



Modelación estocástica y componentes estructurales de la planta en la estimación no destructiva de la producción en *Coffea arabica*

Luis Carlos Imbachí Quinchúa

1er Simposio Colombo - Brasileño

Ecofisiología del Café

Avances científicos e innovaciones para enfrentar los desafíos actuales y futuros en los agrosistemas

Organizan



Apoyan





¿Por qué es importante estimar la carga de frutos en café?



Planificación de la cosecha

- Permite determinar el momento óptimo de la cosecha
- Facilita la organización de la cosecha
- Reduce riesgo de pérdidas por maduración



Mano de obra

- Permite estimar la cantidad de trabajadores necesarios
- Mejora la asignación de tareas y turnos
- Optimiza costos laborales



Comercialización

- Facilita la proyección de volúmenes disponibles
- Mejora la negociación con compradores
- Permite planificar entregas

Una estimación temprana y precisa de la carga de frutos es clave para una gestión eficiente y rentable de toda la cadena productiva del café



Método tradicional

Pesaje de frutos



Destructivos | Corte de ramas | Retiro de frutos

Altamente demandante de tiempo | Mano de obra considerable | dispendiosos en cultivos extensos

Eficiencia

Intervención no destructiva

Reducción de costos





Modelos estadísticos de rendimiento

Se desarrollaron modelos basados en variables fisiológicas, vegetativas y climáticas para explicar el rendimiento del café mediante regresiones lineales y análisis multivariados. Aunque proporcionaron interpretabilidad, asumían relaciones lineales y requerían un muestreo intensivo en campo.

Machine Learning en la agricultura

El aprendizaje automático permitió modelar relaciones no lineales e integrar simultáneamente variables climáticas, fisiológicas, edáficas y de manejo. Estos enfoques incrementaron la capacidad predictiva y redujeron la dependencia de supuestos estadísticos tradicionales.

ML aplicado a café

Algoritmos como Random Forest, SVM, Redes Neuronales y XGBoost mejoraron la estimación del rendimiento al combinar información climática, satelital, fisiológica y de manejo. Sin embargo, la mayoría de los estudios continúan dependiendo de variables indirectas o mediciones destructivas.

Modelación probabilística

Distribuciones Gamma, Weibull y Log-normal permiten representar la variabilidad inherente de la producción agrícola y cuantificar la incertidumbre del rendimiento. Estos modelos generan pronósticos probabilísticos más robustos para apoyar la toma de decisiones bajo condiciones variables.



Antecedentes

Su enfoque determinístico y lineal impulsó el desarrollo de métodos más flexibles basados en aprendizaje automático.

Hasta 2015

Castro-Tanzi et al. (2014)
Vaast et al. (2002)

Facilita el procesamiento de grandes volúmenes de datos provenientes de múltiples fuentes.

Hasta 2015-2020

Araújo et al. (2023)
Lionel et al. (2025)
Abreu Júnior et al. (2022)
Martello et al. (2022)

La integración de sensores remotos y aprendizaje automático permitió estimaciones más precisas del rendimiento, aunque persisten limitaciones para cuantificar directamente la carga de frutos.

Hasta 2018-2022

Martello et al. (2022)
Freitas et al. (2024)
Liakos et al. (2018)
Abreu Júnior et al. (2022b)

Incorporan la incertidumbre como parte del proceso de predicción, superando los enfoques determinísticos.

Hasta 2021-2024

Reynolds et al. (2021)
Montes-Pajuelo et al. (2024)
Tesfa et al. (2023)
Bak et al. (2025)



Fundamento fisiológico de la estimación de la carga de frutos

La carga de frutos del cafeto es el resultado del equilibrio entre la capacidad fotosintética de la planta, el desarrollo de su arquitectura productiva y la disponibilidad de fotoasimilados. Estos procesos fisiológicos sustentan el uso de variables morfológicas para la estimación no destructiva del rendimiento.

Formación foliar

El desarrollo del follaje determina la capacidad fotosintética y la disponibilidad de fotoasimilados para el crecimiento y llenado de los frutos. En consecuencia, una mayor área foliar suele asociarse con una mayor carga de frutos, siempre que no existan limitaciones ambientales. León-Burgos et al. (2024a,b); Almeida et al. (2021)

Crecimiento plagiotrópico

Las ramas plagiotrópicas constituyen la principal estructura productiva del cafeto, ya que en sus nudos se desarrollan flores y frutos. Su crecimiento incrementa el número de sitios potenciales de fructificación y explica parte de la variación en la carga de frutos. Unigarro et al. (2022); Rakocevic et al. (2023)

