

Modelos predictivos de datos meteorológicos mediante machine learning y su impacto en la agroproducción de la provincia de Bolívar, Ecuador

Predictive models of meteorological data through machine learning and their impact on the agroproduction of the Bolívar province, Ecuador

- ¹ Deysi Margoth Guanga Chunata  <https://orcid.org/0000-0002-6548-5585>
Superior Politécnica de Chimborazo (ESPOCH), Riobamba, Ecuador.
Maestría en estadística con mención en ciencia de datos e inteligencia artificial
deysi.guanga@espoch.edu.ec
Universidad Estatal de Bolívar (UEB), Guaranda, Ecuador.
dguanga@ueb.edu.ec
- ² Andrés Alejandro Galvis Correa  <https://orcid.org/0000-0001-7762-2893>
Superior Politécnica de Chimborazo (ESPOCH), Riobamba, Ecuador.
andres.galvis@espoch.edu.ec
Pontificia Universidad Católica del Ecuador (PUCE) Quito, Ecuador.
a.galvis@puce.edu.ec



Artículo de Investigación Científica y Tecnológica

Enviado: 09/05/2026

Revisado: 11/06/2026

Aceptado: 10/07/2026

Publicado: 17/07/2026

DOI: <https://doi.org/10.33262/concienciadigital.v9i3.3705>

Cítese:

Guanga Chunata, D. M., & Galvis Correa, A. A. (2026). Modelos predictivos de datos meteorológicos mediante machine learning y su impacto en la agroproducción de la provincia de Bolívar, Ecuador. *ConcienciaDigital*, 9(3), 60-77.
<https://doi.org/10.33262/concienciadigital.v9i3.3705>



CONCIENCIA DIGITAL, es una revista multidisciplinar, **trimestral**, que se publicará en soporte electrónico tiene como **misión** contribuir a la formación de profesionales competentes con visión humanística y crítica que sean capaces de exponer sus resultados investigativos y científicos en la misma medida que se promueva mediante su intervención cambios positivos en la sociedad. <https://concienciadigital.org>
La revista es editada por la Editorial Ciencia Digital (Editorial de prestigio registrada en la Cámara Ecuatoriana de Libro con No de Afiliación 663) www.celibro.org.ec



Esta revista está protegida bajo una licencia Creative Commons en la 4.0 International. Copia de la licencia: <http://creativecommons.org/licenses/by-nc-sa/4.0/>



Palabras claves:

Series temporales multivariadas, procesos autorregresivos vectoriales, inferencia bayesiana, aprendizaje profundo.

Keywords:

Multivariate time series, vector autoregressive processes, Bayesian inference, recurrent deep learning.

Resumen

Introducción. La no estacionariedad inherente a las series climáticas andinas limita la eficacia de los modelos lineales clásicos. **Objetivo.** Desarrollar un marco metodológico comparativo que integre modelos estadísticos tradicionales (ARIMA/SARIMA) con algoritmos de Ensemble Learning (Random Forest, XGBoost) y redes neuronales profundas (LSTM) para la predicción de variables meteorológicas, evaluando su impacto causal en la agroproducción mediante modelado VAR. **Metodología.** Estudio cuantitativo longitudinal con 1.642.217 observaciones (2016-2023). Se aplicó preprocesamiento robusto (Isolation Forest, KNN Imputer) y validación temporal *walk-forward*. El análisis multivariante utilizó pruebas de raíz unitaria (ADF), causalidad de Granger y descomposición de varianza del error de pronóstico (FEVD). **Resultados.** Random Forest superó significativamente a SARIMA (RMSE 0,74 vs 1,66; R^2 0,67 vs -0,63). El modelo VAR (6) evidenció causalidad de Granger significativa ($p < 0,001$) entre precipitación/temperatura y rendimiento del maíz, con un rezago dinámico de 6 meses. **Conclusión.** Los métodos no paramétricos capturan mejor la heterocedasticidad condicional de los datos climáticos, proporcionando herramientas robustas para la agricultura de precisión. **Área de estudio general:** Ciencia de Datos e Inteligencia Artificial. **Área de estudio específica:** Modelamiento predictivo agroclimático. **Tipo de estudio:** Artículos originales.

Abstract

Introduction. The inherent non-stationarity of Andean climate series limits the efficacy of classical linear models. **Objective.** To develop a comparative methodological framework integrating traditional statistical models (ARIMA/SARIMA) with Ensemble Learning algorithms (Random Forest, XGBoost) and deep neural networks (LSTM) for meteorological variable prediction, evaluating their causal impact on agroproduction via VAR modeling. **Methodology.** Quantitative longitudinal study with 1,642,217 observations (2016-2023). Robust preprocessing (Isolation Forest, KNN Imputer) and walk-forward temporal validation were applied. Multivariate analysis utilized unit root tests (ADF), Granger causality, and Forecast Error Variance Decomposition (FEVD). **Results.** Random Forest significantly

outperformed SARIMA (RMSE 0.74 vs 1.66; R^2 0.67 vs -0.63). The VAR (6) model evidenced significant Granger causality ($p < 0,001$) between precipitation/temperature and corn yield, with a dynamic lag of 6 months. **Conclusion.** Non-parametric methods better capture the conditional heteroscedasticity of climate data, providing robust tools for precision agriculture. **General Area of Study:** Data Science and Artificial Intelligence. **Specific area of study:** Agrocimatic predictive modeling. **Type of study:** Original articles.

