






Comparación de Modelos Estadísticos para la Predicción del Nicho Ecológico de *Fusarium oxysporum* f. sp. *cupense* y su Implicación en la Fitosanidad de Suelos Agrícolas

Comparison of Statistical Models for the Prediction of the Ecological Niche of *Fusarium oxysporum* f. sp. *cupense* and its Implication in the Phytosanitation of Agricultural soils

Angelita Leonor Bosquez-Mestanza¹ , Yarelys Ferrer-Sánchez^{2†} ,
Alexis Herminio Plasencia-Vázquez³ , Jorge Josué Jacho-Saa²  y
Anay Serrano-Rodríguez³ 

¹ Universidad Técnica Estatal de Quevedo, Facultad de Ciencias de la Computación y Diseño Digital; (A.L.B.M.). ² Facultad de Ciencias de la Ingeniería; (Y.F.S.), (J.J.J.S.). Av. Quito km 1.5, Vía a Santo Domingo de los Tsáchilas. 120501, Quevedo, Los Ríos, Ecuador.

[†] Autora para correspondencia (yferrer@uteq.edu.ec)

³ Universidad Autónoma de Campeche, Centro de Investigaciones Históricas y Sociales. Avenida Agustín Melgar s/n, Col. Buenavista. 24039 Campeche, Campeche, México; (A.H.P.V.), (A.S.R.).

RESUMEN

Es importante evaluar la eficacia de diferentes modelos de distribución de especies, ya que son instrumentos valiosos para predecir distribuciones futuras. Se evaluó la eficacia del modelo bayesiano, *Bayesian Additive Regression Trees* (BART), frente a modelos estadísticos: *Generalized Linear Model* (GLM), *Generalized Additive Model* (GAM) y algoritmos de aprendizaje automático: *Random Forest* (RF), *Maximum Entropy Model* (MaxEnt) y *Boosted Regression Tree* (GBM) para modelar la distribución potencial del patógeno *Fusarium oxysporum* f. sp. *cupense* en Ecuador continental, fitopatógeno de suelo que afecta significativamente la producción de banano. Se emplearon 92 registros de presencia de la especie y cuatro variables bioclimáticas (de WorldClim). Se obtuvieron los modelos a una resolución espacial de 10 × 10 km², y fueron evaluados con diferentes métricas de rendimiento. Los mapas geográficos de probabilidad de presencia generados en R para cada modelo se exportaron a ArcGis, para elaborar mapas binarios y calcular el área de presencia y ausencia de la especie. La eficacia de los modelos varía significativamente según la métrica utilizada. A pesar de ello, el modelo Random Forest tuvo buen rendimiento en la mayoría de las métricas empleadas, principalmente en discriminación; sin embargo, el Modelo Lineal Generalizado presentó mejor precisión y confiabilidad, con una calibración excelente. Con MaxEnt se obtuvo la mayor área de presencia del fitopatógeno (106 800 km²). Todos los modelos mostraron consistentemente una favorabilidad ambiental > 0.90 en la región Sierra, asociado a las características geográficas, climáticas y ecológicas de la región, tales como la combinación de temperatura y humedad y la disponibilidad de hospederos. Los modelos demostraron su eficacia al predecir la distribución potencial de *F. oxysporum*; sin embargo, se sugiere continuar la investigación con un mayor número de registros de presencia. El estudio es relevante como línea base, especialmente al explorar modelos bayesianos (BART).

Palabras clave: aprendizaje automático, árboles de regresión aditivos bayesianos, enfermedad de Panamá, MaxEnt.

SUMMARY

It is important to evaluate the effectiveness of different species distribution models, as they are valuable tools for predicting future distributions. The efficiency of the Bayesian model Bayesian Additive Regression Trees (BART) was evaluated against statistical models, including the Generalized Linear Model (GLM) and Generalized



Cita recomendada:

Bosquez-Mestanza, A. L., Ferrer-Sánchez, Y., Plasencia-Vázquez, A. H., Jacho-Saa, J. J., & Serrano-Rodríguez, A. (2026). Comparación de Modelos Estadísticos para la Predicción del Nicho Ecológico de *Fusarium oxysporum* f. sp. *cupense* y su Implicación en la Fitosanidad de Suelos Agrícolas. *Terra Latinoamericana*, 44, 1-15. e2417. <https://doi.org/10.28940/terralatinoamericana.v44i.2417>

Recibido: 16 de octubre de 2025.
Aceptado: 5 de diciembre de 2025.
Artículo. Volumen 44.
Abril de 2026.