Relación área foliar / fruto

Es un parámetro útil que relaciona el balance de formación foliar para el desarrollo del fruto. Cuando la relación disminuye aumenta la competencia por fotoasimilados, afectando el crecimiento y desarrollo de los frutos. Bote y Jan (2016); DaMatta et al. (2008)

Fundamento para la estimación no destructiva

Las variables de formación foliar y crecimiento de ramas integran el efecto del ambiente sobre la planta y pueden utilizarse como predictores confiables de la carga de frutos mediante modelos de aprendizaje automático. Almeida et al. (2021); DaMatta et al. (2008)



1. Lugar del estudio

- Estación experimental de Naranjal – Cenicafé
- Chinchiná-Caldas
- 1381 msnm
- Variedad Castillo®Centro
- Inv. 2021 a 2024



2. Clima

- Temp: 21,35 °C
- PAR 383,72 $\mu\text{mol m}^{-2} \text{s}^{-1}$
- 1573 h de luz / año
- 2821 mm precipitación / año
- Datos red Agroclima



3. Establecimiento de cultivo

- Siembra: Marzo 2021
- 1,25 x 1,0 m
- 8000 plantas ha^{-1}
- Manejo agronómico Cenicafé
- Cultivo a libre exposición solar



4. Diseño experimental

- 40 parcelas
- Area total: 2200 m^2
- 52 m^2 / parcela
- 30 plantas efectivas



5. Evaluaciones

- Después de 16 meses se siembra-Inicio de etapa productiva
- 2022 - 2024
- Los puntos de tiempo de medición para cada año se definieron con base en los eventos de floración
- Las etapas fenológicas del desarrollo del fruto del café



Medidas del crecimiento evaluadas

Diseño

- 120 plantas
- 3 plantas seleccionadas aleatoriamente por parcela

Selección ramas

- Por cada planta y año se seleccionaron 2 ramas plagiotrópicas con el mayor número de frutos en el tercio productivo de la copa

Muestro

- Muestreo entre junio y julio.
- Representa mas del 80% de frutos cuajados

Años de estudio

- 2022-2024: evaluación de producción y desarrollo.
- 2024: Medidas vegetativas complementarias al crecimiento de la planta.

Longitud de la rama



Se recolectaron 100 frutos de la cosecha total por parcela (4000)/año



Área foliar de cada hoja (ELA)

Unigarro et al. 2015

L es la longitud de la hoja y W es el ancho de la hoja.

$$ELA = 0.99927 * (L * (-0.14757 + 0.60986 * W))$$

Variables evaluadas

- **LR:** Longitud de la rama (cm)
- **NTN:** Número total de nudos por rama
- **NTL:** Número total de hojas por rama
- **AFA:** Área foliar acumulada por rama (cm²)
- **RAF:** Relación área foliar por fruto (cm² Fruto⁻¹)
- **NNP:** Número de nudos productivos por rama
- **FB:** Número de frutos por rama
- **MF:** Masa por fruto (g)

Se suman las ELA de todas las hojas de la rama

=AFA



Componentes estructurales de la planta



Estimación no destructiva de la producción en
Coffea arabica

Modelos evaluados

Entradas del modelo: medidas de crecimiento (2022-2024)

LR | NTN | NTL | AFA | RAF | NNP

Variable respuesta
FB

SVR

Encuentra la función optima que relaciona las variables de crecimiento con FB utilizando un Kernel

Ventajas

- Modela relaciones altamente no lineales
- Gran capacidad de aprendizaje

ANN

Aprende relaciones no lineales mediante capas de neuronas y retropropagación

Ventajas

- Robusto frente al sobreajuste
- Estimar la importancia de las variables

RF

Combina múltiples arboles de decisión contruidos con muestras aleatorias para mejorar la predicción

Ventajas

- Alta precisión en relaciones complejas
- Buen desempeño con pocos datos

SVR: Hiperparámetros (C y γ) optimizados con Kernel radial en el bucle interno

ANN: Normalización de las variables de entrada y una arquitectura multicapa con seis neuronas ocultas

RF: 1000 árboles y numero de variables seleccionadas aleatoriamente optimizado en el bucle interno

Validación

K-folds

k-fold interno para el ajuste de hiperparámetros

k-fold externo para la evaluación del modelo

Evaluación de rendimiento RMSE y R^2



De una predicción puntual a una predicción probabilística

Los modelos de Machine Learning tradicionales hacen estimaciones puntuales

¿Por qué estudiar el error?

- variabilidad biológica
- clima
- manejo
- errores de medición
- heterogeneidad del cultivo

En otras palabras,
todo aquello que el modelo no pudo explicar.

realidad: Pasar de un dato estático a un escenario probabilístico, permitiendo que las decisiones agronómicas se tomen conociendo el margen de riesgo o incertidumbre.

