativas usadas (35), valor muy cercano a los tamaños de muestra utilizados en cada lote (39, 41 y 47) y/o a la presencia de colinealidad entre variables explicativas, lo que puede inducir a que el método de selección nunca encuentre el mejor modelo y a que los modelos obtenidos tengan coeficientes de determinación altos (Hastie *et al.*, 2008; Miller, 2002). Los autores mencionados también comentan que un tamaño de muestra grande disminuye el error de la predicción por disminución de la varianza, con efecto considerable sobre el coeficiente de determinación. En efecto, el aumento del tamaño de muestra ($n = 129$) incrementó la precisión del método de ajuste por mínimos cuadrados ordinarios (*RMSPD* = 6,96 kg) y del método de selección de variables por pasos hacia adelante y hacia atrás, basado en el Criterio

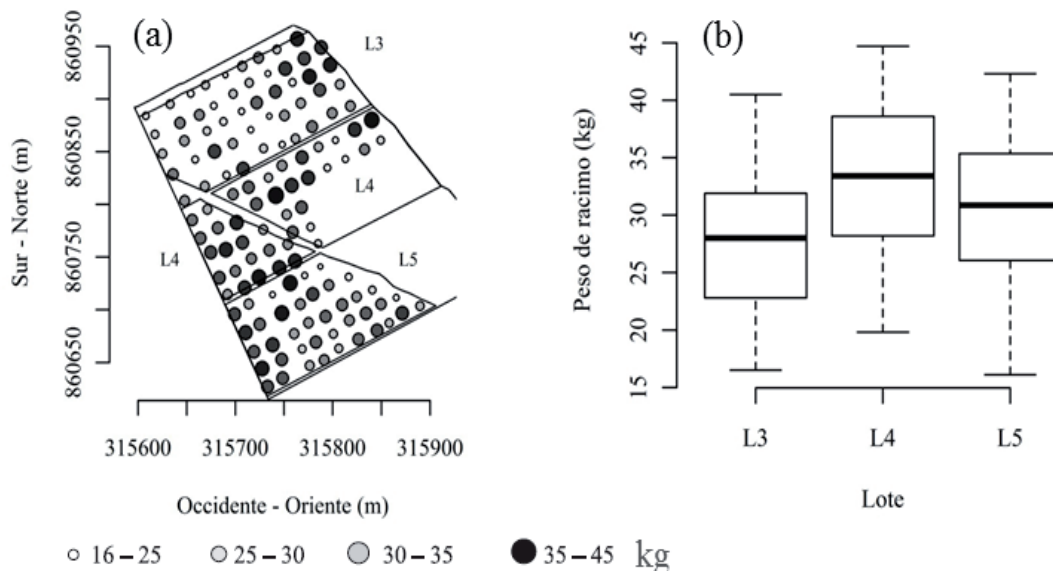


Figura 1. Comportamiento de la producción de banano en tres lotes del Campo experimental de Augura, Carepa, Colombia. (a). Gráfico de burbujas de la distribución espacial del peso de los racimos (kg). Las líneas delimitan los lotes. (b). Diagrama de cajas y bigotes para el peso de los racimos en los tres lotes (L3, L4 y L5).

Figure 1. Banana yield behavior in three lots from experimental field, Augura, Carepa, Colombia. (a). Bubble plot of spatial distribution of weight of banana bunches (kg). The lines delimit the lots. (b). Box and whisker plots of weight of banana bunch in the three lots (L3, L4 y L5).

Tabla 2. Estadísticos descriptivos de las propiedades del suelo y comparaciones pareadas de los promedios, en los lotes 3, 4 y 5 de un campo experimental de banano, Carepa, Colombia.

Table 2. Descriptive statistics of soil properties and paired comparisons of averages for plots 3, 4 and 5 in the banana experimental farm, Carepa, Colombia.

