

# PROYECCIÓN DE LA PRODUCCIÓN BANANERA EN EL ORO: DESARROLLO DE UN MODELO DE REGRESIÓN Y SIMULACIÓN BASADA EN LA RELACIÓN HISTÓRICA ÁREA-RENDIMIENTO (2002-2024)

## PROJECTION OF BANANA PRODUCTION IN EL ORO: DEVELOPMENT OF A REGRESSION AND SIMULATION MODEL BASED ON THE HISTORICAL AREA-YIELD RELATIONSHIP (2002- 2024)

Jhon Paulino Camacho Azan<sup>1</sup>, Alfredo Rodrigo Colcha Ortiz<sup>2</sup>

{[jhon.camacho@unach.edu.ec](mailto:jhon.camacho@unach.edu.ec)<sup>1</sup>, [alfredo.colcha@esPOCH.edu.ec](mailto:alfredo.colcha@esPOCH.edu.ec)<sup>2</sup>}

Fecha de recepción: 14/09/2026 / Fecha de aceptación: 08/10/2026 / Fecha de publicación: 14/10/2026

**RESUMEN:** El objetivo de este estudio fue evaluar la relación entre superficie plantada y producción bananera en la provincia de El Oro y construir proyecciones exploratorias con simulación de residuos, se utilizaron 23 registros anuales de 2002 - 2024 procedentes de una base de datos de estadísticas agropecuarias oficiales en el Ecuador. La serie completa se empleó para la descripción histórica y el período 2014 - 2024 para el ajuste de regresiones lineal y cuadrática, considerando el cambio metodológico documentado en la ESPAC en 2014. El rendimiento se calculó como producción dividida por superficie cosechada y se analizó como variable complementaria. Los modelos de producción obtuvieron coeficientes de determinación de 0,2300 y 0,2323, respectivamente. El término cuadrático no aportó una mejora estadísticamente significativa y el modelo lineal se conservó por parsimonia. En una validación temporal de cinco años, la regresión lineal presentó un RMSE de 194.160 t, frente a 117.191 t del promedio histórico. La simulación de 10.000 réplicas situó la mediana condicional en 1.405.420 t para 2025 y 1.387.037 t para 2030. El rango P5–P95 de 2030 fue de 1.000.564–1.545.976 t. Estos resultados muestran una capacidad predictiva limitada y deben interpretarse como escenarios sujetos a una trayectoria supuesta de superficie, sin incluir toda la incertidumbre de la proyección. El principal aporte consiste en delimitar el alcance de la relación área–producción y establecer un procedimiento reproducible para su evaluación regional.

**Palabras clave:** Producción bananera, Provincia El Oro, superficie plantada, rendimiento agrícola, regresión lineal, simulación de Monte Carlo, proyección de producción, validación temporal

<sup>1</sup>Facultad de Ingeniería, Universidad Nacional de Chimborazo, <https://orcid.org/0009-0000-6413-6378>

<sup>2</sup>Escuela Superior Politécnica de Chimborazo (ESPOCH), <https://orcid.org/0009-0005-2280-5189>

**ABSTRACT:** This study evaluated the relationship between planted area and banana production in El Oro province and developed exploratory projections through residual simulation. Twenty-three annual records from 2002 - 2024 were obtained from a cleaned dataset of official agricultural statistics. The full series was used for historical description, whereas linear and quadratic regressions were fitted to 2014 - 2024 data, acknowledging the methodological change documented for the ESPAC survey in 2014. Yield was calculated as production divided by harvested area and analysed as a complementary variable. The production models achieved  $R^2$  values of 0.2300 and 0.2323, respectively. The quadratic term provided no statistically significant improvement, and the linear model was retained for parsimony. In a five-year temporal validation, linear regression had an RMSE of 194,160 t, compared with 117,191 t for the historical mean. A simulation with 10,000 replicates produced conditional medians of 1,405,420 t in 2025 and 1,387,037 t in 2030. The 2030 P5–P95 range was 1,000,564–1,545,976 t. Predictive performance was limited, and these results represent scenarios conditional on an assumed area trajectory rather than projections incorporating all sources of uncertainty. The study clarifies the scope of the area–production relationship and provides a reproducible procedure for regional assessment.

**Keywords:** *Banana production, El Oro Province, planted area, agricultural yield, linear regression, Monte Carlo simulation, production forecasting, temporal validation*

## INTRODUCCIÓN

La producción bananera ocupa un lugar relevante en el análisis agrícola de la provincia de El Oro por la extensión del cultivo y su vinculación con la economía de los territorios productores. Según la Encuesta de Superficie y Producción Agropecuaria Continua (ESPAC) de 2024, la provincia concentró el 18,3 % de la superficie nacional cosechada de banano. En conjunto, El Oro, Guayas y Los Ríos reunieron el 87,1 % de esa superficie (1). Esta concentración territorial justifica estudiar las variaciones de la producción provincial y disponer de herramientas que permitan examinar sus posibles trayectorias con información estadística verificable (2).

La importancia del banano también se expresa en las condiciones de vida de las familias vinculadas al cultivo. Vega Armijos et al. analizaron la dependencia económica y social de productores de banano orgánico en el sitio La Palestina, cantón El Guabo, e identificaron la relación de esta actividad con los ingresos y el empleo local (3). Aunque sus resultados corresponden a una población y a un sistema productivo específicos, aportan un antecedente regional para comprender por qué las variaciones del volumen producido pueden ser relevantes más allá de su dimensión agrícola. El seguimiento de la producción resulta, por tanto, de interés para productores, asociaciones e instituciones que necesitan organizar sus actividades con información sobre el contexto provincia (4,5).

La planificación productiva requiere considerar, además, la forma en que se gestionan las unidades agrícolas. Palomeque Solano et al. estudiaron 74 unidades vinculadas a una asociación de productores de El Oro y analizaron dimensiones de planificación, control, soporte tecnológico

y rentabilidad (6). Su trabajo relaciona la gestión de los procesos con el funcionamiento económico de las unidades productivas. Estos antecedentes muestran que la extensión de las plantaciones es solo una parte de la información necesaria para comprender el sector (7). Sin embargo, los resultados de encuestas de gestión no permiten sustituir el análisis de una serie anual de superficie y producción: ambas aproximaciones responden a preguntas y escalas diferentes (8).