De esta forma, el modelo final no solo predice lo que una persona contó en campo, sino que predice la carga de frutos después de haber incorporado el margen de error del conteo humano y la variabilidad del cultivo

Ahora ya no existe una estimación puntual.

Existe una distribución de posibles cargas.

la simulación de producción de café cereza por árbol

Para cada año productivo se realiza la simulación por árbol

Salida final: Producción promedio de café cereza por árbol

Variables usadas en la simulación

- Contribución de la carga de frutos de las dos ramas mas productivas al árbol (CPBT)
- Masa por fruto
- Carga de frutos

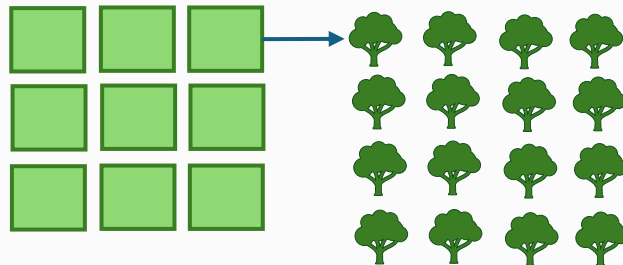
Distribuciones candidatas
Normal | Beta | Gamma | Weibull

Ajuste de distribuciones (Cullen & Frey); AIC, BIC; Anderson Darling al 5%

Simulación Monte Carlo

- Se generan valores aleatorios para cada variable desde las distribuciones ajustadas
- Permite reflejar la variabilidad e incertidumbre en la producción

Adaptado a 40 parcelas y cada una con 30 arboles cada una (1200 Arboles de café)



Estructura de simulación



Simular la carga de frutos por rama

Frutos + incertidumbre (ML)