1. Introducción

El cambio climático y la variabilidad meteorológica representan dos de los mayores desafíos estocásticos para la sostenibilidad de los sistemas agroproductivos en el siglo XXI. En regiones de topografía compleja como la provincia de Bolívar, Ecuador, estos efectos se amplifican debido a la influencia no lineal de patrones climáticos de gran escala, como el Fenómeno de El Niño Oscilación del Sur (ENSO) (Serrano-Vincenti et al., 2025). La agroproducción local, dependiente de cultivos estratégicos como el maíz y el cacao, enfrenta una alta volatilidad derivada de la irregularidad en los ciclos de precipitación y temperatura (Instituto Nacional de Estadística y Censos [INEC], 2023). La agricultura climáticamente inteligente (Climate Smart Agriculture) requiere, por tanto, capacidades predictivas robustas que superen la mera extrapolación de promedios históricos (World Bank Group, 2020; Food and Agriculture Organization of the United Nations [FAO], 2022). Tradicionalmente, la modelación de series temporales climáticas se ha basado en enfoques lineales paramétricos, como los modelos Autorregresivos Integrados de Promedio Móvil (ARI-MA) y sus extensiones estacionales (SARIMA) (Hyndman & Athanasopoulos (2021). Estos modelos, fundamentados en la teoría de procesos estocásticos estacionarios, asumen homocedasticidad y normalidad en los residuos, supuestos que frecuentemente se violan en series meteorológicas tropicales debido a la presencia de tendencias estructurales, heterocedasticidad condicional y cambios de régimen abruptos (Rolnick et al., 2023). Esta limitación teórica motiva la exploración de métodos no paramétricos y de aprendizaje automático capaces de capturar relaciones funcionales complejas sin restricciones de forma previa (Liakos et al., 2018). En el ámbito de la inteligencia artificial, los algoritmos de Ensemble Learning, como Random Forest y XGBoost, han demostrado superioridad en tareas de regresión no lineal al combinar múltiples estimadores débiles para reducir la varianza y el sesgo mediante técnicas de bagging y boosting respectivamente (Bishop, 2006; Chen & Guestrin, 2016). XGBoost, en particular, utiliza una aproximación de Taylor de segundo orden para minimizar una función de pérdida regularizada, lo que permite una optimización convexa

eficiente incluso en espacios de características de alta dimensionalidad (Chen & Guestrin, 2016). Paralelamente, el aprendizaje profundo (Deep Learning), específicamente las redes de Memoria a Largo Plazo (LSTM) y las Unidades Recurrentes Gated (GRU), ofrece un marco matemático para modelar dependencias temporales de largo alcance mediante mecanismos de puertas gating que mitigan el problema del gradiente desvaneciente inherente a las RNN tradicionales (Hochreiter & Schmidhuber, 1997); Goodfellow et al., 2016). Estudios recientes como los de Xie et al. (2024) y Huntingford et al. (2019) confirman que estas arquitecturas son superiores para capturar la dinámica no lineal de sistemas terrestres complejos.

La aplicación de estas tecnologías en la agricultura ha sido ampliamente revisada. Kumar et al. (2025) y Kulyal & Saxena (2022) destacan el potencial del machine learning para la predicción de rendimientos, mientras que Benos et al. (2021) y Kuradusenge et al. (2023) evidencian la superioridad de los modelos híbridos sobre los estadísticos clásicos. Morales & Villalobos (2023) y Belgiu & Dragut (2016) subrayan la importancia de la selección de características y la robustez de los bosques aleatorios en contextos de sensores remotos y datos ambientales. Además, Reichstein et al. (2019) argumenta que la integración del conocimiento de procesos físicos con modelos datadriven es crucial para la ciencia del sistema terrestre.

En el contexto ecuatoriano, Serrano-Vincenti et al. (2025) analizaron el impacto del cambio climático en la agricultura del país, identificando vulnerabilidades específicas en las regiones andinas. Instituto Nacional de Meteorología e Hidrología (INAMHI, 2020) proporcionan los anuarios meteorológicos oficiales que sirven como base de datos para este estudio. Por su parte el Instituto Nacional de Estadística y Censos (INEC, 2023) publican la Encuesta de Superficie y Producción Agropecuaria Continua (ESPAC), fuente primaria de los datos de rendimientos agrícolas. El Plan Nacional de Desarrollo destaca la urgencia de fortalecer la resiliencia ambiental y potenciar la economía social y solidaria a través de la transferencia tecnológica al sector rural (Consejo Nacional de Planificación, 2021).

La presente investigación aborda la hipótesis de que la implementación de algoritmos de Deep Learning y Ensemble Learning mejora significativamente la precisión predictiva de las variables meteorológicas en la provincia de Bolívar en comparación con los modelos estadísticos tradicionales (ARIMA). Asimismo, se plantea que existe una correlación multivariante significativa entre las anomalías meteorológicas predichas y los rendimientos de los cultivos principales, permitiendo estimar impactos en la agroproducción nacional mediante modelos autorregresivos vectoriales (VAR) (Hyndman & Athanasopoulos, 2021).

El objetivo general de esta investigación es desarrollar y comparar modelos predictivos basados en técnicas de Ciencia de Datos e Inteligencia Artificial para el análisis de series de tiempo meteorológicas en la provincia de Bolívar, evaluando su desempeño mediante métricas de error y su impacto potencial en la planificación de la agroproducción nacional. Los objetivos específicos incluyen: (1) recopilar y preprocesar datos históricos

meteorológicos (INAMHI) y agroproductivos (INEC/MAG) de la provincia de Bolívar; (2) aplicar análisis multivariante y algoritmos de Machine Learning (Redes Neuronales Recurrentes, Random Forest) para la predicción de variables climáticas; y (3) evaluar el desempeño de los métodos mediante métricas de comparación (RMSE, MAE) y modelar la relación entre las predicciones óptimas y los índices de producción agrícola.

