



Maestría en Matemáticas Aplicada

CARACTERIZACIÓN DE LA PRODUCTIVIDAD
DEL BANANO: UN ENFOQUE BASADO EN
VARIABLES QUÍMICAS DEL SUELO

Autor:

Juan Diego Valenzuela Cobos

Director:

David Christopher Siddons

Cuenca – Ecuador

2024

DEDICATORIA

A mis Padres: MILTON y CINTHIA, a mi abuelita
MARÍA, y a mi tío DAVID que me acompaña desde
el cielo.

AGRADECIMIENTO

A mis hermanos.

A mi Director: David Christopher Siddons.

A la Familia Garcés Moncayo.

RESUMEN:
CARACTERIZACIÓN DE LA PRODUCTIVIDAD DEL BANANO: UN ENFOQUE BASADO EN VARIABLES QUÍMICAS DEL SUELO

Los recursos agrícolas limitados imponen la necesidad a los productores de banano de optimizar sus insumos. En este estudio, se propone un modelo de regresión múltiple para caracterizar la productividad del cultivo de banano, basado en las características químicas del suelo. Se seleccionaron seis campos bananeros que exhibían diferentes niveles de producción. Para cada campo, se clasificaron dos muestras representativas, que incluyeron datos biométricos para calcular el índice de productividad. La variable de respuesta fue el índice de productividad, mientras que se utilizaron 17 variables independientes para medir las propiedades químicas del suelo, como el pH, la conductividad eléctrica, la capacidad de intercambio catiónico, la materia orgánica y diversos macro y microelementos. Se empleó el método de regresión por pasos para estimar diez modelos de regresión lineal múltiple. Posteriormente, se llevó a cabo un análisis de desempeño para evaluar la capacidad predictiva de los modelos, tanto para el quinto como para el décimo modelo respectivamente. El modelo final incluyó 5 variables cuantitativas y dos de clasificación, las cuales explican la variabilidad con un R^2 de 0.92 y un F de 59.84 (valor $p < 2.2e-16$).

Palabras clave: Banano, Regresión lineal múltiple, Regresión por pasos, Suelo, Productividad.

ABSTRACT:
CHARACTERISING BANANA PRODUCTIVITY: AN APPROACH BASED ON SOIL CHEMISTRY VARIABLES

Limited agricultural resources necessitate banana producers to optimize their inputs. In this study, a multiple regression model is proposed to characterize banana crop productivity based on soil chemical properties. Six banana fields exhibiting different production levels were selected. For each field, two representative samples were classified, including biometric data to calculate the productivity index. The response variable was the productivity index, while 17 independent variables were used to measure soil chemical properties such as pH, electrical conductivity, cation exchange capacity, organic matter, and various macro and microelements. The stepwise regression method was employed to estimate ten multiple linear regression models. Subsequently, a performance analysis was conducted to evaluate the predictive capacity of the models, for both the fifth and tenth models, respectively. The final model included 5 quantitative variables and two categorical ones, which explain the variability with an R^2 of 0.92 and an F of 59.84 (p-value $< 2.2e-16$).

Key words: Banana, Multiple linear regression, Stepwise regression, Soil, Productivity.

ÍNDICE

INTRODUCCIÓN.....	1
1.1 Problemática	1
1.2 Estado del arte y Marco Teórico.....	1
MÉTODO.....	3
2.1 Descripción de los datos	4
2.2 Procedimiento para el muestreo.....	4
2.3 Datos biométricos e índice de productividad.....	5
2.4 Análisis estadístico	5
RESULTADOS	8
3.1 Índice de productividad.....	8
3.2 Regresión lineal múltiple	11
3.3 Estimación del modelo por regresión por pasos	12
DISCUSIÓN Y CONCLUSIONES	15
REFERENCIAS	16

ÍNDICE DE TABLAS Y FIGURAS

Figura 1. Proyección de las variables del índice de productividad	9
Figura 2. Representación de las plantaciones en los dos primeros planos factoriales ...	10
Figura 3. Predicción del modelo 5.....	14
Figura 4. Predicción del modelo final	14
Tabla 1 Análisis de multicolinealidad e independencia de las principales variables explicativas.....	10
Tabla 2. Resumen estadístico de las principales variables explicativas	11
Tabla 3 Coeficientes del modelo	12
Tabla 4. Evaluación de los modelos de Regresión	12
Tabla 5. Error de predicción de los modelos de regresión	13

INTRODUCCIÓN

1.1 Problemática

La conjunción de la crisis económica post pandémica representa una problemática crítica para el sector bananero en Ecuador. La incapacidad de acceder a insumos agrícolas esenciales y la proyección de pérdida de vastas extensiones de cultivos intensifican la urgencia de una evaluación integral de las variables que afectan la productividad del suelo. Este análisis profundo es fundamental para orientar estrategias que optimicen recursos y permitan superar desafíos específicos, asegurando la resiliencia y continuidad del sector en el contexto nacional e internacional. El suelo de cultivo, determinante para el éxito de las plantaciones de banano, está influenciado por la interacción de variables físicas, biológicas y químicas.

1.2 Estado del arte y Marco Teórico

El banano es una fruta cultivada en las regiones tropicales del mundo, considerada como un alimento primordial de exportación que sostiene la economía y la seguridad alimentaria de un país (Galarza Suárez, 2019). Entre los principales países exportadores de banano en la última década se encuentran Ecuador, Filipinas y Costa Rica, mientras que los principales importadores son Bélgica, Alemania y Estados Unidos (Cedeño, 2023; León Ajila et al., 2023).