La superficie constituye solo uno de los factores relacionados con la producción bananera, las condiciones climáticas, incluidas las alteraciones de las precipitaciones y la temperatura asociadas con El Niño (9), pueden afectar el desarrollo del cultivo y las condiciones de cosecha. Enfermedades como la sigatoka negra pueden comprometer la productividad, mientras que las variaciones de precios y costos de insumos pueden influir en la toma de decisiones de inversión y manejo agrícola (10). Estos factores ofrecen posibles explicaciones de las fluctuaciones productivas que una regresión basada exclusivamente en superficie no representa. Sin embargo, su contribución a la variabilidad de la producción de El Oro no se cuantifica en este estudio, pues la base analizada no incorpora indicadores climáticos, fitosanitarios ni económicos (11).

Desde el punto de vista productivo, el volumen anual de banano está vinculado con la superficie que se cosecha y el rendimiento obtenido en ella. Cuando el rendimiento agregado se calcula dividiendo producción entre superficie cosechada, la igualdad producción igual a área cosechada por rendimiento constituye una identidad de cálculo. La superficie plantada describe la extensión ocupada por el cultivo, pero no coincide necesariamente con la superficie cosechada en un año (12). Por ello, el análisis debe distinguir expansión del área, participación del área cosechada y productividad por hectárea. Una producción mayor puede coexistir con menor superficie si el rendimiento agregado aumenta, aunque los datos provinciales por sí solos no identifican las causas de ese cambio (13).

La interpretación histórica también depende de la comparabilidad de las estadísticas, la documentación del INEC registra una actualización del marco muestral y cambios en la recolección de información de la ESPAC en 2014 (1). Estos antecedentes son relevantes al examinar la discontinuidad observada entre 2013 y 2014, pues una variación en los registros no debe atribuirse automáticamente a un cambio físico de la misma magnitud en las plantaciones. De ahí la necesidad de distinguir el período utilizado para describir la evolución histórica del período empleado para estimar los modelos. La delimitación temporal reduce la mezcla de condiciones de medición diferentes, pero no elimina todas las limitaciones de comparabilidad (14).

Los modelos de regresión ofrecen una forma sencilla de resumir la asociación entre superficie y producción, una especificación lineal estima una pendiente constante dentro del rango analizado, mientras que una cuadrática permite representar curvatura (15). La mayor flexibilidad de esta última incorpora un parámetro adicional, cuya utilidad debe contrastarse con el desempeño observado (16). En una serie corta, añadir complejidad puede mejorar el ajuste a los datos disponibles sin aportar precisión en años posteriores. Por esta razón, la evaluación debe considerar tanto la capacidad explicativa como la sencillez del modelo y su comportamiento fuera de la muestra de estimación (17).

El ajuste histórico y el pronóstico responden a tareas distintas, los residuos muestran las diferencias entre observaciones y valores ajustados con información que el modelo ya utilizó; los errores de pronóstico se calculan sobre observaciones posteriores (18). Una validación temporal respeta esta separación al ajustar el modelo con los años anteriores a cada año evaluado (19). En este estudio, la superficie futura también debe estimarse dentro de cada entrenamiento, porque no estaría disponible al formular el pronóstico. Comparar la cadena completa con referencias sencillas, como el promedio histórico o el último valor observado, permite establecer si la incorporación del área mejora realmente la anticipación de la producción (20).

La simulación de Monte Carlo complementa la regresión al representar una distribución de resultados bajo supuestos definidos, mediante el remuestreo de residuos, o bootstrap, pueden generarse valores compatibles con la variabilidad de los errores históricos, sin imponer que estos sigan una distribución normal (21). Los percentiles facilitan comunicar esa variabilidad, pero su significado depende del procedimiento aplicado, si se mantienen fijos los coeficientes y la superficie futura, el rango simulado expresa una incertidumbre condicionada a esos valores; no incorpora automáticamente el error de estimación del área, cambios del entorno ni factores agrícolas que no se midieron (22).

Los antecedentes regionales revisados aportan evidencia sobre dependencia económica, gestión productiva y cadena de valor, pero no responden directamente a la pregunta de cuánto contribuye la superficie plantada a anticipar la producción anual agregada de El Oro. La presente investigación aborda esa pregunta mediante una base histórica provincial y una evaluación explícita de los modelos, la relación histórica área rendimiento que enmarca el tema se examina distinguiendo el rendimiento como indicador complementario de la superficie plantada como predictor de la producción. Esta distinción evita utilizar como información independiente un rendimiento calculado a partir de la misma producción que se intenta estimar (23).

El objetivo del estudio fue evaluar la asociación entre superficie plantada y producción total de banano en El Oro, comparar modelos de regresión lineal y cuadrática y construir escenarios de producción del 2025 al 2030 mediante simulación de residuos. La serie desde el 2002 al 2024 se utilizó como contexto histórico y el período 2014 al 2024 como base de ajuste. Para un mejor contexto se incluyó el análisis del rendimiento y una validación temporal con métodos de referencia, con el presente estudio se pretende desarrollar un procedimiento reproducible para estudiar estas relaciones y establecer el alcance de sus proyecciones como apoyo exploratorio a la planificación regional.

## MATERIALES Y MÉTODOS

La investigación tuvo un enfoque cuantitativo y un diseño no experimental, basado en una serie anual de estadísticas agregadas para la provincia de El Oro, la unidad de análisis fue el año, no la finca ni el productor, se trabajó con 23 observaciones del 2002 al 2024 y con un subconjunto de 11 observaciones del 2014 al 2024 para estimar los modelos. El objetivo de la simulación fue desde el 2025 al 2030, las proyecciones se calcularon con información cerrada en 2024, de modo que no incorporan observaciones posteriores a ese año. El insumo analítico fueron los datos

obtenidos desde el Sistema de Información Pública Agropecuaria del ministerio competente en agricultura, con estadísticas de la ESPAC del INEC (1,14).