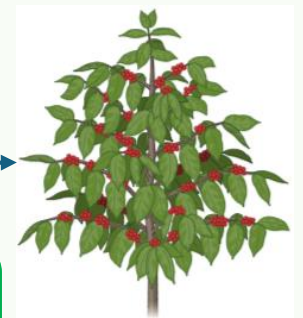
Asignar a cada fruto de café un peso

Tenemos la producción por rama

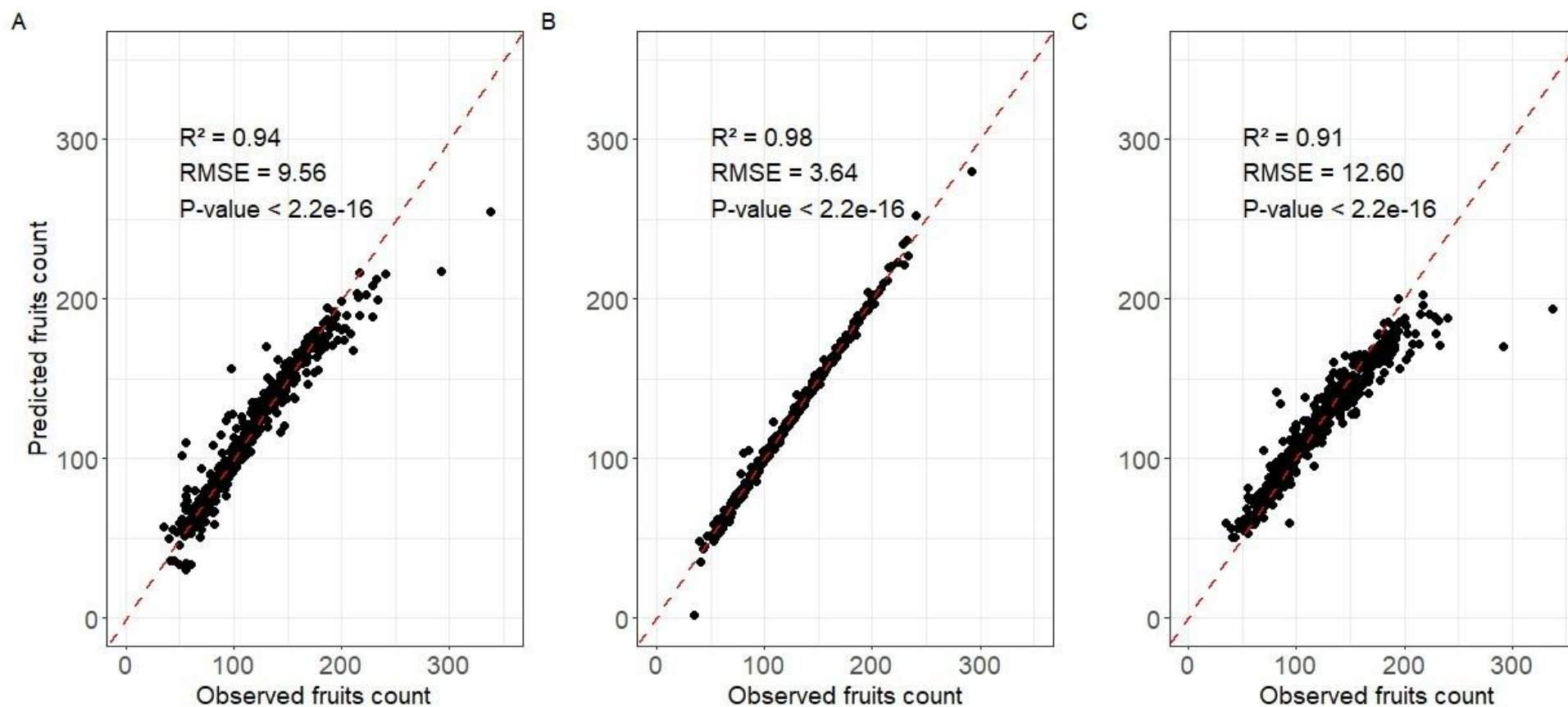
La producción por rama

CPBT

Pr1 + Pr2
CPBT



Modelación del número frutos por rama con base a variables de crecimiento



Resultados

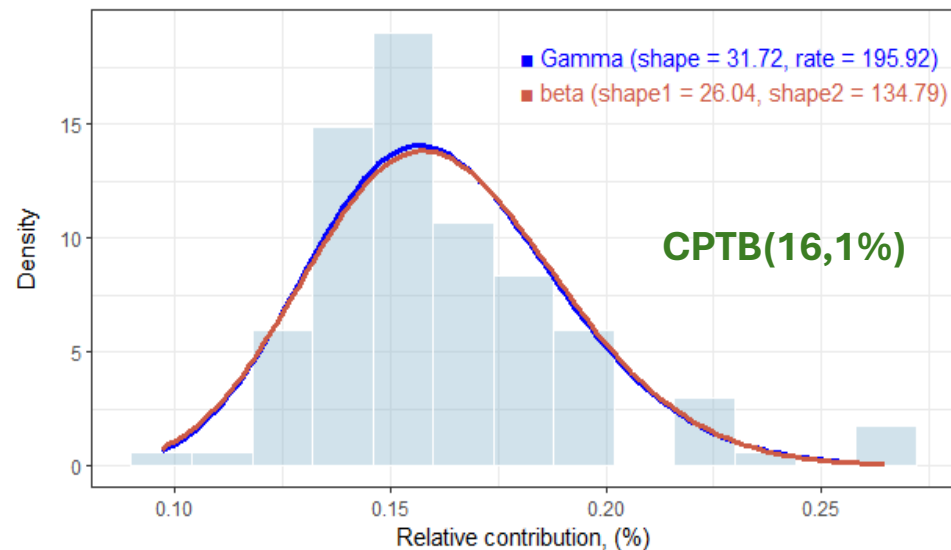
Resultados de la validación cruzada de las técnicas de aprendizaje automático evaluadas para la predicción de la carga de fruta a través del número de frutos por rama en los tres años productivos. **(A) Regresión de vectores de soporte (SVR), (B) red neuronal artificial (ANN) y (C) bosque aleatorio (RF).**

Ajuste de distribuciones para el número de frutos por rama y peso del fruto cereza