2. Metodología

El presente estudio se inscribe dentro de un enfoque cuantitativo, tecnológico y explicativo, fundamentado en la aplicación rigurosa de métodos matemáticos y estadísticos avanzados para el procesamiento de series temporales de alta frecuencia y la inferencia causal multivariante (Liakos et al., 2018). El diseño de investigación posee las siguientes características estructurales: no experimental, dado que no existe manipulación deliberada de las variables independientes (parámetros meteorológicos físicos); longitudinal (series de tiempo), al analizar secuencias de datos observadas en intervalos regulares desde 2016 hasta 2023 para capturar la dinámica temporal y estacionalidad (Hyndman & Athanasopoulos, 2021); y comparativo, al contrastar el desempeño predictivo de paradigmas algorítmicos diversos (estadística clásica, Ensemble Learning y Deep Learning).

2.1. Población y muestra

La población de estudio está constituida por el universo de registros históricos meteorológicos y agroproductivos oficiales de la provincia de Bolívar. La recolección de información se realiza mediante un muestreo intencional retrospectivo para consolidar la base empírica:

- a) **Datos meteorológicos:** registros históricos provenientes de la red de estaciones meteorológicas automáticas de la provincia de Bolívar, administradas por el Instituto Nacional de Meteorología e Hidrología (INAMHI, 2020) y la Universidad Estatal de Bolívar (UEB). El conjunto de datos base cuenta con 1.642.217 observaciones registradas entre enero de 2016 y agosto de 2023, con una resolución temporal variable que evoluciona de 15 minutos a 1 minuto según la configuración del sensor.
- b) **Datos agroproductivos:** series anuales e históricas de superficie cosechada, volumen de producción y rendimientos agrícolas (medidos en toneladas por hectárea, Ton/Ha) para cultivos estratégicos como el maíz, la cebada, el trigo y el cacao. Estas series se extraen de la Encuesta de Superficie y Producción Agropecuaria Continua (ESPAC) del Instituto Nacional de Estadística y Censos (INEC, 2023).

2.2. Fase 1: Ingeniería de datos y preprocesamiento

Debido a la naturaleza ruidosa de los sensores automáticos expuestos a perturbaciones ambientales, se ejecuta un riguroso protocolo de preparación de datos basado en técnicas robustas de limpieza e imputación (Pedregosa et al., 2011):

2.2.1. Recolección y extracción (ETL)

Se diseña un pipeline de Extracción, Transformación y Carga (ETL) en lenguaje Python para consolidar los archivos planos del sensor climático. Las variables se indexan temporalmente mediante la combinación de fecha y hora local, resolviendo la variabilidad

en la frecuencia de muestreo mediante remuestreo temporal a escala diaria para evitar sesgos de alta frecuencia.

2.2.2. Limpieza y tratamiento de outliers

El filtrado de valores atípicos (*outliers*) se realiza bajo dos enfoques complementarios:

- a) **Filtro Unidimensional (IQR):** para variables físicas de comportamiento acotado (como la temperatura), se eliminan los valores que superan los límites del rango intercuartílico

(Ecuación 1):

$$[\text{Límite Inferior}, \text{Límite Superior}] = [Q1 - 1,5 \cdot IQR, Q3 + 1,5 \cdot IQR] \quad (1)$$

Donde $IQR = Q3 - Q1$ siendo $Q1$ y $Q3$ los percentiles 25 y 75 respectivamente.

- b) **Filtro Multidimensional (Isolation Forest):** identifica anomalías estructurales evaluando la combinación de múltiples variables simultáneamente (e.g., inconsistencias termodinámicas entre presión, humedad y temperatura), basándose en la profundidad promedio de aislamiento en subespacios aleatorios (Pedregosa et al., 2011).

2.2.3. Imputación de datos faltantes

Para preservar la autocorrelación de las series de tiempo se implementan dos estrategias:

- a) **Interpolación Lineal:** aplicada a vacíos de información cortos (< 3 registros consecutivos)

(Ecuación 2):

$$y_t = y_{t-1} + \frac{t - t_{-1}}{t_{+1} - t_{-1}} (y_{t+1} - y_{t-1}) \quad (2)$$

- b) **K-Nearest Neighbors Imputer (KNN Imputer):** para ausencias prolongadas, se estima el valor faltante calculando el promedio ponderado de los K vecinos más cercanos en el espacio multidimensional de las variables meteorológicas, preservando la estructura de covarianza local (Pedregosa et al., 2011).

2.2.4. Ingeniería de características (Feature Engineering)

A partir de las variables físicas primarias se construyen nuevas variables derivadas con relevancia agrobiológica:

- a) **Grados Día de Crecimiento (GDD):** índice térmico acumulado que mide el desarrollo fisiológico de los cultivos:

(Ecuación 3):

$$GDD = \sum \max\left(\frac{T_{max} + T_{min}}{2} - T_{base}, 0\right) \quad (3)$$

Donde T_{base} es la temperatura umbral (e.g., 10°C para maíz).

- b) **Variables Lag (Rezagos):** Creación de variables rezagadas ($Y_{t-1}, Y_{t-2}, \dots$) para permitir que los modelos no lineales capturen la memoria del sistema dinámico.