### Variables y delimitación temporal

La superficie plantada A se expresó en hectáreas, la superficie cosechada H en hectáreas y la producción P en toneladas. El rendimiento Y se calculó como  $P/H$  y se expresó en toneladas por hectárea. Este indicador es un cociente agregado provincial, no el promedio simple de rendimientos de fincas. La variación anual se obtuvo como 100 multiplicado por el cociente entre producciones consecutivas menos uno. La columna de variación de la base almacena proporciones, por lo que se multiplicó por 100 al informar porcentajes.

$$Y = P / H$$

La discontinuidad entre 2013 y 2014 se examinó junto con la documentación histórica de la ESPAC. El INEC informa una actualización del marco muestral y cambios de recolección en 2014 (1). Por ello, los datos previos se mantuvieron como contexto y el ajuste se restringió del 2014 al 2024. Esta decisión reduce la mezcla de períodos con condiciones de medición diferentes, aunque no demuestra homogeneidad completa del subconjunto ni permite atribuir toda la discontinuidad a una única causa administrativa.

### Ajuste y comparación de regresiones

Se estimaron por mínimos cuadrados ordinarios una regresión lineal de producción sobre superficie plantada y una regresión cuadrática que incorporó el área al cuadrado. Para facilitar el cálculo se utilizó el área centrada y escalada z. Ambos modelos conservan la misma información que las ecuaciones escritas en hectáreas (24).

$$z = (A - 40.000) / 1.000$$

$$P = \beta_0 + \beta_1 z + \varepsilon$$

$$P = \gamma_0 + \gamma_1 z + \gamma_2 z^2 + \varepsilon$$

Como análisis complementario se ajustó una regresión del rendimiento sobre z. No se incorporó el rendimiento contemporáneo como predictor de la producción porque se calcula a partir de esa misma producción. Su inclusión podría introducir una relación contable en lugar de aportar información independiente para predecir un año futuro.

Se calcularon  $R^2$ ,  $R^2$  ajustado, MAE, RMSE y MAPE. MAE corresponde al promedio del valor absoluto de los errores; RMSE, a la raíz del promedio de sus cuadrados; y MAPE, al promedio del error absoluto relativo multiplicado por 100. Los errores de producción se informaron en toneladas y los de rendimiento en t/ha. La comparación del término cuadrático se efectuó mediante una prueba F entre modelos anidados. Se revisaron normalidad residual con Shapiro–Wilk, el estadístico descriptivo de Durbin–Watson y la distancia de Cook. Por el tamaño de la muestra, estos diagnósticos se interpretaron con cautela (25).

## Validación temporal de las proyecciones

Se aplicó una evaluación con origen móvil y ventana de entrenamiento expansiva, el primer ajuste utilizó datos del 2014 al 2019 para estimar 2020; luego se incorporó un año al entrenamiento hasta evaluar 2024. En cada una de las cinco iteraciones, la superficie del año evaluado se estimó exclusivamente con los años de entrenamiento mediante una regresión de área sobre tiempo. Esa superficie estimada alimentó el modelo de producción, de esa manera se evaluó la cadena de proyección sin utilizar como predictor el área observada del año que se intentaba anticipar (26).

El entrenamiento con seis observaciones responde a la extensión de la serie seleccionada y permite reservar cinco años para evaluar los pronósticos respetando su orden, la ventana expansiva incorpora progresivamente información adicional en cada ajuste (27). El tamaño inicial exige cautela al interpretar los coeficientes y las diferencias de desempeño, especialmente en la especificación cuadrática, que estima tres parámetros y dispone inicialmente de tres grados de libertad residuales, al evaluar la cadena completa de proyección, los errores de estimación de la superficie pueden trasladarse al pronóstico de producción, la evaluación se interpreta como una comparación exploratoria y no como evidencia concluyente de capacidad predictiva (28).

Las regresiones se compararon con dos referencias calculadas en cada entrenamiento: la producción del último año disponible y el promedio de todas las producciones del período de entrenamiento. Las métricas se agregaron sobre los cinco años evaluados, esta prueba analiza pronósticos a un año y no establece el desempeño de proyecciones a seis años. Su reducido número de casos limita la estabilidad de las comparaciones (29).

## Simulación condicional y reproducibilidad

Para construir el ejercicio prospectivo se ajustó una regresión de superficie plantada sobre año con los 11 registros de desde el 2014 al 2024, la superficie estimada para los años 2025 al 2030 se incorporó al modelo lineal de producción, seleccionado por parsimonia entre las regresiones evaluadas. Se generaron 10.000 réplicas mediante muestreo con reemplazo de los 11 residuos históricos, utilizando la semilla 123 en R, en cada réplica se añadió un residuo con una nueva muestra a la predicción de cada año. Este procedimiento aproxima la distribución empírica de las perturbaciones observadas: para cada año, la producción simulada puede tomar como máximo 11 valores distintos, correspondientes a la predicción puntual más cada residuo disponible. Con esto aumentar el número de réplicas reduce la variación atribuible al muestreo de la simulación, pero no incorpora información adicional ni genera perturbaciones fuera del rango histórico de los residuos (30).

$$P^* = \hat{P}(A \text{ proyectada}) + \varepsilon^*$$

Los percentiles 5, 50 y 95 describen los valores inferior, central y superior de la distribución simulada, el rango P5 a P95 contiene el 90 % central de esa distribución condicional. No es un intervalo de confianza de los coeficientes ni un intervalo predictivo con cobertura empírica validada, en este caso el procedimiento mantiene fijos el área futura y los parámetros estimados, por lo que no propaga su incertidumbre, de igual manera no incorpora el error de muestreo de

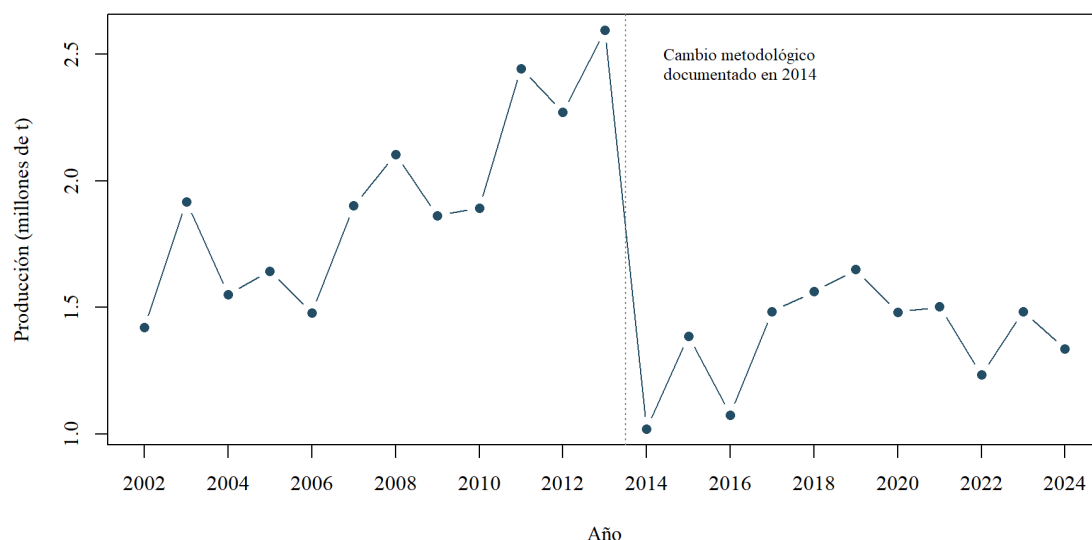
la estadística oficial, tampoco la posible dependencia temporal de los residuos ni cambios climáticos, fitosanitarios o económicos, el remuestreo supone que los residuos futuros pueden representarse mediante los históricos (31).