En el caso de Ecuador, el cultivo de banano se produce en abundancia, con una superficie de siembra de cerca de 230000 ha, concentradas en las regiones del Guayas, los Ríos y el Oro principalmente (Navarro, 2023). El rendimiento promedio de banano en Ecuador corresponde a 5 t/ha/año, lo cual se considera relativamente bajo a diferencia de otros países que reportan hasta 20 t/ha/año (Ministerio de comercio exterior, 2017). El bajo rendimiento del cultivo depende de diversos factores bióticos, abióticos y tecnológicos, por un limitado manejo en el control de plagas, riego, fertilización y control de los factores internos o externos que puedan afectar el cultivo como las condiciones del suelo (Barrezueta-Unda et al., 2023).

El cultivo convencional de banano es un factor desencadenante de la degradación del suelo, por el empleo de prácticas agrícolas inapropiadas que afectan directamente en su fertilidad natural, lo que conlleva al empleo intensivo de agroquímicos para maximizar los rendimientos en las áreas de cultivo (Barrezueta-Unda et al., 2023; Iori et al., 2020). Dentro de las prácticas agrícolas empleadas en los cultivos de banano en Ecuador se encuentran los monocultivos, labranza, irrigación y fertilización química. En estudios previos sobre cultivo de banano se ha demostrado que existe una relación directa entre la pérdida de calidad del suelo y la reducción de la productividad del cultivo por el manejo de prácticas agrícolas inadecuadas (Clercx & Huyghe, 2013; Segura et al., 2015). En el caso de otros tipos de cultivos, se han explorado diferentes variables como las actividades enzimáticas, de biomasa microbiana, abundancia bacteriana, pH, conductividad eléctrica, contenido de agua, entre otros, y la manera en que se relacionan con la calidad del suelo y del fruto (Bonilla-Bedoya et al., 2023). Los estudios reportados hasta el momento miden el potencial productivo de los suelos bananeros acorde a sus propiedades físicas, topografía, geomorfología, condiciones climáticas, entre otros (Segura et al., 2015).

En la búsqueda constante de métodos efectivos para mejorar la productividad agrícola, la aplicación de modelos estadísticos se ha convertido en una herramienta valiosa para comprender las complejas interacciones entre las características químicas del suelo y el rendimiento de los cultivos (Patrick et al., 2023). Entre estos modelos, la regresión múltiple se presenta como una herramienta robusta para analizar cómo diversas variables independientes, en este caso, las propiedades químicas del suelo, influyen en una variable dependiente clave, como el índice de productividad (Olivares, 2023).

El cultivo sostenible y eficiente demanda un conocimiento detallado de los factores que afectan el rendimiento de los cultivos, y las propiedades químicas del suelo desempeñan un papel fundamental en este contexto (Olivares et al., 2020). La utilización de múltiples variables en un modelo de regresión permite capturar la complejidad de estas relaciones, ofreciendo una visión más completa y precisa de cómo los componentes químicos del suelo pueden incidir en la productividad de los cultivos.

En este contexto, este estudio se enfoca en la presentación de un modelo de regresión múltiple que emplea diversas características químicas del suelo, con el objetivo de estimar el impacto de estas variables en el índice de productividad. A través de este enfoque, se busca proporcionar información valiosa para agricultores, investigadores y

tomadores de decisiones, contribuyendo así a la optimización de prácticas agrícolas y al fomento de sistemas de cultivo más eficientes y sostenibles.

1.2.1 Pregunta de investigación

Es posible caracterizar las áreas con diferente índice de productividad bananera en la zona 5 en Ecuador a partir de un índice derivado de un modelo matemático que abarque las propiedades químicas del suelo.

1.2.2 Objetivos

1.2.2.1 General

Presentar un modelo de regresión múltiple que caracterice la productividad del cultivo de banano en función de las características químicas del suelo de cultivo.

1.2.2.2 Específicos

Caracterizar las principales variables químicas del suelo de cultivo de banano correlacionadas con la productividad bananera en la zona 5 en Ecuador.

Determinar el índice de productividad agrícola del banano en las zonas de estudio a partir de datos biométricos de las plantas en fase de cosecha.

Estimar la validez del modelo seleccionado mediante la aplicación de un modelo alternativo (SARIMA).

MÉTODO

El presente estudio es de naturaleza cuantitativa, el alcance del estudio es la caracterización de los niveles de productividad de las áreas bananeras a partir de las propiedades fisicoquímicas del suelo. Este tiene un enfoque cuantitativo y correlacional, con un diseño transversal para asegurar la variabilidad de los datos en una escala de tiempo única.

2.1 Descripción de los datos

Se seleccionaron 7 campos bananeros ubicados en la zona 5 del Ecuador, de las cuales se nombraron en forma de secuencia alfanumérica (A1 al A7). Los suelos predominantes en esta área poseen gran cantidad de materia orgánica y humedad relativa media. Para la selección de la muestra se siguieron los lineamientos propuestos por Rosales et al. (2008). En este sentido, se escogieron áreas de alta y baja productividad basándose en la información recuperada de una evaluación preliminar de diferentes variables del suelo y a partir de registros disponibles por los agricultores de cada área. Los niveles de productividad se clasificaron en 4 categorías como alto-alto (≥ 40 t/ha/año), alto-bajo (35– 40 t/ha/año), bajo-bajo (≤ 30 t/año ha/año) y bajo-alto (30–35 t/ha/año).