Editor de Sección:
Dr. Fernando Abasolo Pacheco



Copyright: © 2026 by the authors.
Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY NC ND) License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

Additive Model (GAM), and machine learning algorithms, including Random Forest (RF), Maximum Entropy (MaxEnt), and Boosted Regression Tree (GBM), to model the potential distribution of the pathogen *Fusarium oxysporum* f. sp. *cubense* in mainland Ecuador, a soil phytopathogen that significantly affects banana production. Ninety-two presence records and four bioclimatic variables (WorldClim) were used. The models were generated at a spatial resolution of 10×10 km² and evaluated using different performance metrics. Probability-of-presence maps generated in R for each model were exported to ArcGIS to produce binary maps and calculate the area of presence and absence of the species. Model performance varied depending on the evaluation metric. Despite this, the Random Forest model performed well across most metrics, particularly in discrimination; however, the Generalized Linear Model showed higher accuracy and reliability, with excellent calibration. MaxEnt produced the largest predicted area of presence of the phytopathogen (106 800 km²). All models consistently showed environmental favorability > 0.90 in the Sierra region, associated with its geographical, climatic, and ecological characteristics, such as the combination of temperature and humidity and host availability. The models demonstrated efficiency in predicting the potential distribution of *F. oxysporum*; however, further research with a greater number of presence records is recommended. The study is relevant as a baseline, particularly for the application of Bayesian models such as BART.

Index words: machine learning, Bayesian additive regression trees, Panama disease, MaxEnt.

INTRODUCCIÓN

En las últimas décadas, se ha implementado un gran número de métodos estadísticos y de aprendizaje rápido en la modelación de la distribución de especies (MDE). De estos, solo un pequeño grupo ha logrado prevalecer por su fácil aplicación, velocidad computacional y su fuerte rendimiento predictivo en rigurosas predicciones cruzadas, como MaxEnt, uno de los más populares gracias a los excelentes resultados proporcionados (Carlson, 2020). Entre los algoritmos de aprendizaje automático más potentes utilizados para obtener MDE destacan los Modelos Aditivos Generalizados (GAM) y *Random Forest* (RF). Sin embargo, la efectividad de estos algoritmos para proyectar diferentes niveles dentro de los sistemas ecológicos, como la densidad y riqueza de especies, no ha sido completamente confirmada (Kosicki, 2020). Los árboles de regresión aditiva bayesiana (BART) han aparecido como una nueva alternativa a otros métodos populares de árboles de clasificación (Konowalik y Nosol 2021).

En la actualidad, ante esta gran cantidad de herramientas estadísticas disponibles para predecir la distribución de las especies, se ha hecho necesario evaluar su pertinencia bajo el enfoque de diferentes investigaciones. De esta manera se ha determinado que, cuando no se incluyen métodos analíticos variados en el proceso de predicción de MDE, el valor ecológico de los resultados puede ser limitado (Konowalik y Nosol 2021). Además, es difícil decidir cuál MDE es el óptimo para cada una de las situaciones que se pueden presentar. En ausencia de métricas de rendimiento del modelo específicas de la aplicación, a menudo se busca que el algoritmo genere predicciones más precisas (Hao, Elith, Lahoz y Guillera, 2020). Sin embargo, el rendimiento predictivo no depende exclusivamente del modelo, sino que en él influyen, entre otros factores, la calidad de los registros de las especies en estudio, el tipo de especie, la ecología de estas, así como las variables ambientales que se utilicen, por lo que no se ha podido establecer que un algoritmo sea superior a otro (Pearson *et al.* 2006). Ante esta situación, se sugiere como una alternativa construir diferentes modelos utilizando varios algoritmos, los cuales se comparan en función de cuál logra un mejor rendimiento y predicción para la especie en cuestión.