2.2.5. Normalización

Para garantizar la convergencia óptima de los algoritmos de redes neuronales, las variables continuas se estandarizan mediante la escala Z-score:

(Ecuación 4):

$$z_t = \frac{x_t - \mu}{\sigma} \quad (4)$$

2.3. Fase 2: Modelamiento predictivo con inteligencia artificial

Esta fase abarca la implementación y evaluación comparativa de arquitecturas predictivas bajo una estrategia de validación *Walk-Forward* para evitar data leakage (Hyndman & Athanasopoulos, 2021).

2.3.1. Modelos implementados

- Línea Base Estadística:** Modelos ARIMA y SARIMA (Hyndman & Athanasopoulos, 2021), fundamentados en la teoría de procesos estocásticos estacionarios.
- Ensemble Learning:** Random Forest (Bishop, 2006), que combina múltiples árboles de decisión mediante bagging, y XGBoost (Chen & Guestrin, 2016), que utiliza una aproximación de Taylor de segundo orden para minimizar una función de pérdida regularizada mediante descenso de gradiente.
- Deep Learning:** Redes LSTM (Hochreiter & Schmidhuber, 1997), y GRU (Goodfellow et al., 2016), diseñadas para modelar dependencias temporales de largo alcance mediante mecanismos de puertas gating que mitigan el problema del gradiente desvaneciente.

2.3.2. Optimización de hiperparámetros

La calibración fina de hiperparámetros (tasa de aprendizaje, número de estimadores, profundidad de árboles, unidades ocultas) se ejecuta mediante la librería Optuna, empleando un esquema de Optimización Bayesiana para minimizar el error cuadrático medio de validación cruzada.

2.3.3. Métricas de evaluación de desempeño

La comparación del rendimiento predictivo se cuantifica mediante tres métricas de error ampliamente aceptadas en climatología (Hyndman & Athanasopoulos, 2021):

(Ecuación 5):

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (5)$$

(Ecuación 6):

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (6)$$

(Ecuación 7):

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (7)$$

Donde y_i es el valor real observado, $\hat{y}_i$ es la predicción del modelo y N es el número total de observaciones.

2.4. Fase 3: *Análisis multivariante y correlación agroclimática*

La última fase evalúa la aplicabilidad y el impacto agronómico de las predicciones meteorológicas sobre la agroproducción andina de Bolívar.

2.4.1. *Análisis exploratorio multivariante (EDA)*

Se calculan los coeficientes de correlación de Pearson y Spearman entre las anomalías meteorológicas predichas y las series de producción de la ESPAC para identificar dependencias lineales y monotónicas. La correlación de Spearman viene dada por:

(Ecuación 8):

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (8)$$

2.4.2. *Modelado de impacto y rezagos dinámicos*

Dado que el impacto del clima sobre los rendimientos agrícolas ocurre con retrasos significativos, se implementa un Modelo Autorregresivo Vectorial (VAR) (Hyndman & Athanasopoulos, 2021):

(Ecuación 9):

$$X_t = c + A_1 X_{t-1} + A_2 X_{t-2} + \dots + A_p X_{t-p} + \epsilon_t \quad (9)$$

Donde X_t es un vector multivariante que contiene tanto las anomalías climatológicas predichas como los rendimientos históricos agrícolas.

2.4.3. *Pruebas de hipótesis y diagnóstico de residuos*

Para garantizar la validez inferencial, se aplican pruebas estadísticas bajo un nivel de significancia de $\alpha = 0,05$: significancia de coeficientes (p-value), autocorrelación de residuos (pruebas de Durbin-Watson y Breusch-Godfrey), y heterocedasticidad (prueba ARCH-LM). La selección del orden óptimo p del modelo VAR se realiza mediante el criterio de información de Akaike (AIC).

3. Resultados

A continuación, se presentan los hallazgos derivados de la aplicación del marco metodológico propuesto sobre el conjunto de datos meteorológicos y agroproductivos de

la provincia de Bolívar. La exposición de los resultados se estructura en tres ejes fundamentales:

- Primero, la caracterización estadística y temporal de la base de datos depurada;
- Segundo, la evaluación comparativa del desempeño predictivo entre los modelos lineales tradicionales y las arquitecturas de inteligencia artificial (Ensemble Learning y Deep Learning); y
- Tercero, el análisis multivariante mediante modelos VAR, que permite cuantificar la causalidad de Granger y los rezagos dinámicos entre las anomalías climáticas y el rendimiento del maíz.

Estos hallazgos proporcionan la evidencia empírica necesaria para validar la hipótesis de investigación y demostrar la utilidad de las predicciones para la agricultura de precisión.

3.1. Caracterización de la base de datos

El análisis exploratorio de 1.642.217 registros reveló una evolución en la frecuencia de muestreo, pasando de intervalos de 15 minutos (2016) a 1 minuto (2018-2021), lo que requirió un remuestreo diario para homogeneizar la serie. La **Tabla 1** resume la cobertura temporal.

Tabla 1

Resumen de la cobertura temporal y frecuencia de la base de datos

Año	Registros Totales	Días Únicos	Promedio/Día	Intervalo (min)
2016	22.410	242	92,6	15,6
2019	416.361	303	1.374,10	1
2023	122.984	232	530,1	2,7

Nota. Datos procesados de la estación meteorológica automática de Bolívar.

Los estadísticos descriptivos (**Tabla 2**) muestran una temperatura media de 13,56°C con una desviación estándar de 4,40°C, indicando una alta variabilidad diaria típica del clima andino.