2.2 Procedimiento para el muestreo

En cada sitio de muestreo se realizarán pozos de suelo (60 cm de ancho, 60 cm de largo y 60 cm de profundidad) para caracterizar los suelos en estudio siguiendo la metodología de Rosales et al. (2008). En las parcelas de diferente productividad de cada campo bananero se desarrollarán 4 pozos y se tomarán 25 muestras, en las cuáles se determinarán las diferentes propiedades químicas del suelo: pH (1:2,5), conductividad eléctrica (CE) (dS/m), capacidad de intercambio catiónico en el complejo de intercambio (ECEC) (cmol/kg), materia orgánica (MO) (%), nitrógeno total (N) (%), carbono total (C) (%), fósforo (P) (ppm), potasio (K) (cmol /kg), calcio (Ca) (cmol/kg), magnesio (Mg) (cmol/kg), azufre (S) (ppm), hierro (Fe) (ppm), zinc (Zn) (ppm), cobre (Cu), manganeso (Mn), boro (B)(ppm) y sodio (Na) (ppm). Acorde con la metodología dispuesta por la USDA (1995), el pH se medirá a partir del método basado en el multiparamétrico, la conductividad eléctrica a partir de un método basado en el conductímetro y la materia orgánica siguiendo el método del carbono orgánico oxidable por Walkley y Black. La capacidad de intercambio catiónico se podrá medir a partir de la metodología del acetato de amonio (Zhang & Gong, 2012). La cantidad de nitrógeno, fósforo, potasio, calcio, magnesio, azufre, hierro, zinc, cobre, manganeso, boro y sodio se obtendrá a partir de la metodología de NTC ISO/IEC 17025(ISO, 2018). La toma de medición se hará por triplicado.

2.3 Datos biométricos e índice de productividad

El índice de productividad (IP) se desarrollará a partir de datos biométricos de productividad que se calcularán del promedio de los datos muestreados de características de plantas con racimos próximos a ser cosechados (10 a 13 semanas de edad) de cuatro parcelas representativas: circunferencia del pseudotallo de la planta madre a 1 m de altura (CP, cm), número de manos por racimos (Nr, n) y la altura de la planta sucesora (AP, cm). Para obtener el índice, se aplicará un análisis de componentes principales (ACP) siguiendo la metodología aplicada por (Olivares et al., 2020). El ACP busca reducir la dimensionalidad de los datos recopilados. La variable de síntesis para el análisis será la combinación lineal de las variables biométricas, en la que los coeficientes representan las cargas de las variables PC1. A partir de un bosque aleatorio se seleccionan las variables más relevantes, a partir de métodos como la estimación progresiva y la prueba de multicolinealidad.

2.4 Análisis estadístico

2.4.1 Modelo de regresión lineal múltiple

En este trabajo, el modelo de regresión muestra una relación entre el rendimiento del cultivo de banano, denotado por la variable respuesta Y y diecisiete variables explicativas: pH (X_1), conductividad eléctrica (X_2), capacidad de intercambio catiónico en el complejo de intercambio (X_3), materia orgánica (X_4), nitrógeno total (X_5), carbono total (X_6), fósforo (X_7), potasio (X_8), calcio (X_9), magnesio (X_{10}), azufre (X_{11}), hierro (X_{12}), zinc (X_{13}), cobre (X_{14}), manganeso (X_{15}), boro (X_{16}) y sodio (X_{17}) (Bhausheb et al., 2023). La ecuación de regresión poblacional, en particular, describe la conexión real entre las variables explicativas y la variable de respuesta. Sin embargo, como la ecuación de regresión poblacional sigue siendo desconocida, debemos estimarla con base en datos muestreados (Hanson, 2010).

Consideremos una muestra de “n” observaciones, cada una de las cuales contiene valores para la variable de respuesta Y y “p” variables explicativas X_i . Podemos representar los valores de i observaciones como $Y_i, X_{i1}, X_{i2}, \dots, X_{ip}$. Así, la ecuación de regresión múltiple para estos valores viene dada por: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \epsilon_i$, donde Y_i representa el valor de la variable de respuesta para las i

observaciones, y $(X_{i1}, X_{i2}, \dots, X_{ip})$ representa los valores de las variables explicativas de las i observaciones. Los coeficientes del modelo de regresión se denotan por $\beta_0, \beta_1, \beta_2, \dots, \beta_p$, y el término ε representa el término de error para la i ésima observación (Sagamiko et al., 2020).

Si tenemos más puntos de datos (n) que las variables explicativas (p), formando un sistema sobre determinado con ecuaciones linealmente dependientes, podemos representar el i ésima observación de la variable X_j como X_{ij} , dónde $j = 1, 2, \dots, p$ y $i = 1, 2, \dots, n$. En este caso, el modelo poblacional para todas las observaciones de la muestra se puede expresar como el siguiente sistema de ecuaciones (Hanson, 2010; Sagamiko et al., 2020):

$$\begin{cases} Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \dots + \beta_p X_{1p} + \varepsilon_1 \\ Y_2 = \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + \dots + \beta_p X_{2p} + \varepsilon_2 \\ \vdots \\ Y_n = \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + \dots + \beta_p X_{np} + \varepsilon_n \end{cases}$$

El sistema de las ecuaciones. Se puede representar en notación matricial de la siguiente manera:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1p} \\ 1 & X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

El objetivo principal del análisis de regresión es seleccionar las variables explicativas que tienen un impacto significativo en la productividad. A la luz del supuesto de que las variables respuesta y explicativas tienen una conexión lineal, podemos expresar la ecuación matemáticamente como:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon$$

Dónde $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ son los coeficientes o parámetros asociados a cada variable explicativa, y ε representa el término de error o residual, capturando la variabilidad en el rendimiento del cultivo que no es explicada por el modelo (Rathod & Mishra, 2018).

2.4.2 Arima estacional (SARIMA)

ARIMA, uno de los modelos de series temporales más populares y eficaces, es uno de los clásicos. El modelo ARIMA ha ganado considerable popularidad debido a sus características estadísticas lineales y al enfoque Box-Jenkins comúnmente utilizado para la creación de modelos creado por Box y Jenkins en la década de 1970. La forma estándar del modelo ARIMA se escribe entonces como ARIMA (p, d, q) donde las letras p representa el orden de términos autorregresivo, d para el orden de los términos de diferenciación, y q para el orden de los plazos de media móvil (Box et al., 2015). Matemáticamente, el ARIMA (p, d, q). El modelo se puede expresar como:

$$\phi_p(B)(1-B)^d X_t = \mu + \theta_q(B)\varepsilon_t$$

Donde $\phi_p(B)$ Representa el operador de orden autorregresivo (AR). $p(1-B)^d$ representa el operador diferenciador, donde d representa el orden de diferenciación, X_t representar la variable de serie temporal en el momento t, cuál es la variable que se está modelando, μ es un término constante en la ecuación, representa cualquier componente determinista o compensación en la serie de tiempo, $\theta_q(B)$ representa (MA) el operador de orden de media móvil q, y ε_t es el término de error en el momento t, que denota el componente aleatorio o inexplicable de la serie temporal (Arunraj et al., 2016).