El análisis se ejecutó con el software R 4.5.1. El código de apoyo incluye controles de integridad, comparación de modelos, validación temporal, simulación y exportación de figuras y tablas, la base original se conservó sin cambios. Aumentar el número de réplicas estabiliza la aproximación a los percentiles de los residuos disponibles, pero no amplía la información histórica ni corrige las limitaciones del modelo.

## RESULTADOS

### Evolución histórica de la producción y del rendimiento

En el presente estudio la serie desde el 2002 al 2024 presentó una producción máxima de 2.594.000 t en 2013 y una mínima de 1.020.381 t en 2014, en la que se observa que entre ambos años la producción registrada disminuyó 60,66 % y la superficie plantada, 47,40 %. La coincidencia con el cambio metodológico documentado en la ESPAC impide interpretar esa discontinuidad únicamente como una contracción real del cultivo. La **Figura 1** muestra la serie completa y señala el límite utilizado para delimitar el período de ajuste.



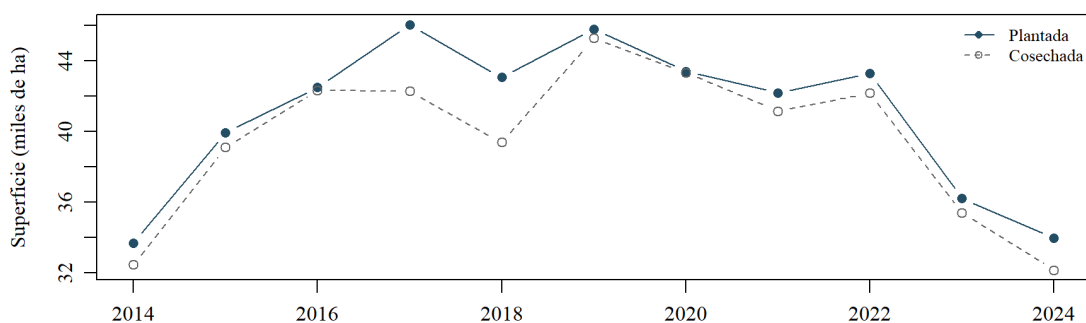
**Figura 1.** Producción bananera registrada en El Oro durante 2002–2024. La línea vertical identifica la separación entre los períodos descriptivo y de ajuste.

Durante el 2014 al 2024, la producción anual media fue de 1.383.404 t y varió entre 1.020.381 y 1.649.079 t. La superficie plantada media fue de 40.928 ha y la cosechada, de 39.562 ha, el rendimiento anual medio fue de 35,17 t/ha, calculado como promedio de los cocientes de cada año, el mínimo correspondió a 2016, con 25,40 t/ha, y el máximo a 2023, con 41,90 t/ha. La Tabla 1 permite distinguir las diferencias entre extensión y productividad.

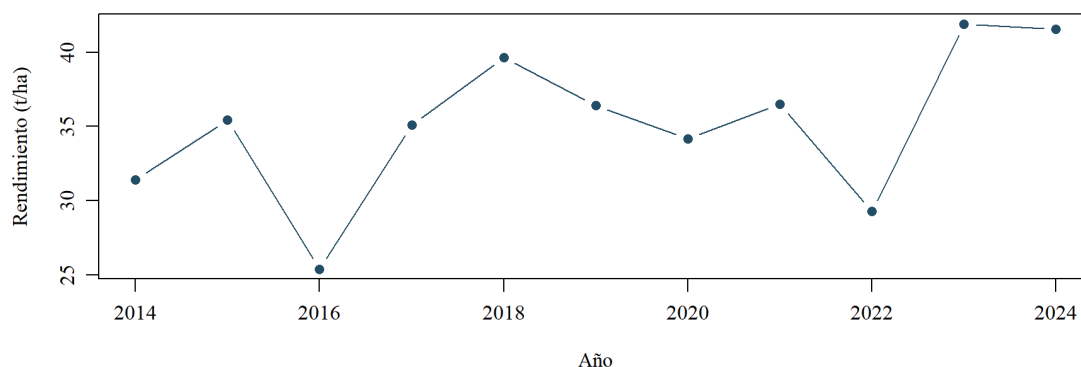
**Tabla 1. Variables anuales utilizadas en el ajuste.**

| Año  | Plantada<br>ha | Cosechada<br>ha | Producción<br>t | Rendimiento<br>t/ha |
|------|----------------|-----------------|-----------------|---------------------|
| 2014 | 33.711         | 32.468          | 1.020.381       | 31,43               |
| 2015 | 39.953         | 39.118          | 1.387.003       | 35,46               |
| 2016 | 42.513         | 42.340          | 1.075.395       | 25,40               |
| 2017 | 46.056         | 42.292          | 1.484.078       | 35,09               |
| 2018 | 43.066         | 39.403          | 1.562.203       | 39,65               |
| 2019 | 45.801         | 45.287          | 1.649.079       | 36,41               |
| 2020 | 43.416         | 43.344          | 1.481.452       | 34,18               |
| 2021 | 42.202         | 41.169          | 1.502.098       | 36,49               |
| 2022 | 43.307         | 42.179          | 1.235.002       | 29,28               |
| 2023 | 36.215         | 35.421          | 1.484.179       | 41,90               |
| 2024 | 33.968         | 32.161          | 1.336.569       | 41,56               |

En 2024 se registraron 33.968 ha plantadas, 32.161 ha cosechadas y 1.336.569 t de producción, con un rendimiento de 41,56 t/ha. La producción descendió 9,95 % respecto de 2023. La Figura 2 muestra que la reducción reciente de superficie coexistió con rendimientos elevados dentro del período de ajuste. Esta coexistencia permite describir diferencias productivas, pero no identificar los factores que las provocaron.