Variable	Unidad	Lote: L3				Lote: L4				Lote: L5				
		Prom ⁽¹⁾	SD ⁽²⁾	CV ⁽³⁾	DTK ⁽⁴⁾ p < 0,05	Prom	SD	CV	DTK p < 0,05	Prom	SD	CV	DTKp < 0,05	
pH		5,15	0,36	0,07	a ⁽⁵⁾	5,10	0,36	0,07	a	5,13	0,34	0,07	a	
MO ⁽⁶⁾	%	1,51	0,66	0,44	b	2,44	0,82	0,34	a	2,18	0,94	0,43	b	
Aluminio	cmol(+) ⁻¹ kg ⁻¹	1,09	0,72	0,66	a	1,12	0,89	0,79	a	0,96	0,73	0,76	a	
Calcio		16,49	2,88	0,17	a	14,01	3,33	0,24	b	15,50	4,25	0,27	ab	
Magnesio		5,49	0,86	0,16	a	5,22	0,65	0,13	a	5,40	0,73	0,13	a	
Potasio		0,36	0,10	0,29	a	0,37	0,10	0,28	a	0,38	0,14	0,37	a	
CICE ⁽⁷⁾		23,54	2,77	0,12	a	20,60	3,11	0,15	b	22,36	4,11	0,18	ab	
CIC7 ⁽⁸⁾		22,22	2,98	0,13	a	21,25	3,26	0,15	a	21,75	3,24	0,15	a	
Fósforo	mg kg ⁻¹	9,44	2,78	0,29	b	10,93	3,67	0,34	ab	12,55	4,88	0,39	a	
Azufre		13,48	7,09	0,53	a	12,67	7,43	0,59	a	16,57	11,34	0,68	a	
Hierro		139,9	46,62	0,33	a	163,48	42,06	0,26	a	149,00	39,35	0,26	a	
Manganeso		22,48	11,39	0,51	a	21,19	11,31	0,53	a	20,60	10,52	0,51	a	
Cobre		5,67	1,37	0,24	b	6,90	1,48	0,21	a	7,20	1,71	0,24	a	
Zinc		1,88	0,61	0,32	a	2,07	0,71	0,34	a	2,08	0,80	0,38	a	
Boro		0,26	0,14	0,53	ab	0,29	0,13	0,45	a	0,21	0,08	0,40	b	
Arena			15,04	3,48	0,23	a	15,52	4,00	0,26	a	14,90	3,92	0,26	a
Limo			44,83	3,37	0,08	a	45,29	3,53	0,08	a	44,73	3,10	0,07	a
Arcilla			39,54	4,21	0,11	a	39,19	3,59	0,09	a	39,73	5,72	0,14	a
ArD ⁽⁹⁾	%	5,17	1,78	0,34	a	3,77	1,90	0,50	b	4,01	2,18	0,54	b	
HO.3 ⁽¹⁰⁾		36,41	3,64	0,10	a	35,38	3,40	0,10	a	36,58	4,19	0,11	a	
H15 ⁽¹¹⁾		23,65	2,84	0,12	a	22,59	2,67	0,12	a	23,35	3,22	0,14	a	
Da ⁽¹²⁾	g cm ⁻³	1,23	0,10	0,08	b	1,24	0,11	0,08	ab	1,29	0,09	0,07	a	
CD ⁽¹³⁾	%	12,72	3,54	0,28	a	9,30	3,77	0,41	b	9,96	5,31	0,53	b	
CP30 ⁽¹⁴⁾	kg cm ⁻²	2,10	0,82	0,39	a	2,27	0,82	0,36	a	2,51	0,77	0,31	a	
CP60 ⁽¹⁵⁾		2,07	0,70	0,34	a	2,23	0,78	0,35	a	2,46	0,86	0,35	a	
CP100 ⁽¹⁶⁾		2,05	0,78	0,38	a	2,07	0,67	0,32	a	2,24	0,88	0,39	a	
DPMS ⁽¹⁷⁾	mm	4,70	0,26	0,05	a	3,97	0,59	0,15	b	4,66	0,36	0,08	ab	
AFS ⁽¹⁸⁾	%	2,03	1,64	0,81	b	3,72	2,37	0,64	a	3,51	2,57	0,73	a	
AES ⁽¹⁹⁾		83,10	4,55	0,05	a	82,70	5,15	0,06	a	79,40	5,55	0,07	b	
IES ⁽²⁰⁾		20,70	6,75	0,33	b	20,60	6,34	0,31	b	26,56	9,08	0,34	a	
DPMH ⁽²¹⁾	mm	1,99	0,79	0,40	b	2,47	0,68	0,27	a	1,71	0,82	0,48	b	
AFH ⁽²²⁾	%	44,94	16,37	0,36	a	35,88	12,49	0,35	b	52,64	21,06	0,40	a	
AEH ⁽²³⁾		79,50	3,73	0,05	ab	78,42	4,63	0,06	b	81,74	6,05	0,07	a	
IEH ⁽²⁴⁾		26,05	5,84	0,22	ab	27,97	7,92	0,28	a	22,98	8,93	0,39	b	
RF ⁽²⁵⁾		58,55	21,64	0,37	b	51,45	24,89	0,48	b	72,01	16,92	0,23	a	