El modelo ARIMA se puede ampliar como SARIMA (p, d, q) (P, D, Q)_s para adaptarse a las variaciones estacionales, donde s es un término que considera la duración del período estacional. El modelo SARIMA se puede representar como:

$$\phi(B)\Phi_p(B^S)(1-B)^d(1-B^S)^D X_t = \theta_q(B)\Theta_Q(B^S)\varepsilon_t$$

Dónde $\phi(B)$ representan (AR) el operador autorregresivo estacional de orden p, $\Phi_p(B)$ representa (MA) el operador de orden de media móvil estacional q, $(1-B)$ representa el operador de diferenciación aplicado d veces, $(1-B^S)^D$ denota el operador de diferenciación estacional aplicado D veces, y S representan la duración estacional (digamos, s=4 en datos trimestrales, y s=12 en datos mensuales) (Neog et al., 2022). Dado que SARIMAX (p, d, q) (P, D, Q)_s modelo, donde (X) es el vector de variables externas, las técnicas de regresión lineal múltiple se utilizan para modelar las variables externas. En este estudio, podemos expresar matemáticamente un modelo de regresión múltiple como:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + w_t$$

Dónde $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ y β_5 son los coeficientes o parámetros asociados a cada variable explicativa, w_t representa el término de error o residual, capturando la variabilidad en el rendimiento del cultivo que no es explicada por el modelo (Arunraj et al., 2016). El término de error w_t se puede expresar en forma de modelo SARIMA como:

$$w_t = \frac{\theta_q(B)\Theta_Q(B^S)}{\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D} \varepsilon_t$$

De esta ecuación se deriva a la siguiente:

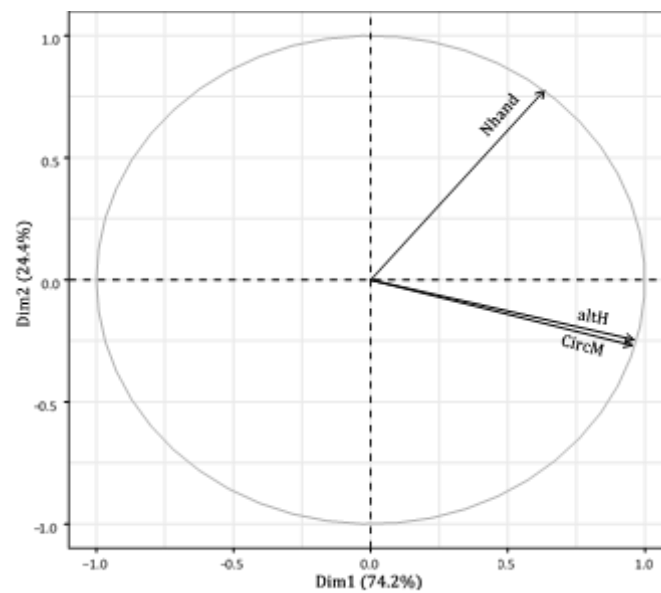
$$Y_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \dots + \frac{\theta_q(B)\Theta_Q(B^S)}{\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D} \varepsilon_t$$

RESULTADOS

3.1 Índice de productividad

La mayor parte de la variación de los datos biométricos, con un total de 98,61% se relaciona con los datos biométricos. La primera dimensión está relacionada con las variables de altura de la planta y circunferencia del pseudotallo, por el contrario, la dimensión 2 está relacionado con el número de manos. Para este estudio, se seleccionó únicamente la primera dimensión puesto que explica en un 74,21% la variabilidad total de los datos, a diferencia de la dimensión dos que aporta con un 24,4%, entonces, el IP propuesto para este estudio dependerá de la magnitud y dirección de las variables (vectores) representados (Figura 1).

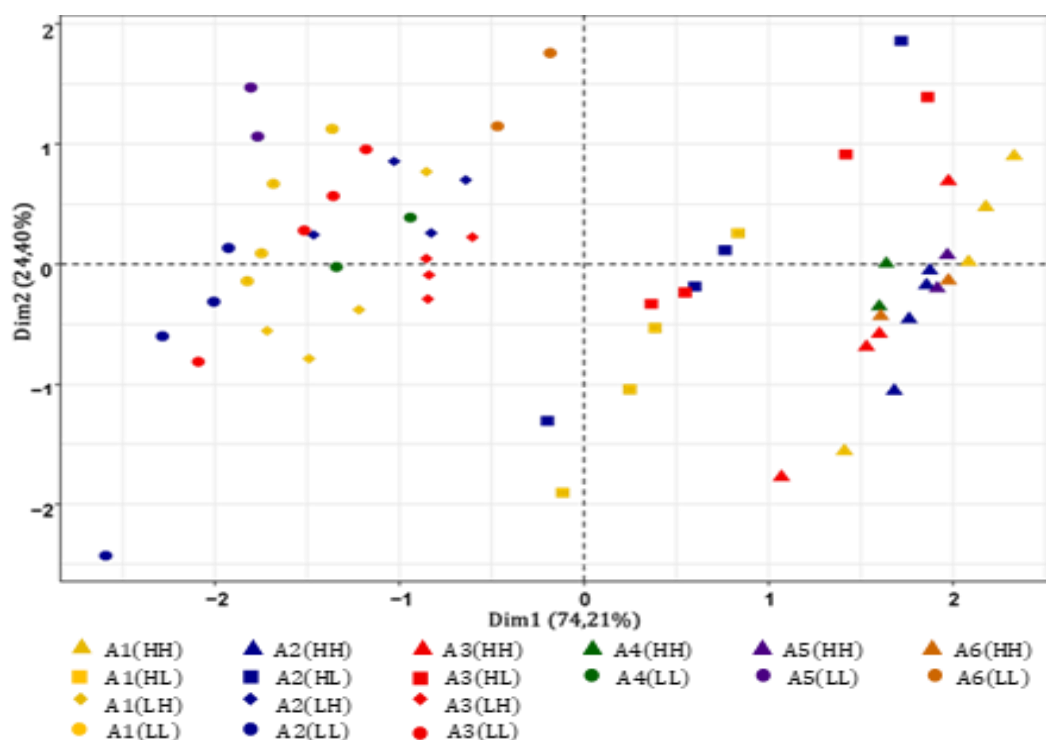
Figura 1. Proyección de las variables del índice de productividad