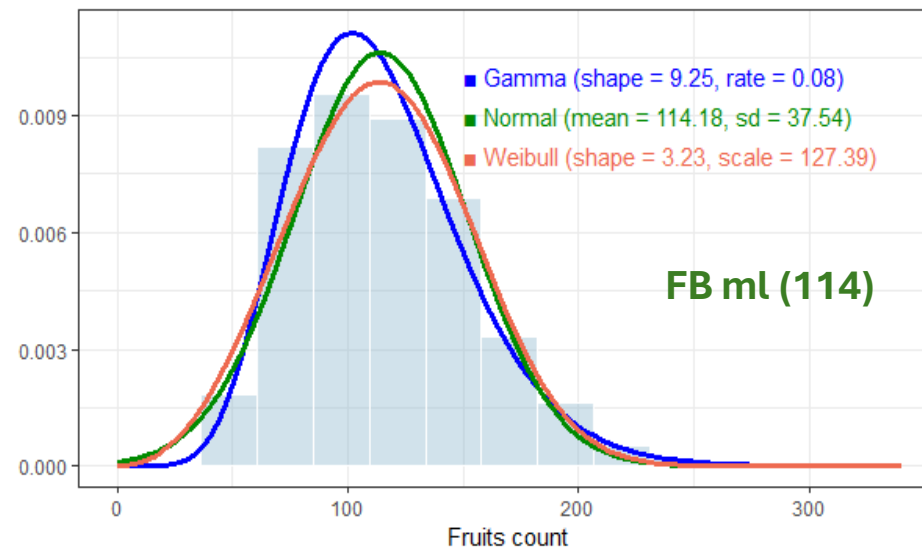
Variable	Distribución	Parámetro	AIC	BIC	Anderson–Darling Value	P value
CPBT _i	Gamma	$\alpha = 31.72$ $\lambda = 195.92$	-509.83	-504.25	0.092073	0.261
	Beta	$\alpha = 26.04$ $\beta = 134.79$	-507.23	-501.66	0.09638	0.214
FB _{ml}	Gamma	$\alpha = 9.25$ $\lambda = 0.081$	5742.15	5750.85	0.97083	0.3729
	Weibull	$\lambda = 3.23$ $k = 127.39$	5785.03	5793.70	3.3804	0.01763
	Normal	$\mu = 114.18$ $\sigma = 37.54$	5780.23	5788.93	3.2943	0.001946
FM	Gamma	$\alpha = 41.33$ $\lambda = 18.97$	3198.76	3211.74	2.712	0.05
	Weibull	$\lambda = 6.63$ $k = 2.32$	3730.53	3743.5	5.9571	0.001
	Normal	$\mu = 2.17$ $\sigma = 0.33$	3236.27	3249.25	46.534	1.23E-07