⁽¹⁾Promedio. ⁽²⁾Desviación estándar. ⁽³⁾Coefficiente de variación. ⁽⁴⁾Comparación pareada múltiple de Dunnett-Tukey-Kramer, para tamaños de muestras diferentes (Matthew 2013). ⁽⁵⁾Diferencias significativas (p < 0,05) para promedios con letras diferentes por filas. ⁽⁶⁾Materia orgánica. ⁽⁷⁾Capacidad de intercambio catiónico efectiva. ⁽⁸⁾Capacidad de intercambio catiónico a pH 7. ⁽⁹⁾Arcilla dispersa. ⁽¹⁰⁾Retención de humedad a 33 kPa. ⁽¹¹⁾Retención de humedad a 1500kPa. ⁽¹²⁾Densidad aparente del suelo. ⁽¹³⁾Coefficiente de dispersión. ⁽¹⁴⁾Compactación del suelo a 30 cm de la planta. ⁽¹⁵⁾Compactación del suelo a 60 cm de la planta. ⁽¹⁶⁾Compactación del suelo a 100 cm de la planta. ⁽¹⁷⁾Diámetro ponderado medio de agregados en seco. ⁽¹⁸⁾Agregados finos del suelo en seco. ⁽¹⁹⁾Agregados extremos del suelo en seco. ⁽²⁰⁾Índice de estructuración del suelo en seco. ⁽²¹⁾Diámetro ponderado medio de agregados en húmedo. ⁽²²⁾Agregados finos en húmedo. ⁽²³⁾Agregados extremos en húmedo. ⁽²⁴⁾Índice de estructuración del suelo en húmedo. ⁽²⁵⁾Raíz funcional de la planta.

Tabla 3. Precisión de la predicción y número de propiedades del suelo relacionadas con producción de banano, usando de seis métodos de selección de variables, agrupadas por lotes, según división de la finca para administración y manejo agronómico, con diferentes tamaños de muestra. Variables evaluadas en un campo experimental de banano, Municipio de Carepa, Colombia.

Table 3. Precision of prediction and number of soil properties related with banana yield, using six selection methods of variables, grouped by fields with different sample sizes. Variables evaluated in the bananas experimental field, Municipality of Carepa Colombia.

Lote	n ⁽¹⁾	OLS ⁽²⁾		ESTEP AIC ⁽³⁾		LASSO ⁽⁴⁾			LAR ⁽⁵⁾			RIDGE ⁽⁶⁾			FORWARD STAGewise ⁽⁷⁾		
		RMSPD ⁽⁸⁾	NV ⁽⁹⁾	RMSPD	NV	RMSPD	NV	λ_m ⁽¹⁰⁾	RMSPD	NV	λ_m	RMSPD	NV	λ_m	RMSPD	NV	λ_m
L3	48	49,74	35	47,38	22	6,28	3	10,78	6,28	3	10,8	6,42	35	248,2	6,28	3	10,8
L4	42	1585	35	16,45	28	6,91	4	12,28	6,9	2	13,5	6,54	35	2,15	6,73	5	12,5
L5	40	36,16	35	32,99	27	6,83	1	12,28	6,94	1	13,5	6,46	35	2,15	6,77	0	12,5
C.Exp. ⁽¹¹⁾	130	6,96	35	6,87	11	6,51	7	12,41	6,56	5	13,8	6,34	35	107,3	6,53	5	13,6

⁽¹⁾Tamaño de muestra. ⁽²⁾Ajuste por mínimos cuadrados ordinarios. ⁽³⁾Método de selección de variables basado en el criterio de Akaike. ⁽⁴⁾Método de encogimiento LASSO. ⁽⁵⁾Mínimo ángulo de regresión. ⁽⁶⁾Regresión ridge. ⁽⁷⁾Avance por etapas en la selección de variables. ⁽⁸⁾Raíz cuadrada media de los errores de predicción del peso de los racimos de banano, obtenidos por el procedimiento de validación cruzada, dejando uno por fuera, expresada en kg. ⁽⁹⁾Número de variables explicativas promedio de las validaciones cruzadas o el modelo de regresión, después del proceso de selección.