Nota. Nhand, Número de manos; altH, Altura de la planta, CircM, circunferencia del pseudortallo

Por otro lado, es evidente como categorías de igual producción se superponen entre sí, en este sentido, las categorías HH tienen una alta relación con los valores más altos de altura, circunferencia y números de manos de banana (Figura 2). En general, las zonas con alta productividad se encuentran del lado derecho de los planos factoriales con valores de IP más alto.

Figura 2. Representación de las plantaciones en los dos primeros planos factoriales



Nota. Las representaciones bivariadas corresponden a la producción promedio de frutos de las parcelas evaluadas. (HH, Alto - Alto rendimiento. ≥ 40 t/ha/año); HL, alto - bajo rendimiento (35–40 t/ha/año); LL, Bajo - Bajo rendimiento (≤ 30 t/ha/año); LH, bajo - alto rendimiento (30–35 t/ha/año)) dentro de las dimensiones 1 y 2.

Se llevó a cabo un análisis para identificar posibles problemas de multicolinealidad e independencia en las variables. Para evaluar la multicolinealidad, se calculó la tolerancia, mientras que para verificar la independencia de los residuos, se utilizó la prueba de Durbin-Watson con respecto al modelo final.

Tabla 1. Análisis de multicolinealidad e independencia de las principales variables explicativas

Variable	1/VIF	DWTEST
Nmb	.539	.329
Np	.549	
ce	.859	
M.o.	.879	
K	.836	
S	.645	
Na	.789	

Nota. Nmb, campos bananeros (A1, A2, A3, A4, A5, A6); np., parcelas evaluadas (HH, HL, LH, LL); ce., Conductividad eléctrica; M.o., Materia orgánica; K, Potasio; S, Azufre; Na, Sodio.

Para las variables “nmb” y “np”, el valor de la tolerancia es .539 y .549, lo que indica cierta multicolinealidad no tan significativa, mientras que las otras variables como “ce”, “M.o”, “K”, “S”, “Na” indican baja multicolinealidad. Además, la prueba de Durbin-Watson del modelo final arrojó un valor de 0.329, lo que sugiere que los residuos podrían no ser independientes.

3.2 Regresión lineal múltiple

En total en el análisis de regresión por pasos se seleccionaron las 17 variables predictoras incluyendo las categorías de clasificación de producción, con el objetivo de representar la mayor proporción de la varianza del IP.

Se estableció una relación funcional entre el IP, las variables propuestas por el algoritmo y aquellas estadísticamente significativa ($p < .05$). Un total de 7 variables identificadas (tabla 1) en la regresión por pasos se incluyeron en el modelo final, la cual mostró la mayor bondad de ajuste ($R^2 = 0.9284$), indicando que existe un buen grado de asociación entre el IP y las variables predictoras del modelo, como se indica en la siguiente ecuación:

$$y = 1.36 - 0.09(\mathbf{A2}) + 0.23(\mathbf{A3}) - 0.22(\mathbf{A4}) - 0.02(\mathbf{A5}) + 0.58(\mathbf{A6}) - 0.99(\mathbf{HL}) \\ - 3.48(\mathbf{LH}) - 2.96(\mathbf{LL}) - 0.26(\mathbf{C.e.}) - 0.61(\mathbf{M.o}) + 0.21(\mathbf{K}) \\ + 0.021(\mathbf{S}) + 0.0039(\mathbf{Na})$$

Donde, A2, A3, A4, A5 y A6 es la ubicación geográfica de las parcelas, HL, LH y LL son los niveles de productividad históricos reportados por los agricultores, C.e. es la conductividad eléctrica, M.o. es la materia orgánica, K es el nivel de potasio, S es el nivel de azufre y Na los niveles de sodio.

Tabla 2. Resumen estadístico de las principales variables explicativas

Parámetro	Mean	S.D.	Range (max – min)
Constant			
C.e.	.994	.329	.260 – 1.793
M.o.	1.479	.156	1.103 – 1.783

K	1.011	.502	.000 – 2.000
S	29.81	6.723	13.67 – 46.00
Na	177.16	44.375	78.33 – 287.67

Nota. Constant, Intercepto; Ce., Conductividad eléctrica; M.o., Materia orgánica; K, Potasio; S, Azufre; Na, Sodio.

En la tabla 2 se muestra los estimados y estadísticos cuantitativos de las 7 variables que explican mejor la predicción del IP, además se muestra una alta significancia en los niveles de la variable de ubicación y nivel de productividad, por lo que es evidente la dependencia del índice de productividad en el tipo de suelo para cada zona.