**Figura 2. Superficie plantada y cosechada, y rendimiento agregado provincial durante 2014–2024 (miles de ha).**



**Figura 3. Superficie plantada y cosechada, y rendimiento agregado provincial durante 2014–2024 (t/ha).**

### Capacidad explicativa de las regresiones

La ecuación lineal estimada, expresada en hectáreas, fue la siguiente. El coeficiente del área representa una asociación agregada de aproximadamente 21,69 t por hectárea adicional de superficie plantada dentro del rango observado. No equivale a un rendimiento por hectárea cosechada ni demuestra el efecto de aumentar el área en una finca.

$$\hat{P} = 495.558,8077 + 21,69283162 A$$

El modelo cuadrático y el análisis complementario del rendimiento obtuvieron las siguientes ecuaciones. Se muestran coeficientes con precisión suficiente para evitar que el redondeo del término cuadrático altere las estimaciones.

$$\hat{P} = 1.527.224,115 - 31,09267322 A + 0,0006668036192 A^2$$

$$\hat{Y} = 48,75859263 - 0,0003320798029 A$$

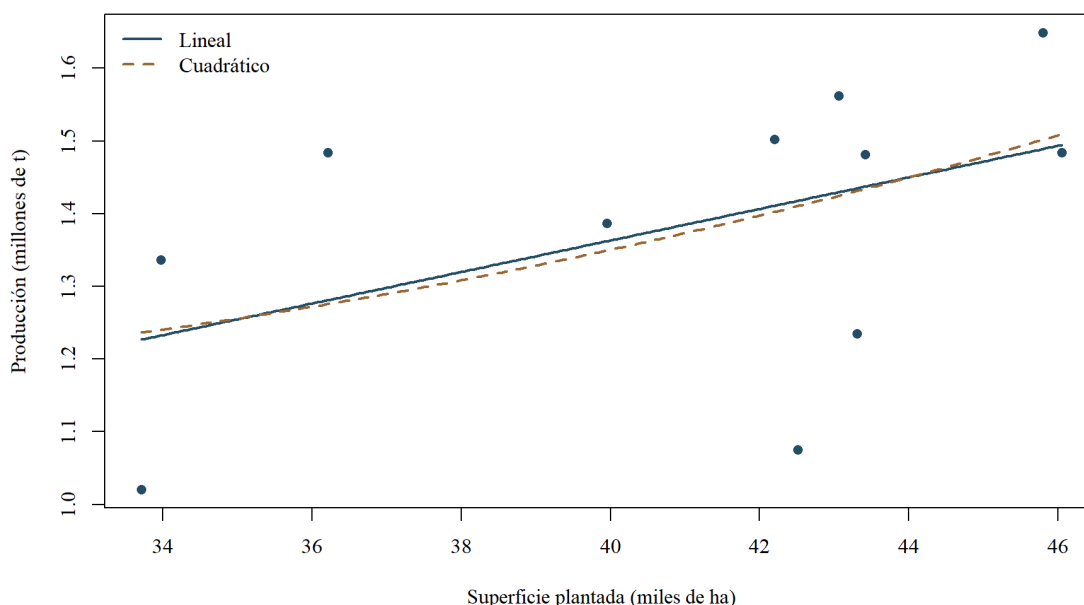
**Tabla 2. Ajuste de los modelos en del 2014 al 2024.**

| Modelo                | R <sup>2</sup> | R <sup>2</sup> ajustado | MAE       | RMSE      | MAPE % |
|-----------------------|----------------|-------------------------|-----------|-----------|--------|
| Producción lineal     | 0,2300         | 0,1445                  | 138.077 t | 166.791 t | 10,86  |
| Producción cuadrática | 0,2323         | 0,0404                  | 140.617 t | 166.540 t | 11,04  |
| Rendimiento lineal    | 0,0841         | -0,0176                 | 3,76 t/ha | 4,60 t/ha | 11,49  |

La regresión lineal explicó el 23,00 % de la variación de la producción en la muestra de ajuste. La regresión cuadrática elevó el R<sup>2</sup> a 23,23 %, pero su R<sup>2</sup> ajustado cayó a 0,0404. La prueba del término cuadrático presentó p = 0,8805, sin evidencia de una mejora suficiente para justificar ese parámetro adicional. Aunque el RMSE cuadrático disminuyó aproximadamente 251 t, sus MAE y

MAPE fueron mayores. Por tanto, la selección no puede basarse únicamente en el pequeño aumento de  $R^2$ .

En la regresión lineal, la pendiente del área presentó  $p = 0,1355$ ; su intervalo de confianza del 95 % fue de  $-8,23$  a  $51,62$  t/ha plantada. En el análisis del rendimiento, la asociación con el área fue débil y no significativa al nivel de 0,05, con  $R^2 = 0,0841$  y  $p = 0,3870$ . Estos resultados no demuestran ausencia de relaciones agronómicas; muestran que la superficie plantada, utilizada aisladamente y con 11 registros agregados, ofrece información limitada sobre la variación observada.



**Figura 4.** Relación entre superficie plantada y producción, con ajustes lineal y cuadrático sobre una malla de áreas dentro del rango observado.

La prueba de Shapiro–Wilk de los residuos lineales obtuvo  $p = 0,157$  y el estadístico descriptivo de Durbin–Watson fue de 1,93. Estas cifras no establecen que todos los supuestos estén satisfechos. El año 2014 presentó la mayor distancia de Cook, de 0,548, superior al umbral orientativo 4/11. Por ello, la influencia de observaciones individuales y la corta longitud de la serie deben considerarse al interpretar coeficientes y proyecciones.

### Desempeño en años no utilizados para el ajuste

El promedio histórico del entrenamiento presentó el menor MAE, RMSE y MAPE. Frente a esa referencia, el RMSE de la regresión lineal fue aproximadamente 65,7 % mayor. La regresión cuadrática también presentó mayor error que la lineal en las tres métricas. La validación indica que incorporar una trayectoria estimada del área no mejoró el pronóstico en estos cinco casos, considerando que el conjunto para la evaluación es pequeño, esta comparación debe entenderse como evidencia limitada al ejercicio realizado, no como una clasificación universal de métodos.