⁽¹⁰⁾Parámetro de penalización ⁽¹¹⁾Campo experimental.

de Información de Akaike ($RMSPD = 6,87$ kg), reduciendo, en este último, a once el número de variables explicativas requeridas.

Las estrategias usadas por los métodos de encogimiento (*Ridge*, *LASSO* y *Lar*), de las variables explicativas, así como por el de avance por etapas, presentaron buena precisión en la predicción, tanto con muestras pequeñas como grandes. Aunque el método *Ridge* pertenece al grupo de los métodos de encogimiento, la estrategia de penalización sobre los coeficientes no es eficiente para seleccionarlos. Los métodos *LASSO*, *Lar* y *Forward Stagewise* cumplen con los requisitos inicialmente buscados de ser parsimoniosos, es decir, utilizar un número pequeño de propiedades en el modelo final, y de tener buena precisión en la predicción de la producción, tanto a nivel de lote como de todo el campo experimental. La precisión de la predicción con el método *LASSO*, por lotes, osciló entre 6,28 y 6,91 kg involucrando entre una y cuatro propiedades de predicción por lote, mientras que para todo el área del campo experimental, la precisión fue de 6,51 kg e involucró siete propiedades. El método *LASSO* evidencia un efecto de lote; sin embargo, la estrategia no perdió precisión cuando se trabajó a nivel de toda el área de experimentación, con un aumento de dos variables con respecto al lote L4 que fue el que más propiedades involucró (4) para la predicción del peso de los racimos. La selección de variables relacionadas con el peso de racimos de banano se realizó finalmente con el método *LASSO* porque además de ser un método parsimonioso con bajo *RMSPD*, Hastie *et al.* (2008), lo reportan como un método con una alta habilidad para recuperar el modelo correcto, incluso, en casos donde el número de variables

explicativas es mayor que el tamaño de la muestra. Por otra parte, como se indicó anteriormente, el objetivo de este trabajo fue el de encontrar un modelo con buena capacidad de predicción, basado en pocas variables, de modo que se pueda implementar el manejo agronómico por sitio específico de aquellas propiedades relacionadas con producción.

Debido a que el error de predicción en cada uno de los lotes fue similar, la selección de variables se realizó tomando toda el área del campo experimental con un tamaño de muestra total $n = 130$, lo cual es ventajoso para el diseño e implementación de la estrategia de manejo por sitio específico. En la Tabla 4 se detallan los coeficientes de regresión de las variables relacionadas con el peso de los racimos, usando el método de encogimiento *LASSO* para seleccionarlas. El modelo final está compuesto por siete variables y la variable con mayor peso fue el diámetro ponderado medio de los agregados cribados en seco (DPMS) con un coeficiente de regresión = -0,96. La relación inversa con el peso de racimos, puede estar mostrando que hay un deterioro importante de la estructura del suelo y confirma lo que puede observarse en la Tabla 2, al comparar la cantidad de agregados finos que tienen los suelos en seco (AFS) y la cantidad de éstos que se producen cuando el suelo es humedecido (AFH). Estos cambios tan drásticos en la agregación del suelo, en presencia de arcillas expansibles, podría estar dañando raíces y afectando la producción, al cohesionarse el suelo cuando pierde humedad (Vaquero, 2003).

Otra variable de alto peso en el modelo fue el aluminio ($\beta = 0,71$). Según el criterio de Bertsch (1986), el aluminio

Tabla 4. Coeficientes de regresión de variables explicativas del suelo relacionadas con el peso de racimos de banano, seleccionadas a través del método de encogimiento LASSO⁽¹⁾.

Table 4. Regression coefficients of soil explanatory variables related with banana bunch weight selected by shrinkage method Lasso.