Tabla 3. Coeficientes del modelo

	Estimación	Error estándar	Estadístico T	Valor P
Constant*	1.365	0.621227	2.198	0.03301
A3	-0.099	0.147105	-0.674	0.50363
A4	0.230	0.144132	1.597	0.11701
A5	-0.227	0.243724	-0.935	0.3549
A6	-0.026	0.250802	-0.108	0.9148
A7*	0.584	0.239398	2.44	0.0186
HL***	-0.997	0.164889	-6.051	2.43·10 ⁻⁷
LH***	-3.482	0.160713	-21.672	2·10 ⁻¹⁶
LL***	-2.967	0.147056	-20.178	2·10 ⁻¹⁶
C.e.	-0.269	0.171482	-1.573	0.12261
M.o.	-0.616	0.358771	-1.719	0.09241
K	0.211	0.11417	1.857	0.06978
S*	0.021	0.009708	2.179	0.0345
Na**	0.0039	0.00133	2.967	0.00476

* $p < .05$, ** $p < .01$, *** $p < .001$

3.3 Estimación del modelo por regresión por pasos

En la tabla 3 se muestra el análisis de rendimiento para los modelos desarrollados, desde el modelo más complejo (modelo 5). En cada paso, se descartó la variable con mayor penalización (menor nivel de significancia) hasta llegar a un modelo con únicamente 7 variables. Los resultados indican que a medida que disminuye el número de variables, el valor de R² aumenta y el criterio de información de Akaike disminuye hasta obtener un modelo con una explicación del 92,84% de variabilidad total.

Tabla 4. Evaluación de los modelos de Regresión

Model	Variabes (número)	R ²	F	AIC	P-value	Variabes excluidas
5	U.G., N.P., C.e., M.o., C.total, K, Ca, S, hierro, cobre, Na (11)	.9264	44.7	79.336	2.2·10 ⁻¹⁶	-
7	U.G., N.P., C.e., M.o., C.total, K, Ca, S, hierro, Na (10)	.9275	48.17	77.864	2.2·10 ⁻¹⁶	Cu

8	U.G., N.P., C.e., M.o., C.total, K, Ca, S, Na (9)	.9284	51.99	76.505	$2.2 \cdot 10^{-16}$	Cu, Fe
9	U.G., N.P., C.e., M.o., K, Ca, S, Na (8)	.9288	56	75.482	$2.2 \cdot 10^{-16}$	Cu, Fe, Cttotal
10	U.G., N.P., C.e., M.o., K, S, Na (7)	.9284	59.84	75.165	$2.2 \cdot 10^{-16}$	Cu, Fe, Cttotal, Ca

Nota. U.G., ubicación geográfica, N.P. nivel de productividad; Cttotal, carbono total; Ce., Conductividad eléctrica; M.o., Materia orgánica; K, Potasio; S, Azufre; Na, Sodio

Los modelos del 5 a 10 siguiendo la regresión por pasos presentaban rendimientos con aceptación ascendente. El modelo 10 explica el mayor porcentaje de variación de Índice de productividad.

En la comparativa del modelo 5 y 10, es evidente que el Modelo final muestra un error estándar residual ligeramente menor (0.4026) en comparación con el Modelo 5 (0.4081), lo que indica una mejor capacidad de ajuste de este. Además, el Modelo 10 tiene un valor de R cuadrado ajustado ligeramente más alto (0.9284) en comparación con el Modelo 5 (0.9264), lo que sugiere que el Modelo 10 explica mejor la variabilidad de la variable dependiente después de ajustar por el número de predictores en el modelo.

A pesar de que ambos modelos tienen valores de estadística F altos y valores p muy bajos ($< 2.2e-16$), lo que indica un valor estadístico significativo del modelo en general, el Modelo 10 parece ser preferible debido a su menor error estándar residual y su R cuadrado ajustado más alto.

Adicional, para este tipo modelos se realizó el calculo de errores de predicción lo cual es un método que permite mayor selección de modelo.

Tabla 5. Error de predicción de los modelos de regresión

Model	Variables (número)	R ²	F	AIC	P-value	Variables excluidas	MCE
5	U.G., N.P., C.e., M.o., C.total, K, Ca, S, hierro, cobre, Na (11)	.9264	44.7	79.336	$2.2 \cdot 10^{-16}$	-	6.99
7	U.G., N.P., C.e., M.o., C.total, K, Ca, S, hierro, Na (10)	.9275	48.17	77.864	$2.2 \cdot 10^{-16}$	Cu	7.06
8	U.G., N.P., C.e., M.o., C.total, K, Ca, S, Na (9)	.9284	51.99	76.505	$2.2 \cdot 10^{-16}$	Cu, Fe	7.13
9	U.G., N.P., C.e., M.o., K, Ca, S, Na (8)	.9288	56	75.482	$2.2 \cdot 10^{-16}$	Cu, Fe, Cttotal	7.25
10	U.G., N.P., C.e., M.o., K, S, Na (7)	.9284	59.84	75.165	$2.2 \cdot 10^{-16}$	Cu, Fe, Cttotal, Ca	7.46

Con un valor de error de predicción de 6.99, este modelo muestra un rendimiento superior en comparación con los demás modelos, los cuales presentan valores de error más altos, considerando las 5 variables con mayor proporción de información y significancia para el modelo.

Además, para visualizar la capacidad predictiva de los modelos propuestos, se trazaron los valores del PI observado y predichos por el modelo 5 y 10 respectivamente, siendo el primero el de peor ajuste.