Resultados

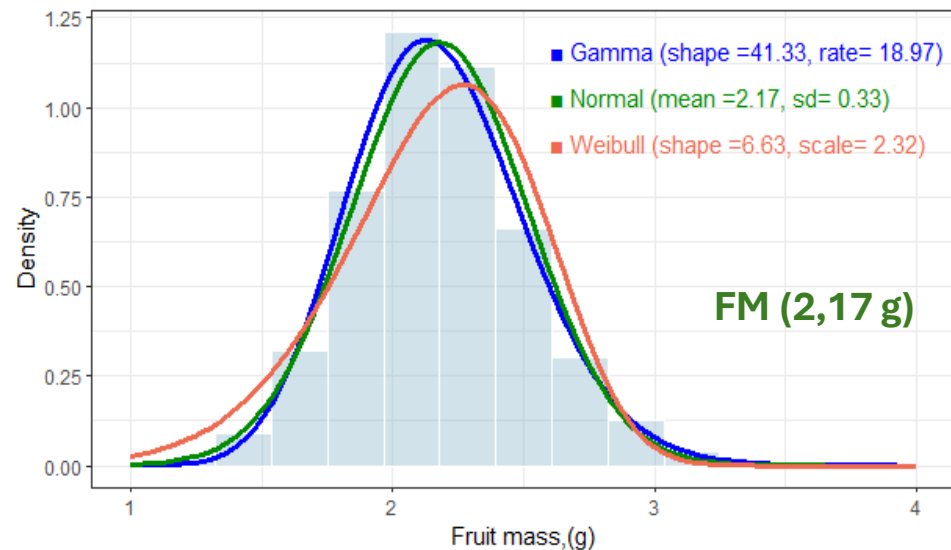
A



B



C



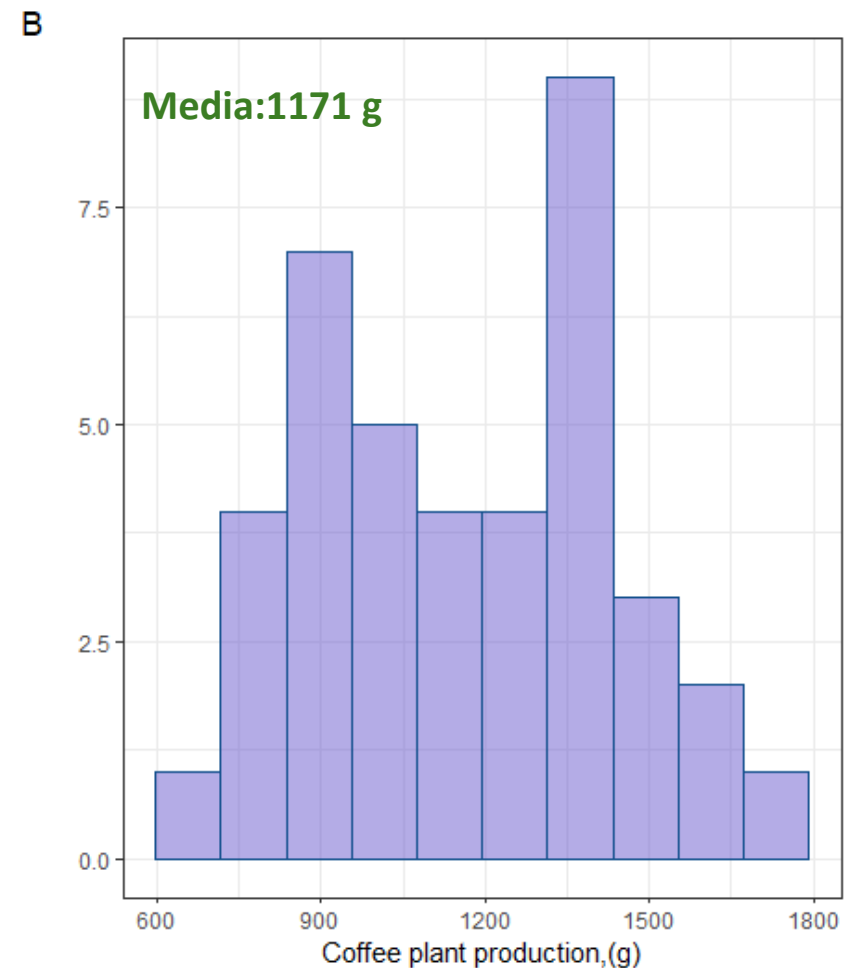
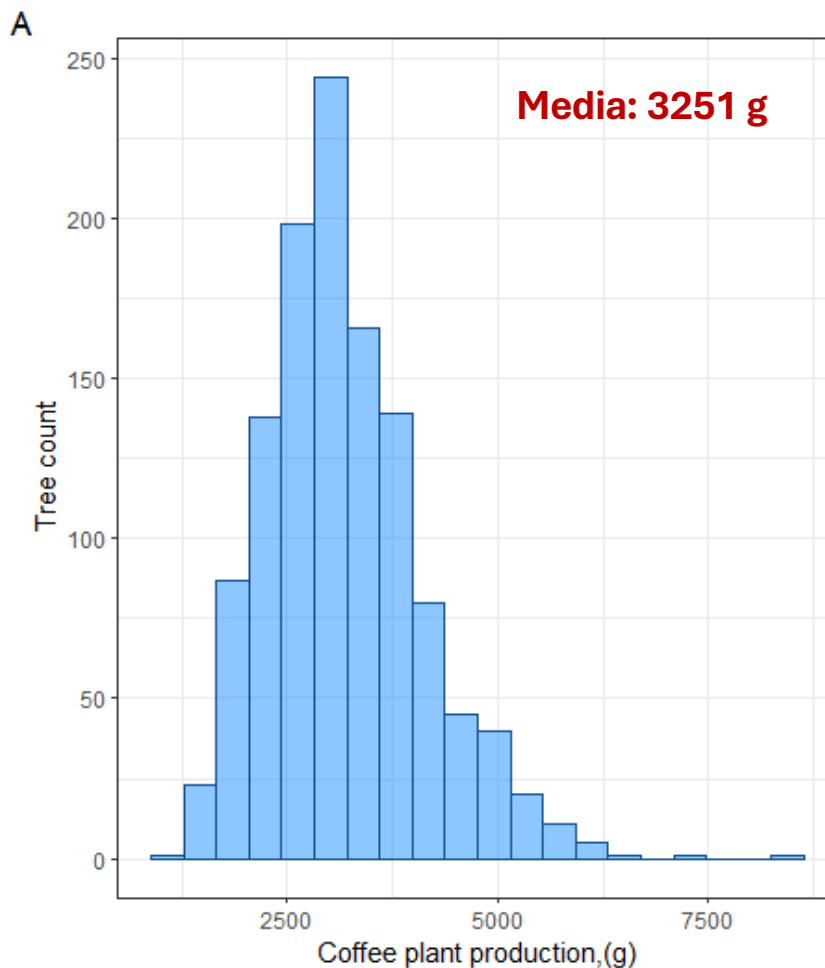
Densidad y ajuste de las distribuciones probabilísticas que indican el comportamiento de **la contribución de las ramas productivas evaluadas en el árbol** (CPBT_i) (A), **el número de frutos por rama** predicho por Aprendizaje Automático (FB_{ml}) (B) y **la masa individual del fruto** (FW) (C) para el primer y segundo año de producción de cultivos.