Variable	Coefficientes de regresión
Intercepto	30,41
Arena	-0,13
Aluminio	0,71
Calcio	-0,13
Hierro	0,3
Cobre	0,27
Boro	0,09
DPMS ⁽²⁾	-0,96
ML ⁽³⁾	37,9
SDEV ⁽⁴⁾	6,40

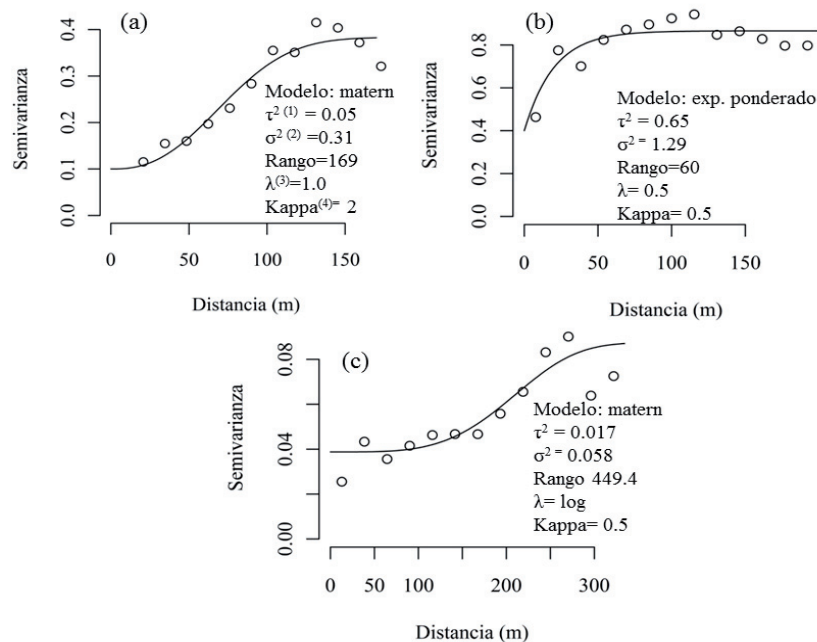
⁽¹⁾La función usó un parámetro $\lambda = 11,62$, obtenido por validación cruzada empleando el método dejando uno por fuera (LOO). ⁽²⁾Diámetro ponderado medio de agregados en seco. ⁽³⁾Máxima Verosimilitud de la matriz de covarianza de los residuales. ⁽⁴⁾Desviación estándar de la estimación.

intercambiable en banano debe estar por debajo de 0,3 cmol(+) kg⁻¹ suelo. Toda el área experimental presentó un contenido de Al intercambiable por encima de este nivel crítico, pero ello no representó un problema para el cultivo,

lo que puede estar mostrando que el nivel crítico establecido por Bertsch (1986) no es adecuado para los suelos de Urabá, muy diferentes a los Andisoles que se utilizan en Costa Rica para cultivar banano.

La relación inversa entre el calcio y la producción está, en cierta manera, confirmando la relación obtenida con el Al y puede estar indicando problemas de antagonismo con las otras bases del suelo, puesto que las relaciones promedias Ca/K (45,52) y (Ca+Mg)/K (61,34), doblaron los valores sugeridos por López y Espinosa (1995), 17-25, para el Ca/K; 20-30, para (Ca+Mg)/K, como relaciones ideales para el cultivo de banano. Los altos contenidos de calcio intercambiable en estos suelos muestran un desbalance entre este nutriente y los demás cationes principalmente con el potasio (Fageria, 2001).

Según criterios de Cambardella *et al.* (1994), el diámetro ponderado medio con cribado en seco presentó una fuerte dependencia espacial ($\tau^2/\sigma^2 \leq 25\%$), mientras que la del aluminio y el cobre fue moderada ($25 < \tau^2/\sigma^2 < 75$). La dependencia espacial del calcio fue débil y prácticamente se puede considerar como una propiedad con variabilidad aleatoria, así como la del hierro, la de la arena y la del boro. En la Figura 2 aparecen los gráficos de los variogramas



⁽¹⁾Pepita. ⁽²⁾Meseta. ⁽³⁾Transformación Box-Cox para normalización de variables. ⁽⁴⁾Parámetro de forma del variograma.