Figura 3. Predicción del modelo 5

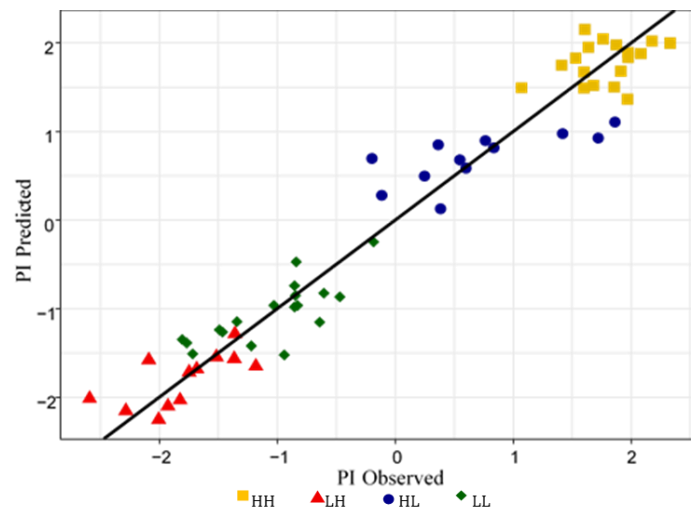
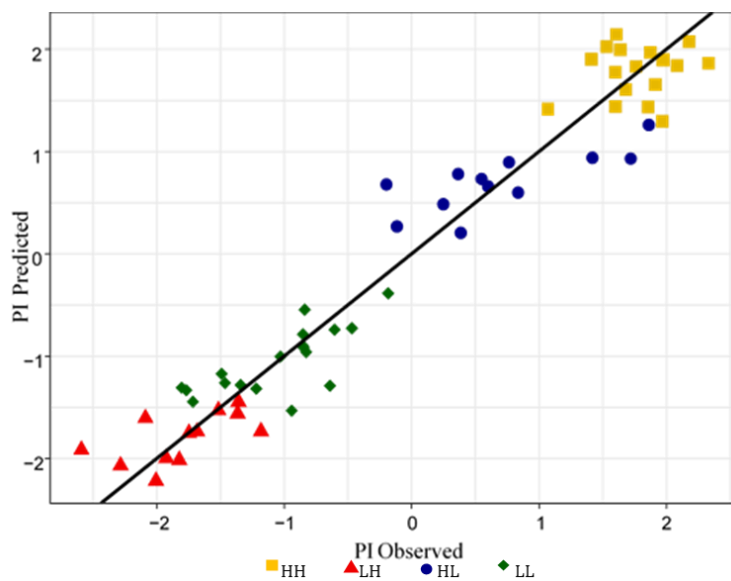


Figura 4. Predicción del modelo final



Ambos modelos muestran valores de R cuadrado múltiple significativamente altos (0.9442 para el Modelo final y 0.9476 para el Modelo 5), Indicando un alto porcentaje de

explicación de la variable dependiente por medio de las variables independientes seleccionadas.

DISCUSIÓN Y CONCLUSIONES

Los productores de banano en Ecuador buscan mejorar el uso de insumos en las plantaciones, motivo por el que es importante la evaluación de áreas de producción. Este estudio se centra en identificar las propiedades del suelo que influyen en la productividad del banano. El objetivo es desarrollar modelos simples para caracterizar la productividad mediante un Índice de Productividad basado en las características del suelo.

Se ha comprobado la factibilidad de crear varios modelos para evaluar el Índice de Productividad (IP) en cultivos de banano, considerando las características del suelo. Esto ofrece una guía para elegir la mejor opción según las necesidades de las partes interesadas. Aunque el modelo 5 mostró una gran capacidad predictiva al igual que el modelo 10, este se descartó por la gran cantidad de variables incluidas.

El modelo obtenido abarco 5 variables predictoras de carácter cuantitativo incluyendo Conductividad eléctrica, Materia orgánica, Potasio; Azufre y Sodio, esto aporta acerca de la naturaleza de los suelos en Ecuador, para explicar que parámetros son los más significativos en la predicción de la productividad bananera. Otros estudios como los de Olivares et al., (2020) en los que evalúan las características fisicoquímicas y microbiológicas del suelo, determinaron que la mayor capacidad de predicción para el índice de productividad, abarcan las características como Resistencia a la penetración, Respiración microbiana total y la cantidad de nematodos de vida libre, obteniendo un modelo con un error absoluto medio de 0,23.

El modelo final obtenido considera a los macroelementos potasio y azufre como significativos dentro del modelo para explicar una parte de la variabilidad del IP, esto se puede deber a que la fluctuación de los niveles de potasio y azufre, resultan en diferentes niveles de desempeño en las funciones fisiológicas de las plantas, debido a que estos macroelementos cumplen papeles como la síntesis de proteínas, fotosíntesis, formación de clorofila y en definitiva son parte importante en la nutrición del banano (Karim et al, 2019).

En definitiva, el estudio abordó la modelación de la capacidad de producción de banano utilizando datos biométricos de las plantas. Se desarrolló un Índice de Productividad (IP) que posteriormente se vinculó con las propiedades del suelo mediante la creación de un modelo empírico. Este modelo de predicción se construyó utilizando cinco atributos del suelo que son comúnmente reconocidos como indicadores clave de su calidad. Estos atributos, que incluyen el contenido de magnesio, materia orgánica, macro y microelementos que son considerados de gran importancia tanto desde un punto de vista agronómico como ambiental.

Finalmente, el modelo resultante tiene la capacidad de identificar de manera confiable áreas con niveles altos y bajos de productividad de bananos en regiones específicas de Ecuador, habiendo sido validado mediante pruebas adicionales en campo. En última instancia, la identificación de los atributos del suelo más relevantes en términos de productividad del banano en este estudio puede ser de gran utilidad para los agricultores, ya que les proporciona orientación sobre las prácticas de manejo sostenible del suelo. Esto, considerando la variabilidad natural de los suelos, podría contribuir significativamente a mejorar la sostenibilidad a largo plazo de la producción de bananos en Ecuador.