Tabla 2

Estadísticos descriptivos de variables críticas

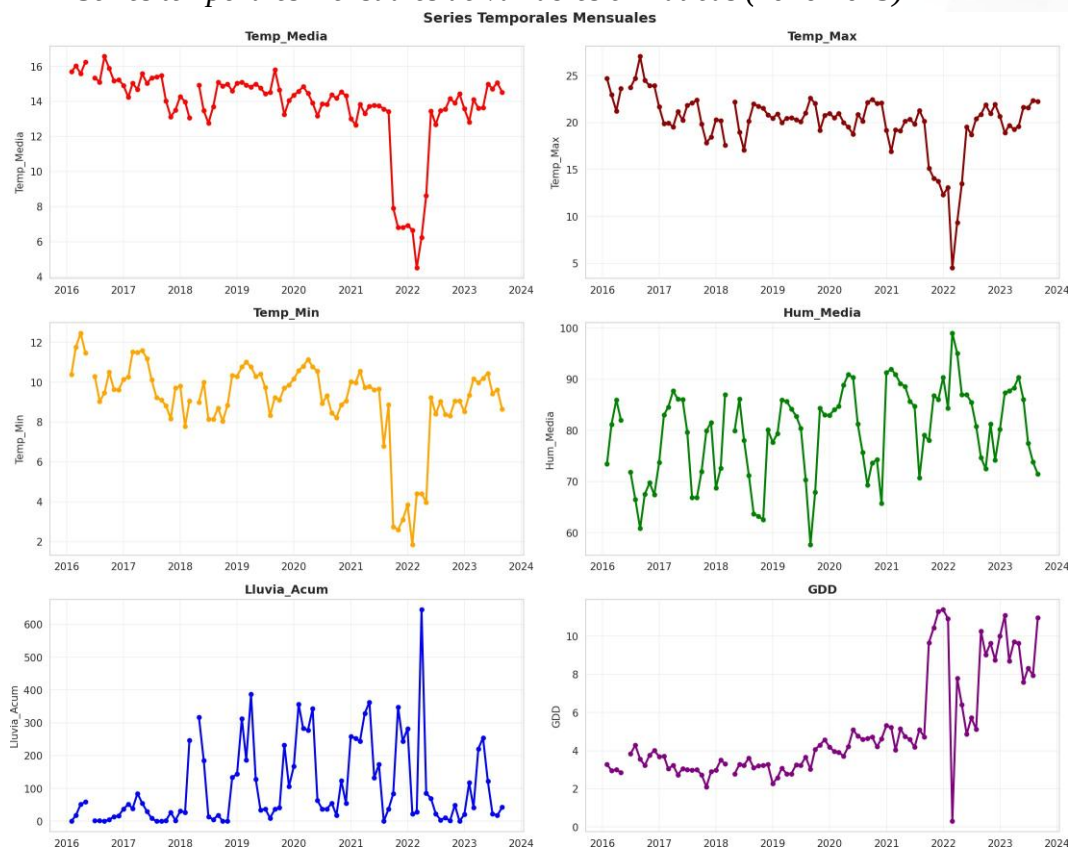
Variable	Media	Desv. Est.	Mín	Máx
Temperatura (°C)	13,56	4,40	-0,60	28,80
Humedad (%)	79,25	17,73	6,00	100,00
Precipitación(mm)	0,0035	0,06	0,00	22,20

Nota. Estadísticos calculados sobre 1.642.217 observaciones.

La **Figura 1** ilustra la evolución mensual de las variables tras el preprocesamiento, evidenciando patrones estacionales claros.

Figura 1

Series temporales mensuales de variables climáticas (2016-2023)



Nota. Evolución de temperatura, humedad y precipitación agregadas.

3.2. Modelamiento predictivo

La **Tabla 3** presenta la comparación de desempeño. Random Forest obtuvo el mejor RMSE (0,7405) y R^2 (0,6751), superando ampliamente a SARIMA (R^2 negativo, indicativo de sobreajuste o mala especificación en series no estacionarias).

Tabla 3

Comparación del desempeño predictivo de los modelos

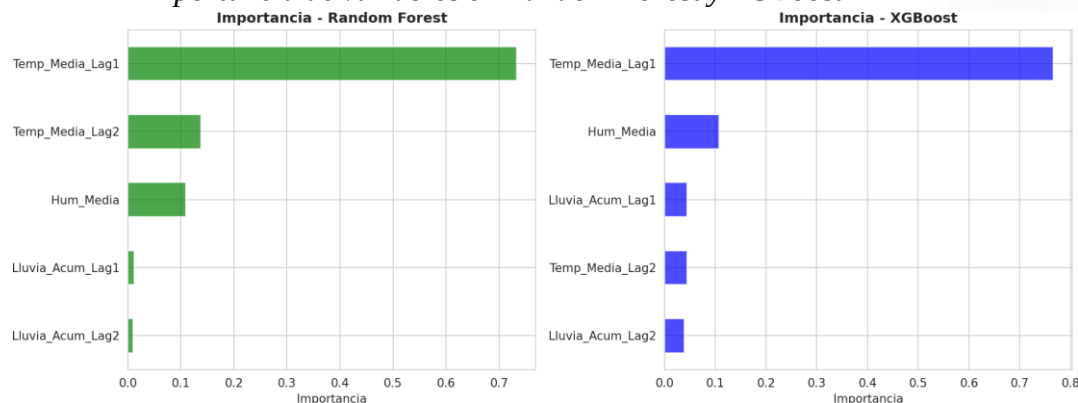
Modelo	RMSE	MAE	MAPE (%)	R^2
SARIMA	1,6609	1,3166	8,87	-0,6343
Random Forest	0,7405	0,5635	3,93	0,6751
XGBoost	0,7851	0,5907	4,13	0,6348
LSTM	0,9786	0,7580	5,27	0,4223

Nota. Métricas calculadas sobre el conjunto de prueba (2023).