La comparación entre las Tablas 2 y 3 evidencia la diferencia entre describir asociaciones históricas y anticipar observaciones nuevas, la superficie es conceptualmente relevante para la producción, pero su relevancia no garantiza que una regresión simple sea la mejor opción de pronóstico, un método sencillo sin área obtuvo mejores resultados temporales. Esto delimita el uso de la regresión como instrumento exploratorio y evita presentar su simulación como un pronóstico operativo validado.

**Tabla 3. Validación temporal de la cadena de proyección en 2020–2024.**

| Procedimiento                  | MAE<br>t | RMSE<br>t | MAPE<br>% |
|--------------------------------|----------|-----------|-----------|
| Regresión lineal               | 162.890  | 194.160   | 12,05     |
| Regresión cuadrática           | 180.581  | 208.786   | 13,22     |
| Última producción<br>observada | 170.432  | 191.700   | 12,43     |
| Promedio del<br>entrenamiento  | 111.816  | 117.191   | 8,03      |

### Proyección condicionada a la superficie plantada

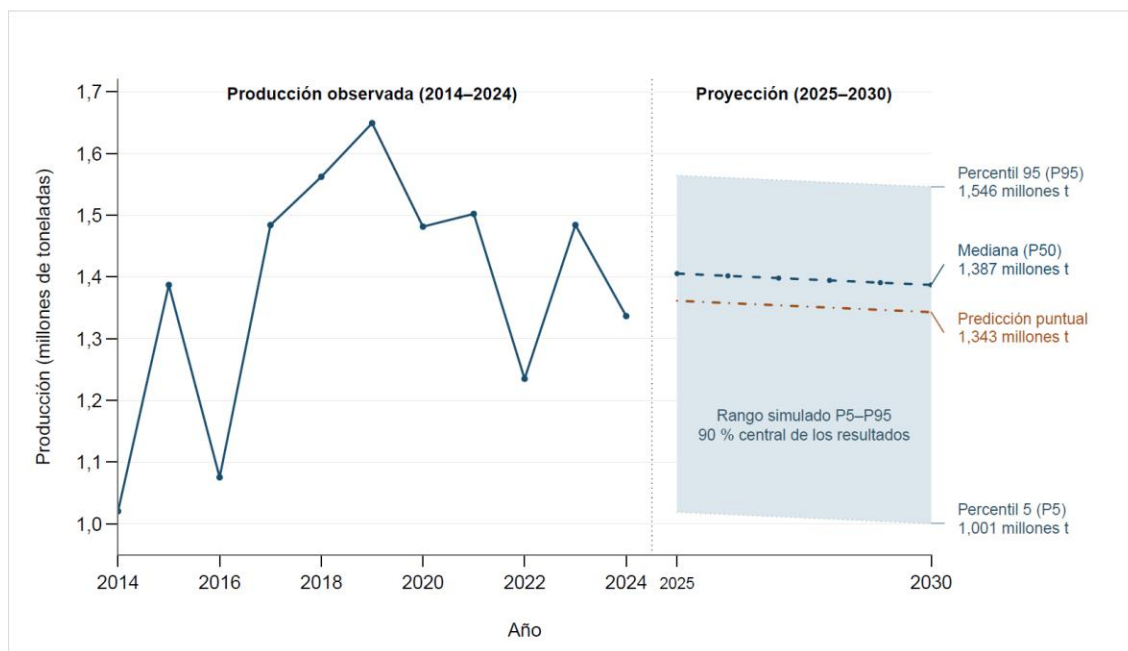
El modelo temporal del área estimó una disminución de 169,48 ha por año. Su  $R^2$  fue de 0,0163 y la pendiente presentó  $p = 0,7086$ , el intervalo de confianza del 95 % de la pendiente, de  $-1.163,39$  a  $824,42$  ha/año, incluyó valores positivos y negativos. Por tanto, los datos no sustentan una tendencia decreciente firme. La trayectoria utilizada en la Tabla 4 es una hipótesis de cálculo derivada del ajuste lineal del período completo del 2014 al 2024.

$$\hat{A} = 40.928,02527 - 169,4844680 (\text{año} - 2019)$$

**Tabla 4. Simulación del modelo lineal con 10.000 réplicas.**

| Año  | Área<br>ha | Ajuste<br>t | P5<br>t   | P50<br>t  | P95<br>t  |
|------|------------|-------------|-----------|-----------|-----------|
| 2025 | 39.911     | 1.361.344   | 1.018.947 | 1.405.420 | 1.564.359 |
| 2026 | 39.742     | 1.357.667   | 1.015.271 | 1.401.743 | 1.560.682 |
| 2027 | 39.572     | 1.353.991   | 1.011.594 | 1.398.067 | 1.557.006 |
| 2028 | 39.403     | 1.350.314   | 1.007.917 | 1.394.390 | 1.553.329 |
| 2029 | 39.233     | 1.346.638   | 1.004.241 | 1.390.713 | 1.549.652 |
| 2030 | 39.064     | 1.342.961   | 1.000.564 | 1.387.037 | 1.545.976 |

La mediana simulada pasó de 1.405.420 t en 2025 a 1.387.037 t en 2030, una reducción de 1,31 % entre esos dos años. Para 2030, el rango de P5 a P95 se situó entre 1.000.564 y 1.545.976 t. La mediana difiere del ajuste puntual porque la distribución empírica de los residuos es asimétrica: su mediana no es cero, aunque su media sí lo es por las propiedades del ajuste con intercepto. Por ello, ambas columnas describen cantidades distintas.



**Figura 5. Producción observada y proyección exploratoria del modelo lineal. La banda muestra los percentiles condicionales P5–P95.**

La superficie proyectada para 2025 fue de 39.911 ha, un 17,49 % superior a las 33.968 ha registradas en 2024, el salto corresponde a una discontinuidad entre el último valor observado y la trayectoria estimada mediante la regresión temporal de 2014 - 2024. Como la recta se ajustó al conjunto de observaciones y no se ancló al registro de 2024, el valor proyectado refleja el nivel y la pendiente del ajuste global, sin prolongar directamente el descenso observado entre 2022 y 2024. En consecuencia, el incremento inicial es un resultado de la especificación matemática utilizada y no constituye evidencia de una expansión efectiva o prevista de las plantaciones. Esta diferencia también influye en la producción proyectada: la mediana simulada de 2025 supera la producción observada en 2024, aunque disminuye posteriormente hasta 2030.