REFERENCIAS

- Arunraj, N. S., Ahrens, D., y Fernandes, M. (2016). Application of SARIMAX model to forecast daily sales in food retail industry. *Int. J. Oper. Res. Inf. Syst.* , 7(2), 1–21.
- Barrezueta-Unda, S. A., Azuero-Caamaño, H., y Gootman Jadan, I. R. (2023). Development of banana (Musa AAA) under different doses of biochar and chemical fertilizers. *Bionatura*, 8(3), 1–10. <https://doi.org/10.21931/RB/2023.08.03.85>
- Bhausahab, A., Lazarus, P., Vijayan, A., Sathayan, A. R., y Joseph, B. (2023). Impact of climate change on banana production in thiruvananthapuram district of kerala, India. *Asian J. Agric. Extens. Econ. Sociol*, 41(3), 114–123.
- Bonilla-Bedoya, S., Valencia, K., Herrera, M. Á., López-Ulloa, M., Donoso, D. A., y Macedo Pezzopane, J. E. (2023). Mapping 50 years of contribution to the development of soil quality biological indicators. *Ecological Indicators*, 148, 110091. <https://doi.org/10.1016/j.ecolind.2023.110091>
- Box, G., Jenkins, G. M., Reinsel, G. C., y Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley y Sons.

- Cedeño, G. (2023). *Banano, plátano y otras musáceas*. <https://www.iniap.gob.ec/banano-platano-y-otras-musaceas/>
- Clercx, L., y Huyghe, B. (2013). TOWARDS A MORE SUSTAINABLE BANANA - LIMITATIONS AND STRENGTHS OF A TERRITORIAL APPROACH. *Acta Horticulturae*, 986, 353–362. <https://doi.org/10.17660/ActaHortic.2013.986.38>
- Galarza Suárez, L. (2019). Tierra, trabajo y tóxicos: Sobre la producción de un territorio bananero en la costa sur del Ecuador. *Estudios Atacameños*. <https://doi.org/10.22199/issn.0718-1043-2019-0034>
- Hanson, T. (2010). *Multiple regression*.
- Iori, P., Silva, R. B. Da, Junior, M. D. S. D., Nakamura, R., y Ferreira De Almeida, L. C. (2020). Soil quality analysis in riparian areas for soil and water resource management. *Archives of Agronomy and Soil Science*, 66(5), 572–585. <https://doi.org/10.1080/03650340.2019.1630822>
- ISO. (2018). *ISO/IEC 17025:2017 General requirements for the competence of testing and calibration laboratories*. ISO.
- León Ajila, J. P., Espinosa Aguilar, M. A., Carvajal Romero, H. R., y Quezada Campoverde, J. (2023). Análisis de la producción y comercialización de banano en la provincia de El Oro en el periodo 2018-2022. *Ciencia Latina Revista Científica Multidisciplinar*, 7(1), 7494–7507. https://doi.org/10.37811/cl_rcm.v7i1.4981
- Karim, A. A., Kumar, M., Mohapatra, S., Singh, S. K., & Panda, C. R. (2019). Co-plasma processing of banana peduncle with phosphogypsum waste for production of lesser toxic potassium–sulfur rich biochar. *Journal of Material Cycles and Waste Management*, 21, 107-115.
- Ministerio de comercio exterior. (2017). *Informe sector bananero ecuatoriano*. <https://www.produccion.gob.ec/wp-content/uploads/2019/06/Informe-sector-bananero- espa%C3%B1ol-04dic17.pdf>
- Navarro, P. (2023). *Production of bananas in Ecuador from 2015-2021*. Statista. <https://www.statista.com/statistics/1054054/ecuador-banana-production-volume/#:~:text=In%202021%2C%20production%20of%20banana,leading%20banana%20producing%20countries%20worldwide.>
- Neog, B., Gogoi Bipin, y Patawary, A. N. (2022). Development of hybrid time series models for forecasting autumn rice using arimax-ann and arimax-svm. *Ann. for. Res*, 65(1), 9119– 9133.
- Olivares, B. O. (2023). *Identification of the Main Soil Variables Correlated with Banana Productivity* (pp. 119–140). https://doi.org/10.1007/978-3-031-34475-6_5
- Olivares, B. O., Araya-Alman, M., Acevedo-Opazo, C., Rey, J. C., Cañete-Salinas, P., Kurina, F. G., Balzarini, M., Lobo, D., Navas-Cortés, J. A., Landa, B. B., y Gómez, J. A. (2020). Relationship Between Soil Properties and Banana Productivity in the Two Main Cultivation Areas in Venezuela. *Journal of Soil Science and Plant Nutrition*, 20(4), 2512– 2524. <https://doi.org/10.1007/s42729-020-00317-8>
- Patrick, S., Mirau, S., Mbalawata, I., y Leo, J. (2023). Time series and ensemble models to forecast banana crop yield in Tanzania, considering the effects of climate

- change. *Resources, Environment and Sustainability*, 14, 100138.
<https://doi.org/10.1016/j.resenv.2023.100138>
- Rathod S., y Mishra G.C. (2018). Statistical models for forecasting mango and banana yield of karnataka, India. *J. Agric. Sci. Technol.*, 20(4), 803–816.
- Rosales, F. E., Pocasangre, L. E., Tejos, J., Peña, W., y Serrano, E. (2008). *Guía de diagnóstico de la calidad y salud de suelos bananeros*. Bioversity International.
- Sagamiko, T., Shaban, N., y Mbalawata, I. (2020). Sensitivity analysis and uncertainty parameter quantification in a regression model: The case of deforestation in tanzania. *Tanzan. J. Sci.*, 46(3), 673–683.
- Segura, R. A., Serrano, E., Pocasangre, L., Acuña, O., Bertsch, F., Stoorvogel, J. J., y Sandoval, J. A. (2015). Chemical and microbiological interactions between soils and roots in commercial banana plantations (Musa AAA, cv. Cavendish). *Scientia Horticulturae*, 197, 66–71. <https://doi.org/10.1016/j.scienta.2015.10.028>
- USDA. (1995). *Soil survey laboratory methods manual*. USDA.
- Zhang, G. L., y Gong, Z. T. (2012). *Soil survey laboratory methods manual* (Vol. 38). Science Press.