Figura 2. Variogramas experimentales y ajustados de las propiedades del suelo relacionadas con producción que presentaron dependencia espacial moderada a fuerte. (a). Diámetro ponderado medio del suelo en seco. (b). Aluminio intercambiable. (c). Cobre disponible.

Figure 2. Experimental and adjusted variograms of soil properties related with banana yield that evidenced moderate to strong spatial dependence. (a). Average weighted diameter, evaluated in dry soil. (b). Exchangeable aluminum. (c). Copper available.

experimentales, y su correspondiente variograma teórico ajustado, para las propiedades del suelo relacionadas con la producción y que presentaron dependencia espacial moderada y fuerte. Los mapas de la Figura 3 muestran la distribución espacial de las variables del suelo relacionadas con el peso de los racimos de banano que presentaron alta y moderada dependencia espacial. Las áreas con los agregados de menor tamaño ($< 4,2$ mm) se encuentran en el lote L4, en el cual se registró el mayor peso promedio de racimos de banano (33 kg), mientras que los lotes L3 y L5 presentaron los agregados con tamaños superiores a 4,4 mm (Fig. 3a) y en estos lotes se registró la menor producción. El mapa de la Figura 3b muestra el estado del aluminio con respecto al punto crítico de $0,3$ cmol (+) kg^{-1} de suelo, obtenido por la diferencia entre el valor predicho por inter-

polación y el valor crítico. Se evidencia que los niveles de aluminio intercambiable sobrepasan en la mayor parte del área el nivel crítico. De manera similar, la Figura 3c, muestra deficiencias de cobre en toda el área experimental, usando como nivel crítico el punto medio del rango óptimo (8 cmol (+) kg^{-1} suelo). Sólo una pequeña área ubicada en lote L5 está por encima de este nivel, en el resto del campo sería necesario aumentar entre $0,5$ y 3 cmol (+) kg^{-1} de suelo para tener un nivel de cobre adecuado. El gráfico de burbujas mostrado en la Figura 3d, para los residuales del peso de los racimos, obtenidos por la diferencia entre el peso de los racimos y el valor predicho estimado con el modelo usando el método de selección *LASSO*. En éste se aprecia que los residuales se distribuyen aleatoriamente, eliminándose la tendencia espacial, observada inicialmente en la Fi-