Resultados



La distribución de la producción simulada osciló entre 735 g y 7000g



La distribución de la producción observada osciló entre 600 g y 1800 g

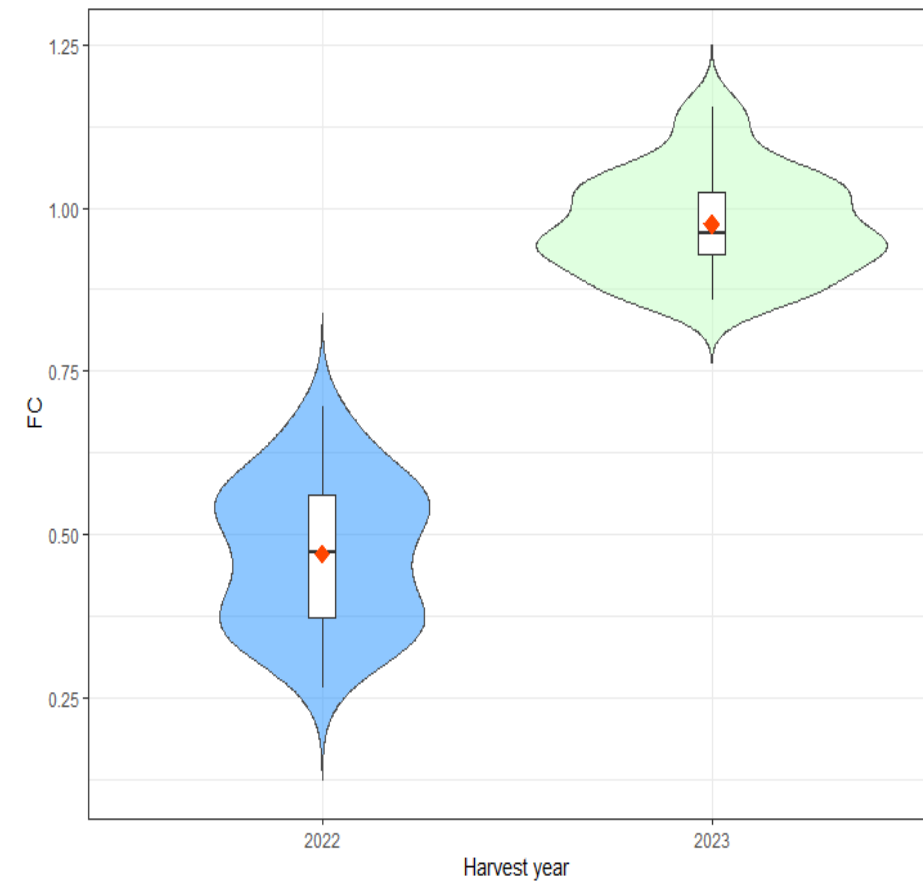
Histogramas que presentan la distribución por árbol en las producciones simuladas (barras azules) (A) y observadas (barras moradas) (B) para el primer año de producción.