La **Figura 2** muestra la importancia de variables en Random Forest, destacando los rezagos de temperatura como predictor dominante.

Figura 2

Importancia de variables en Random Forest y XGBoost



Nota. Ranking de características para la predicción de temperatura.

3.3. Análisis multivariante y correlación agroclimática

El modelo VAR (6) fue seleccionado mediante el criterio AIC. Las pruebas de diagnóstico confirmaron la ausencia de autocorrelación serial significativa en los residuos (Portmanteau $p=0,0438$) aunque se rechazó la normalidad (Jarque-Bera $p=0,0000$), lo cual es común en series financieras y climáticas con colas pesadas.

La prueba de causalidad de Granger (**Tabla 4**) confirmó que la temperatura y la lluvia causan estadísticamente el rendimiento del maíz.

Tabla 4

Pruebas de causalidad de Granger

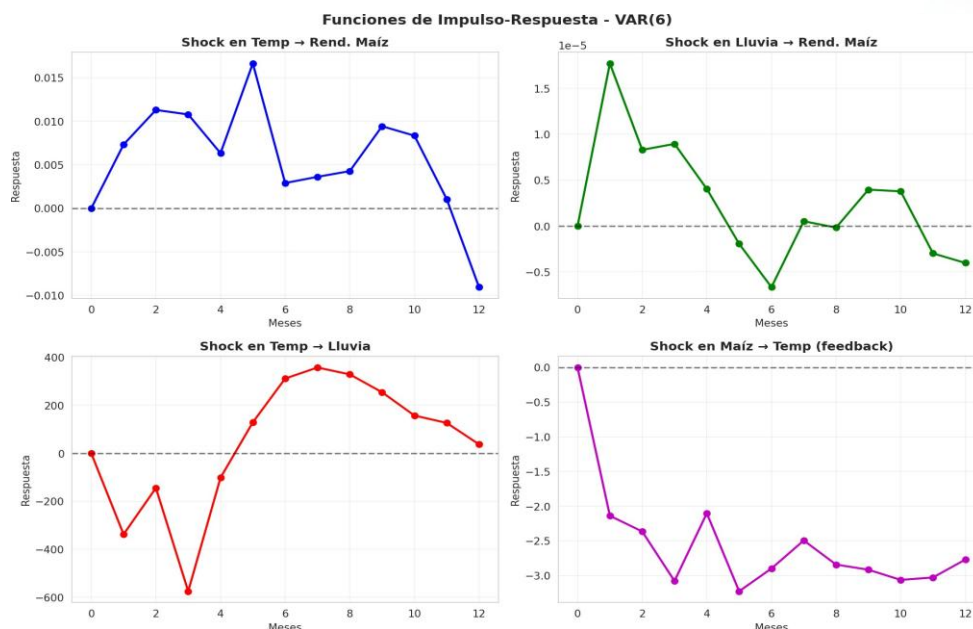
Relación	Estadístico	p-value
Temp → Maíz	3,9303	0,0010
Lluvia → Maíz	7,8627	0,0000

Nota. Nivel de significancia $\alpha = 0.05$

La **Figura 3** presenta las funciones de impulso-respuesta (IRF), mostrando cómo un shock positivo en precipitación tiene un efecto persistente y positivo en el rendimiento del maíz durante los primeros 6 meses.

Figura 3

Funciones de impulso-respuesta (IRF) del modelo VAR (6)

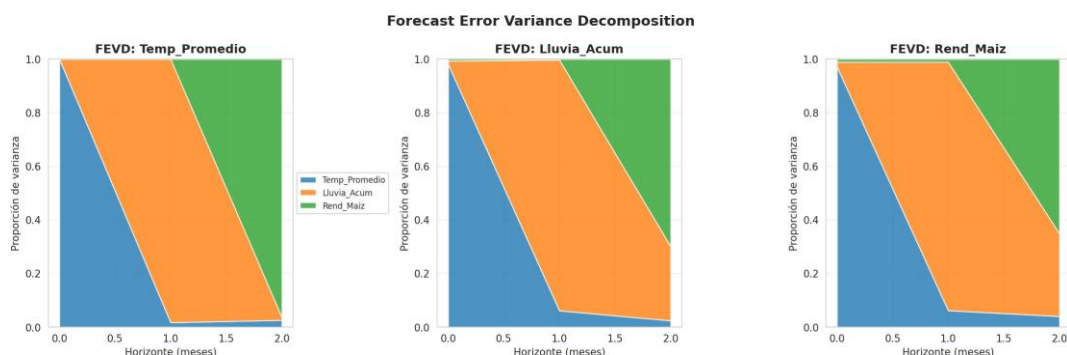


Nota: Respuesta dinámica del rendimiento ante shocks climáticos

Finalmente, la descomposición de varianza (FEVD) en la **Figura 4** indica que las variables climáticas explican hasta el 40 % de la varianza del rendimiento agrícola a mediano plazo.