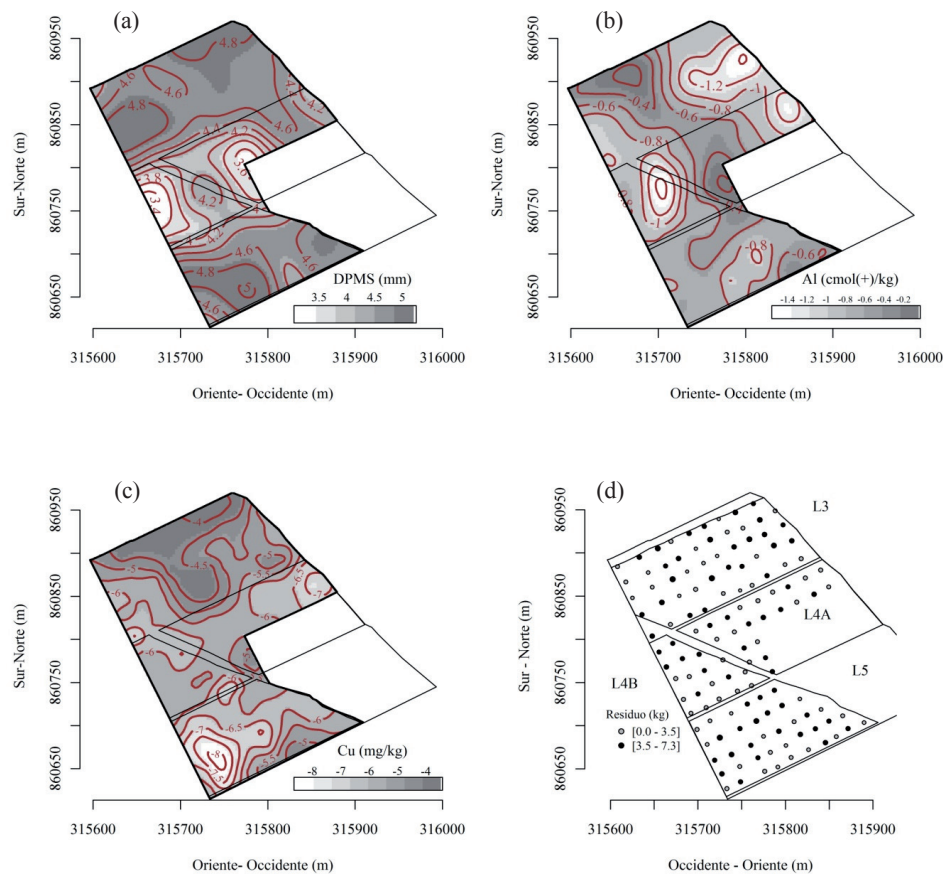


Figura 3. Mapas de la distribución de las propiedades del suelo relacionadas con el peso del racimo de banano. (a). Mapa del diámetro ponderado medio de agregados. (b). Mapa de exceso de aluminio intercambiable tomando como punto crítico $0,3$ cmol (+) kg^{-1} suelo para el cultivo de banano. (c). Mapa de requerimientos de cobre, tomando como punto crítico, el punto central del rango óptimo (8 mg kg^{-1}). (d). Gráfico de burbujas para los residuales del peso de racimos.

Figure 3. Distribution of soil properties maps related with banana bunch weight. (a). Map of weighted average aggregate diameter. (b). Excess aluminum exchangeable map, taking as critical limit $0,3$ cmol (+) kg^{-1} soil for banana growing. (c). Copper requirements map, taking as critical, the central point of the optimal range (8 mg kg^{-1}). (d). Bubble plot for residuals of bunch weights.

gura 1a, de plantas con mayor peso de racimos hacia la diagonal central del lote. Además, las propiedades del suelo implicadas en el modelo encontrado permiten explicar y extraer esa tendencia y son las que mayor potencial tienen en la implementación de un programa de manejo por sitio específico.

CONCLUSIÓN

Fue posible modelar la producción de banano en función de unas pocas propiedades físicas y químicas del suelo, seleccionadas con tres métodos: el Lasso, el de la Arista de regresión (*LAR*) y el de avance por etapas (Forward stage-wise), conservando la precisión de la predicción de la producción, tanto para muestras con tamaños pequeños como grandes. Usando el método *LASSO* se obtuvo un modelo con siete variables, de las cuales el diámetro ponderado medio en seco de agregados (*DPMS*) fue la variable más relacionada con el peso de los racimos. Las variables aluminio, cobre, hierro presentaron una relación positiva con el peso de los racimos, mientras que con el calcio hubo una relación inversa que sugiere, probablemente, un desbalance entre cationes. El análisis espacial permitió detectar estructuración espacial en las propiedades del suelo diámetro ponderado medio, aluminio intercambiable y cobre disponible, así como, identificar las variables con mayor potencial para la implementación en un programa de manejo por sitio específico, a través del cual se puedan establecer áreas para un manejo homogéneo de éstas.

AGRADECIMIENTOS

Los autores expresan su agradecimiento al Ministerio de Agricultura de Colombia y a la Asociación de Bananeros de Colombia, por la cofinanciación y apoyo logístico en la realización del presente proyecto.