Factor de corrección de producción simulada

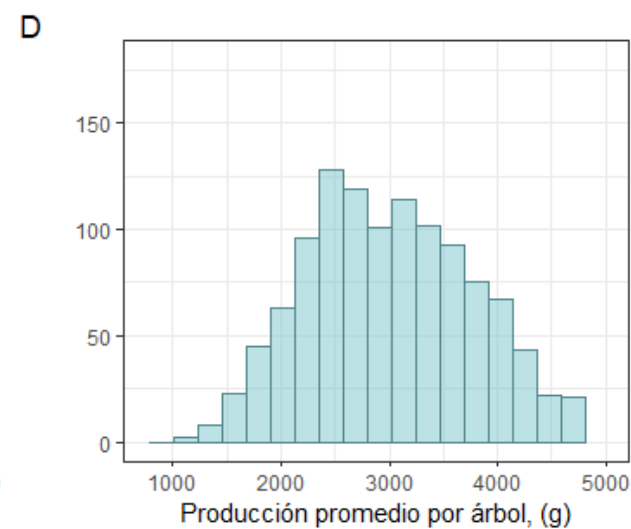
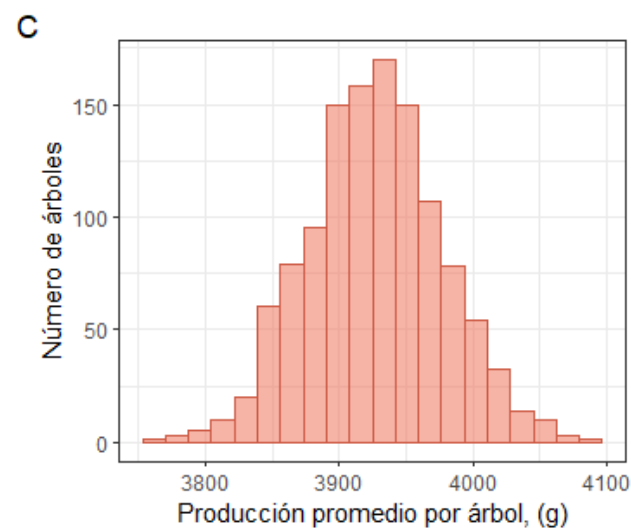
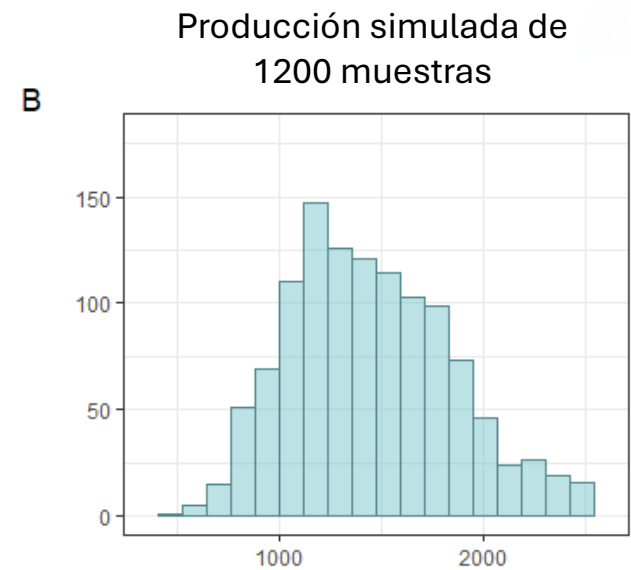
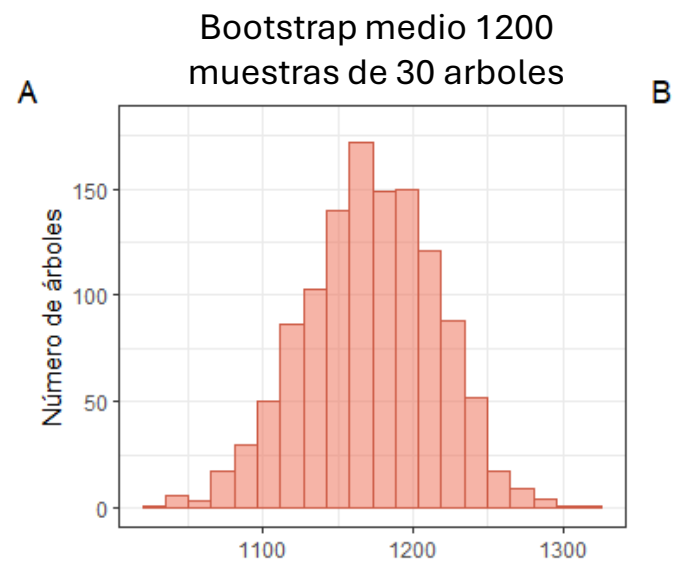
$$\overline{FC} = \frac{1}{m} \sum_{i=1}^m \left(\frac{\frac{P_i}{\overline{R}}}{\frac{1}{n} \sum_{j=1}^n N_{ij}} \right)$$

$$FC = \frac{\text{Producción promedio por rama}}{\text{Promedio de frutos por rama}}$$





Simulación de producción de café cereza por árbol ajustada





Conclusiones

- La integración de técnicas de aprendizaje automático con un enfoque probabilístico sugiere potencial para predecir la carga de frutos y simular los rendimientos de los cultivos de café a partir de mediciones vegetativas no destructivas.
- Entre los modelos evaluados, el enfoque de redes neuronales artificiales (ANN) mostró el mejor desempeño para estimar la carga de frutos y corregir el error de estimación residual a partir de mediciones aéreas del crecimiento vegetativo en plantas de la variedad de café Castillo® Centro.
- En términos generales, estos hallazgos resaltan la dificultad inherente de representar sistemas de producción biológica complejos mediante supuestos estructurales simplificados, un desafío común a todas las especies de cultivos y entornos de producción.
- Futuros estudios deberían abordar la incorporación de información complementaria a nivel de árbol y parcela para caracterizar la variabilidad espacial y productiva con mayor detalle, lo que podría contribuir a mejorar los procesos de calibración y a una representación más precisa de las condiciones de campo.



Agradecimientos

Dirección de Cenicafé
Gerencia Técnica
Carlos Andrés Unigarro Muñoz.
Andrés Felipe León Burgos.
Biometría



<https://doi.org/10.1016/j.atech.2026.102256>



Cenicafé®
Centro Nacional de Investigaciones de Café



cenicafe.org



Cenicafé FNC



@cenicafe



cenicafé



CenicaféFNC



@cenicafefnc