El área proyectada para 2025, de 39.911 ha, supera en 17,49 % la registrada en 2024. Esto ocurre porque la recta temporal se ajustó a todo el período del 2014 al 2024 y no se ancló al último dato. La caída entre 2025 y 2030 no implica una disminución continua desde 2024, la producción mediana de 2025 se sitúa por encima del último registro observado, esa transición es un resultado del supuesto de superficie y debe considerarse antes de utilizar la proyección para planificación.

La amplitud prácticamente constante de la banda se explica por el uso de la misma distribución de residuos en todos los horizontes, esta característica no demuestra estabilidad del riesgo. Una proyección que incorporase incertidumbre del área y de los parámetros podría producir bandas diferentes. La extensión del horizonte hasta 2030 incrementa, además, la dependencia de supuestos que no fueron evaluados mediante la validación a un año.

## DISCUSIÓN

La superficie plantada explicó aproximadamente el 23 % de la variabilidad productiva, y el término cuadrático no mejoró sustancialmente el ajuste, Varma y Bebbber (2019) (32) identificaron influencias climáticas y tecnológicas sobre el rendimiento bananero internacional, mientras que Bebbber (2019) (10) relacionó las condiciones climáticas con el riesgo de sigatoka negra, estos antecedentes fundamentan explorar predictores adicionales, aunque sus escalas y objetivos impiden una comparación directa con los resultados provinciales.

El promedio histórico superó a ambas regresiones en la validación temporal, indicando que incorporar superficie no mejoró la anticipación durante el período evaluado, el entrenamiento inicial con seis observaciones y los cinco pronósticos evaluados exigen cautela al generalizar, la disminución de la mediana simulada entre 2025 y 2030 debe interpretarse junto con el salto inicial del área proyectada, generado por el ajuste temporal global y no por evidencia de expansión agrícola. Asimismo, el usar nuevas muestras en 11 residuos restringe las perturbaciones a los valores disponibles: las 10.000 réplicas no amplían esa información ni incorporan incertidumbre del área o los coeficientes.

Con base en Hyndman y Athanasopoulos (2021) (33), podrían evaluarse modelos ARIMA parsimoniosos y regresiones con precipitación, temperatura e indicadores ENSO, previa ampliación de la información, los modelos estacionales requerirían datos mensuales o trimestrales, estas alternativas deberán contrastarse mediante validación temporal frente a referencias sencillas, el presente estudio aporta una evaluación reproducible del alcance predictivo de la superficie plantada.

## CONCLUSIONES

La producción bananera registrada en El Oro presentó variaciones importantes entre el 2002 y el 2024 y una discontinuidad entre 2013 y 2014, coincidente con cambios documentados en la ESPAC, la delimitación del ajuste a rango desde el 2014 al 2024 fue una decisión de comparabilidad, sin demostrar homogeneidad completa de la serie ni una causa única del cambio observado.

La superficie plantada explicó aproximadamente el 23 % de la variación de la producción dentro del período de ajuste, el término cuadrático no aportó una mejora estadísticamente significativa y la relación entre superficie y rendimiento fue débil. El modelo lineal ofrece una representación más sencilla entre las regresiones evaluadas, pero no permite inferencias causales sobre productividad.

La validación temporal mostró que el promedio histórico del entrenamiento obtuvo menor error que las regresiones, este resultado limita la capacidad predictiva atribuible a la relación área vs producción y debe formar parte de la valoración del modelo, junto con las métricas calculadas dentro de la muestra.

La simulación lineal situó la mediana condicional de 2030 en 1.387.037 t, con un rango de P5 y P95 de 1.000.564–1.545.976 t. La trayectoria depende de un supuesto de superficie cuya pendiente no fue significativa, los percentiles representan únicamente la variabilidad residual incorporada y no toda la incertidumbre futura. El procedimiento es reproducible y útil como evaluación exploratoria; su fortalecimiento requiere información adicional, actualización de la base y validación de horizontes más largos.

## REFERENCIAS BIBLIOGRÁFICAS

1. Boletin\_tecnico\_ESPAC\_2024.pdf [Internet]. [citado 9 de octubre de 2026]. Disponible en: [https://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas\\_agropecuarias/espac/2024/Boletin\\_tecnico\\_ESPAC\\_2024.pdf](https://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas_agropecuarias/espac/2024/Boletin_tecnico_ESPAC_2024.pdf)
2. Coral C, Mithöfer D. Contemporary narratives about asymmetries in responsibility in global agri-food value chains: the case of the Ecuadorian stakeholders in the banana value chain. *Agric Hum Values*. septiembre de 2023;40(3):1019-38. doi:10.1007/s10460-022-10405-3
3. Vega Armijos TG, Vite Cevallos H, Carvajal Romero H, Garzón Montealegre VJ. Dependencia económica y social de la producción de banano orgánico en el sitio La Palestina, cantón El Guabo, período 2017-2020. *Rev Metrop Cienc Apl*. 1 de junio de 2021;4(Suplemento 1):129-36. doi:10.62452/Okcykv41
4. Bonisoli L, Galdeano-Gómez E, Piedra-Muñoz L, Sorroche-del-Rey Y. Evaluating the Sustainability of Rural Agri-Producers: The Application of SAFA Smallholders in Ecuadorian Banana. *J Environ Dev*. junio de 2026;35(2):607-31. doi:10.1177/10704965261417260
5. Borja J. La producción de banano bajo el sistema de comercio justo: un análisis del caso ecuatoriano. *Siembra*. 7 de octubre de 2016;3(1):7-10. doi:10.29166/siembra.v3i1.185
6. Palomeque M, Urdaneta A, Meleán R. Production management in banana production units, El Oro province - Ecuador: Gestión de la producción en unidades de producción bananera, provincia de El Oro–Ecuador. *Rev Fac Agron Univ Zulia*. 6 de junio de 2023;40(2):e234019. doi:10.47280/RevFacAgron(LUZ).v40.n2.09
7. Bonisoli L, Galdeano-Gómez E, Piedra-Muñoz L, Pérez-Mesa JC. Benchmarking agri-food sustainability certifications: Evidences from applying SAFA in the Ecuadorian banana agri-system. *J Clean Prod*. noviembre de 2019;236:117579. doi:10.1016/j.jclepro.2019.07.054
8. Ramírez-Orellana A, Ruiz-Palomo D, Rojo-Ramírez A, Burgos-Burgos JE. The Ecuadorian Banana Farms Managers' Perceptions: Innovation as a Driver of Environmental Sustainability Practices. *Agriculture*. 5 de marzo de 2021;11(3):213. doi:10.3390/agriculture11030213
9. Varma V, Bebbber DP. Climate change impacts on banana yields around the world. *Nat Clim Change*. octubre de 2019;9(10):752-7. doi:10.1038/s41558-019-0559-9
10. Bebbber DP. Climate change effects on Black Sigatoka disease of banana. *Philos Trans R Soc B Biol Sci*. 24 de junio de 2019;374(1775):20180269. doi:10.1098/rstb.2018.0269
11. OECD, Food and Agriculture Organization of the United Nations. OECD-FAO Agricultural Outlook 2024-2033 [Internet]. OECD Publishing; 2024 [citado 13 de octubre de 2026]. (OECD-FAO Agricultural Outlook). Disponible en: [https://www.oecd.org/en/publications/oecd-fao-agricultural-outlook-2024-2033\\_4c5d2cfb-en.html](https://www.oecd.org/en/publications/oecd-fao-agricultural-outlook-2024-2033_4c5d2cfb-en.html) doi:10.1787/4c5d2cfb-en