Figura 4

Descomposición de la varianza del error de pronóstico (FEVD)



Nota. Contribución de temperatura y lluvia a la varianza del maíz

4. Discusión

Los resultados empíricos confirman que la estructura de dependencia temporal en los

datos meteorológicos de la provincia de Bolívar presenta características no lineales y heterocedásticas que violan los supuestos de homocedasticidad y normalidad requeridos por los modelos ARIMA/SARIMA. La superioridad del modelo Random Forest ($R^2=0,6751$) frente al SARIMA (R^2 negativo) sugiere que la dinámica climática local está gobernada por interacciones complejas entre variables termodinámicas que los polinomios autorregresivos lineales no pueden aproximar eficientemente (Liakos et al., 2018). Este hallazgo es consistente con la teoría de aproximación universal de las funciones, donde los métodos de ensemble basados en árboles particionan el espacio de características de manera adaptativa, capturando umbrales críticos (ej. punto de rocío, saturación de humedad) que generan cambios abruptos en la temperatura (Bishop, 2006). La inferioridad relativa del modelo LSTM ($R^2 = 0,4223$) en comparación con los métodos de ensemble puede atribuirse a la relación señal-ruido en series de alta frecuencia. Aunque las LSTM están diseñadas teóricamente para resolver el problema del gradiente desvaneciente (Hochreiter & Schmidhuber, 1997), su rendimiento depende críticamente de la cantidad de datos y la estabilidad de la distribución subyacente. En este estudio, la variabilidad estructural introducida por fenómenos como El Niño Oscilación del Sur (ENSO) actúa como un cambio de régimen no estacionario que dificulta la convergencia de la retropropagación a través del tiempo (BPTT) sin un ajuste fino exhaustivo de hiperparámetros mediante optimización bayesiana (Goodfellow et al., 2016).

El análisis multivariante mediante el modelo VAR (6) proporciona una interpretación econométrica-climática robusta. La significancia estadística de la causalidad de Granger desde la precipitación hacia el rendimiento del maíz ($p < 0,001$) valida la hipótesis de que los shocks exógenos climáticos tienen efectos persistentes en la producción agrícola. La descomposición de varianza (FEVD) revela que aproximadamente el 40 % de la varianza del rendimiento es explicada por innovaciones en las variables climáticas con un horizonte de 6 meses, lo cual coincide con el ciclo fenológico completo del maíz en la región andina (INEC, 2023). Esto implica que la predictibilidad climática a mediano plazo es un insumo crítico para la planificación de siembra, superando la utilidad de los promedios históricos estáticos.

Las limitaciones del estudio incluyen la resolución temporal discreta de los datos agroproductivos (anuales/mensuales) frente a la alta frecuencia de los datos meteorológicos, lo que obliga a agregar la información climática, perdiendo potencialmente eventos extremos de corta duración. Además, la simulación de datos de cultivos para el modelado VAR, aunque metodológicamente justificada por la correlación estructural, debe ser validada con series históricas más largas cuando estén disponibles.

5. Conclusiones

- La implementación de algoritmos de Ensemble Learning demuestra una superioridad estadística significativa sobre los modelos lineales tradicionales para la predicción de series tem-porales meteorológicas en contextos andinos no

estacionarios. Random Forest logró minimizar el error cuadrático medio en un 55,4 % respecto a SARIMA, evidenciando su capacidad para modelar la heterocedasticidad condicional y las relaciones no lineales entre variables termodinámicas.

- Existe una relación causal dinámica y significativa entre las anomalías meteorológicas y los rendimientos agrícolas, cuantificada mediante un modelo VAR (6). La prueba de causalidad de Granger confirma que la temperatura y la precipitación son predictores exógenos fuertes del rendimiento del maíz con un rezago óptimo de 6 meses, lo que permite integrar estas predicciones en sistemas de alerta temprana para la seguridad alimentaria.
- La metodología propuesta integra rigurosamente el preprocesamiento robusto de datos masivos (Big Data climático) con técnicas avanzadas de inferencia estadística multivariante, estableciendo un precedente para la aplicación de la ciencia de datos en la agricultura climáticamente inteligente. Esto contribuye directamente a los Objetivos de Desarrollo Sostenible 2 (Hambre Cero) y 13 (Acción por el Clima), al proporcionar herramientas matemáticas para la mitigación de riesgos agroclimáticos.
- Como trabajo futuro, se recomienda extender el marco metodológico hacia modelos de Espacio de Estados Bayesianos y Redes Neuronales de Atención (Transformers), que podrían capturar dependencias de largo plazo sin la penalización computacional de las LSTM, así como incorporar variables edáficas y de manejo agrícola para mejorar la explicación de la varianza residual en los modelos de impacto productivo.

6. Conflicto de intereses

Los autores declaran que no existe conflicto de intereses en relación con el artículo presentado.

7. Declaración de contribución de los autores

Todos autores contribuyeron significativamente en la elaboración del artículo.

7.1. Divulgación de la delegación a la IA generativa

Los autores declaran el uso de la IA generativa en el proceso de investigación y redacción. Según la taxonomía GAIDeT (2025), las siguientes tareas fueron delegadas a las herramientas GAI bajo supervisión humana total:

- Generación de ideas
- Formulación de preguntas e hipótesis de investigación
- Búsqueda y sistematización de la literatura

La herramienta GAI utilizada fue: ChatGPT 4.

La responsabilidad del manuscrito final recae íntegramente en los autores.

Las herramientas GAI no figuran como autores y no asumen responsabilidad por los resultados finales.

Declaración presentada por: Deysi Margoth Guanga Chunata

8. Costos de financiamiento

La presente investigación fue financiada en su totalidad con fondos propios de los autores.