BIBLIOGRAFÍA

- Araya, M & C Calvo. 2001. Cantidad de raíces en los 10 cantones productores de Costa Rica. *Corbana* 27: 47-64.
- Barrios, MR & A Florentino. 2009. Variabilidad espacial de las propiedades físicas de dos suelos cultivados con palma aceitera en el estado Monagas, Venezuela. *Revista UDO Agrícola* 9(4): 912-924.
- Bertsch, F. 1986. Manual para interpretar la fertilidad de los suelos. Oficina de publicaciones, Universidad de Costa Rica. San José de Costa Rica. 82 pp.
- Cambardella, CA; TB Moorman; TB Parkin & DL Karlen. 1994. Field-scale variability of soil properties in central Iowa soils. *SSSAJ* 58: 501-1511.
- Dias, C & W Krzanowski. 2003. Model selection and cross validation in additive main effect and multiplicative interaction models. *Crop Science* 4: 865-873.
- Diggle, P & PJ Ribeiro. 2007. Model-based geostatistics. Springer Science. New York. 228 pp.
- Espinosa, J & F Mite. 2002. Estado actual y futuro de la nutrición y fertilización del banano. *Informaciones Agonómicas* (48): 4-9.
- Fageria, VD. 2001. Nutrient interactions in crop plants. *Journal of plant nutrition* 24 (8): 1269-1290.
- Gauggel, C; D Moran & E Guardian. 2003. Interrelations between the soil chemical properties and the banana plant root systems. In: Turner D. & Rosales, F. (eds). Proceedings Banana root system: toward a better understanding for its productive management, Pp.13-22. Inibap, Musalac, Corbana. San José de Costa Rica. 261 pp.
- Gramacy, RB. 2009. Monomvn: Estimation for multivariate normal and Student-t data with monotone missingness. R package version 1.7-4.
- Gutiérrez, JC. 2007. Estudio detallado de suelos y clasificación de tierras con fines de riego del Campo Experimental de Augura, Carepa Antioquia. AUGURA. Carepa, Colombia. 28 pp.
- Hastie, T; R Tibshirani & J Friedman. 2008. The elements of statistical learning: data mining, inference, and prediction. 2 ed. Springer Series in statistics. New York. 745 pp.
- Instituto Geográfico Agustín Codazzi (IGAC). 2006. Métodos analíticos del laboratorio de suelos. 6 ed. Imprenta Nacional de Colombia. Bogotá. 648 pp.
- Jaramillo, D. 2011. El suelo: origen, propiedades, espacialidad. Universidad Nacional de Colombia. Medellín, Colombia. 553 pp.
- López, M & J Espinosa. 1995. Manual de nutrición y fertilización del banano, una visión práctica del manejo de la fertilización. Instituto de la potasa y el Fósforo. Quito, Ecuador. 81pp.
- Mattew, KL. 2013. DTK: Dunnett-Tukey-Kramer pairwise multiple comparison test adjusted for unequal variances and unequal sample sizes. R package version 3.5.
- Miao, Y; DJ Mulla & PC Robert. 2006. Spatial variability of soil properties, corn quality and yield in two Illinois, USA fields: Implications for precision corn management. *Precision Agric.* 7: 5-20.
- Miller, A. 2002. Subset selection in regression. 2. ed. New York: Chapman & Hall/CRC eds, 234 p.
- Montgomery, D. 2001. Design and Analysis of experiments. 5. ed. New York: John Wiley & Sons, INC, 684 p.
- Persson, A; P Pilesjö & L Eklundh. 2005. Spatial influence of topographical factors on yield of potato (*Solanum tuberosum* L.) in Central Sweden. *Precision Agric.* 6: 341-357.
- R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Ribeiro, PJ & PD Diggle. 2001. geor: A package for geostatistical analysis. *R-NEWS* 1(2): 15-18.
- Ritter, C; D Dicke; M Weis; H Oebel; HP Piepho; A Büchse & R Gerhards. 2008. An on-farm approach to quantify yield variation and to derive decision rules for site-specific weed management. *Precision Agric.* 9: 133-146.

- Serrano, E; J Sandoval; L Pocasangre; F Rosales & E Delgado. 2006. Importancia de los indicadores físico-químicos, en la calidad del suelo para la producción sustentable del banano en Costa Rica. *En: Soprano, E; F Tcacenco; L Lichtember & C Silva (eds). Memorias XVII Reunión Internacional Acorbat. Vol 1. Pp. 207-221. Acorbat. Joinville, Brasil, 415 pp.*
- Valbuena, CA; ML Martínez & HR. 2008. Giraldo. Variabilidad espacial del suelo y su relación con el rendimiento de mango (*Mangifera indica* L.). *Rev. Bras. Frutic. Jaboticabal-SP. 30(4): 1146-1151.*
- Vaquero, R. 2003. Soil physical properties and banana root growth. *In: Turner, D & F Rosales (eds). Proceedings Banana root system: toward a better understanding for its productive management. Pp. 125-131. Inibap, Musalac, Corbana. San José de Costa Rica. 261 pp.*