12. Guia\_de\_usos\_de\_bases\_de\_datos\_2023.
13. Varma V, Bebbber DP. Climate change impacts on banana yields around the world. *Nat Clim Change*. octubre de 2019;9(10):752-7. doi:10.1038/s41558-019-0559-9
14. Cuichán M, Catota V. Encuesta de Superficie y Producción Agropecuaria Continua (ESPAC) 2024 [Internet]. inec@inec.gob.ec; 2025. Disponible en: <https://www.ecuadorencifras.gob.ec/estadisticas-agropecuarias-2/>
15. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning: with Applications in R [Internet]. New York, NY: Springer US; 2021 [citado 9 de octubre de 2026]. (Springer Texts in Statistics). Disponible en: <https://link.springer.com/10.1007/978-1-0716-1418-1> doi:10.1007/978-1-0716-1418-1
16. Burnham KP, Anderson DR, editores. Model Selection and Multimodel Inference [Internet]. New York, NY: Springer New York; 2004 [citado 9 de octubre de 2026]. Disponible en: <http://link.springer.com/10.1007/b97636> doi:10.1007/b97636
17. Shmueli G. To Explain or to Predict? *Stat Sci*. 1 de agosto de 2010;25(3). doi:10.1214/10-STS330
18. Tashman LJ. Out-of-sample tests of forecasting accuracy: an analysis and review. *Int J Forecast*. octubre de 2000;16(4):437-50. doi:10.1016/S0169-2070(00)00065-0
19. Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. *Int J Forecast*. octubre de 2006;22(4):679-88. doi:10.1016/j.ijforecast.2006.03.001
20. 5.10 Time series cross-validation | Forecasting: Principles and Practice (3rd ed) [Internet]. [citado 9 de octubre de 2026]. Disponible en: <https://otexts.com/fpp3/tscv.html>
21. Davison AC, Hinkley DV. Bootstrap Methods and their Application [Internet]. 1.<sup>a</sup> ed. Cambridge University Press; 1997 [citado 9 de octubre de 2026]. Disponible en: <https://www.cambridge.org/core/product/identifier/9780511802843/type/book> doi:10.1017/CBO9780511802843
22. Efron B. Bootstrap Methods: Another Look at the Jackknife. *Ann Stat*. 1 de enero de 1979;7(1). doi:10.1214/aos/1176344552
23. Shmueli G. To Explain or to Predict? *Stat Sci*. 1 de agosto de 2010;25(3). doi:10.1214/10-STS330
24. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning: with Applications in R [Internet]. New York, NY: Springer US; 2021 [citado 9 de octubre de 2026]. (Springer Texts in Statistics). Disponible en: <https://link.springer.com/10.1007/978-1-0716-1418-1> doi:10.1007/978-1-0716-1418-1
25. Iacobucci D, Schneider MJ, Popovich DL, Bakamitsos GA. Mean centering helps alleviate “micro” but not “macro” multicollinearity. *Behav Res Methods*. diciembre de 2016;48(4):1308-17. doi:10.3758/s13428-015-0624-x
26. Tashman LJ. Out-of-sample tests of forecasting accuracy: an analysis and review. *Int J Forecast*. octubre de 2000;16(4):437-50. doi:10.1016/S0169-2070(00)00065-0
27. Tashman LJ. Out-of-sample tests of forecasting accuracy: an analysis and review. *Int J Forecast*. octubre de 2000;16(4):437-50. doi:10.1016/S0169-2070(00)00065-0
28. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning: with Applications in R [Internet]. New York, NY: Springer US; 2021 [citado 13 de octubre de 2026]. (Springer Texts in Statistics). Disponible en: <https://link.springer.com/10.1007/978-1-0716-1418-1> doi:10.1007/978-1-0716-1418-1

29. Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. *Int J Forecast.* octubre de 2006;22(4):679-88. doi:10.1016/j.ijforecast.2006.03.001
30. Burnham KP, Anderson DR, editores. *Model Selection and Multimodel Inference* [Internet]. New York, NY: Springer New York; 2004 [citado 9 de octubre de 2026]. Disponible en: <http://link.springer.com/10.1007/b97636> doi:10.1007/b97636
31. Davison AC, Hinkley DV. *Bootstrap Methods and their Application* [Internet]. 1.<sup>a</sup> ed. Cambridge University Press; 1997 [citado 9 de octubre de 2026]. Disponible en: <https://www.cambridge.org/core/product/identifier/9780511802843/type/book> doi:10.1017/CBO9780511802843
32. Varma V, Bebbber DP. Climate change impacts on banana yields around the world. *Nat Clim Change.* octubre de 2019;9(10):752-7. doi:10.1038/s41558-019-0559-9
33. *Forecasting: Principles and Practice* [Internet]. [citado 13 de octubre de 2026]. Disponible en: <https://otexts.com/fpp3/>