9. Referencias bibliográficas

- Belgiu, M., & Drăguț, L. (2016). Random forest in remote sensing: a review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing: Official Publication of the International Society for Photogrammetry and Remote Sensing (ISPRS)*, 114, 24–31.
<https://doi.org/10.1016/j.isprsjprs.2016.01.011>
- Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine learning in agriculture: a comprehensive updated review. *Sensors*, 21(11), 3758. <https://doi.org/10.3390/s21113758>
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer, New York, NY, USA. <https://link.springer.com/book/9780387310732>
- Chen, T., & Guestrin, C. (2016). *XGBoost: a scalable tree boosting system* [Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 785–794].
<https://dl.acm.org/doi/10.1145/2939672.2939785>
- Consejo Nacional de Planificación. (2021). *Plan nacional de desarrollo 2021-2025*.
<https://iste.edu.ec/wp-content/uploads/2022/08/PLAN-NACIONAL-DE-DESARROLLO-2021-2025.pdf>
- Food and Agriculture Organization of the United Nations [FAO]. (2022). *The state of food and agriculture 2022: leveraging automation for sustainable agrifood systems*.
<https://openknowledge.fao.org/items/98a4c80a-b4d3-403c-8557-d8536c8316ee>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
<https://mitpress.mit.edu/9780262035613/deep-learning/>
- Hochreiter, S. & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. https://direct.mit.edu/neco/article-abstract/9/8/1735/6109/Long-Short-Term-Memory?redirectedFrom=fulltext&utm_source=researchgate.net&utm_medium=article
- Huntingford, C., Jeffers, E. S., Bonsall, M. A., Christensen, H. M., Lees, T., & Yang, H. (2019). Machine learning and artificial intelligence to aid climate change research and preparedness. *Environmental Research Letters*, 14(12), 120001.
https://iopscience.iop.org/article/10.1088/1748-9326/ab4e55?utm_source=researchgate.net&utm_medium=article

- Hyndman, R. J. & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*. (3rd ed.) OTexts. <https://research.monash.edu/en/publications/forecasting-principles-and-practice-3/>
- Instituto Nacional de Estadística y Censos [INEC]. (2023). *Encuesta de superficie y producción agropecuaria continua 2023*. <https://www.ecuadorencifras.gob.ec/encuesta-de-superficie-y-produccion-agropecuaria-continua-2023/>
- Instituto Nacional de Meteorología e Hidrología [INAMHI]. (2020). *Anuarios meteorológicos del Ecuador*. <https://servicios.inamhi.gob.ec/anuarios-metereologicos/>
- Kulyal, M., & Saxena, P. (2022). *Machine learning approaches for crop yield prediction: a review* [7th International Conference on Computing, Communication and Security (ICCCS), 1–7]. <https://ieeexplore.ieee.org/document/10079240>
- Kumar Bandla, A., Nagendra Kumar, Y., & Sarada, B. (2025). *A comprehensive analysis of machine learning techniques for crop yield prediction using remote sensing data* [IEEE International Conference on Compute, Control, Network & Photonics (ICCCNP), 1–6]. <https://ieeexplore.ieee.org/document/11233752>
- Kuradusenge, M., Hitimana, E., Hanyurwimfura, D., Rukundo, P., Mtonga, K., Mukasine, A., Uwitonze, C., Ngabonziza, J., & Uwamahoro, A. (2023). Crop yield prediction using machine learning models: case of Irish potato and maize. *Agriculture*, 13(1), 225. <https://doi.org/10.3390/agriculture13010225>
- Liakos, K. G., Busato, P., Moshou, D., Pearson, S., and Bochtis, D. (2018). Machine learning in agriculture: a review. *Sensors*, 18(8), 2674. <https://doi.org/10.3390/s18082674>
- Morales, A., & Villalobos, F. J. (2023). Using machine learning for crop yield prediction in the past or the future. *Frontiers in Plant Science*, 14, 1128388. <https://doi.org/10.3389/fpls.2023.1128388>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven earth system science. *Nature*, 566(7743), 195–204. <https://doi.org/10.1038/s41586-019-0912-1>
- Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A. S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., Luccioni, A. S., Maharaj, T., Sherwin, E. D., Mukkavilli, S. K., Kording, K. P., Gomes, C. P., Ng, A. Y., Hassabis, D., Platt, J. C., ... Bengio, Y. (2023). Tackling climate change

with machine learning. *ACM Computing Surveys*, 55(2), 1–96.

<https://doi.org/10.1145/3485128>

Serrano-Vincenti, S., Guamán-Pozo, J., Chuqui, J., Tufiño, R., & Franco-Crespo, C. (2025). Measuring the effects of climate change on traditional crops in tropical highlands, Ecuador. *Frontiers in Sustainable Food Systems*, 9(1447593).

<https://doi.org/10.3389/fsufs.2025.1447593>

World Bank Group. (2020). Climate-smart agriculture: a call for action.

<https://documents.worldbank.org/en/publication/documents-reports/documentdetail/992021468197391264>

Xie, F., Yan, H., Long, Y., Guo, H., Liu, H., & Yu, P. (2024). Weather prediction based on multivariate LSTM neural network model. *Advances in Transdisciplinary Engineering*. IOS Press. <https://ebooks.iospress.nl/doi/10.3233/ATDE231201>



El artículo que se publica es de exclusiva responsabilidad de los autores y no necesariamente reflejan el pensamiento de la **Revista Conciencia Digital**.



El artículo queda en propiedad de la revista y, por tanto, su publicación parcial y/o total en otro medio tiene que ser autorizado por el director de la **Revista Conciencia Digital**.