Los MDE se han utilizado ampliamente para predecir el impacto de especies invasoras y patógenos, ante la grave amenaza que representan para los ecosistemas y para muchas actividades agrícolas productivas. Un enfoque importante y preciso para predecir áreas potencialmente favorables para los patógenos, consiste en trabajar con conjuntos de MDE (Araújo y New 2007), utilizando como variables predictoras métricas de eventos climáticos. Dentro de este grupo, el género *Fusarium* es uno de los más importantes e incluye varias especies de fitopatógenos, que están ampliamente distribuidas en la mayoría de las regiones climáticas del mundo (Burgess, 1981). Estos patógenos representan un problema importante en sistemas agrícolas, ya que su control requiere la implementación de estrategias complementarias, incluyendo el uso de agentes de control biológico (Cabrera, Garmendía, Rufo, Pereyra y Vero, 2020). Una de las especies más conocidas es *Fusarium oxysporum*, la cual tiene cepas patógenas que probablemente causan mayor daño económico a los cultivos agrícolas que cualquier otro patógeno vegetal (Correll, 1991). Existen varias razas que son los agentes de brote de varios cultivos de alto impacto económico como *F. oxysporum* f. sp. *cubense* raza 1 en banano que afectó la variedad Gros Michel, o la raza 4 que destruyó la industria bananera Cavendish de Taiwán en la década de 1970 (Smith, 2007).

Ecuador es un país con una gran biodiversidad, pero con una elevada vulnerabilidad a la invasión de patógenos como *F. oxysporum* f. sp. *ubense*. La aparición de la Raza Tropical 4 (RT4) de este patógeno en países vecinos como Perú y Colombia, ha generado preocupación en la región, ya que este hongo es altamente peligroso para el cultivo de banano Cavendish. Ecuador registró en diciembre de 2025 oficialmente el primer caso de infección por la RT4, además de la existencia de otras cepas del patógeno presentes desde 1930. La incertidumbre sobre la expansión del primer brote de esta raza más peligrosa de *F. oxysporum* f. sp. *ubense* en el país representa una amenaza para el cultivo de banano y la economía de un gran número de personas. Por esta razón, se considera importante aplicar modelos estadísticos confiables y robustos para estimar la distribución de este patógeno en territorio ecuatoriano. En este estudio se comparó la eficacia de diferentes métodos predictivos aplicados a la MDE para predecir la distribución geográfica potencial de *F. oxysporum* f. sp. *ubense* en Ecuador continental, los cuales han mostrado resultados sobresalientes en predicciones similares. Este trabajo es clave para poder tomar medidas eficaces para prevenir su invasión hacia nuevas zonas adecuadas climáticamente y poder controlar su propagación en los sitios donde se ha detectado su presencia.

MATERIALES Y MÉTODOS

Área de Estudio

La República del Ecuador está ubicada en la costa noroeste de América del Sur, entre los 1° 30' N y 3° 23.5' S, y los 75° 12' y 81° 00' O. Limita al norte con Colombia, al sur y este con Perú, y al oeste con el océano Pacífico. Su superficie territorial cubre dos de los cinco puntos críticos de biodiversidad en América del Sur: los Andes tropicales y el Corredor Tumbes-Chocó-Magdalena, brindando una gran variedad de ecosistemas tales como: tundra alpina (Páramo), manglares, bosque nuboso y selva tropical; incluyendo tres biomas: la Costa, los Andes y la Amazonía (Kleemann *et al.*, 2022).

Datos de Presencia de la Especie y Variables Bioclimáticas

Se accedió a los datos de presencia de la especie *F. oxysporum* f. sp. *ubense* a través de la base de datos Global Biodiversity Information Facility (GBIF, 2024). Los registros se complementaron con información de casos de estudio de la especie en Ecuador, proveniente de diferentes instituciones gubernamentales. Las variables climáticas se obtuvieron de la base de datos WorldClim, la cual reúne datos meteorológicos y climáticos globales de alta resolución espacial. En su versión más actual, consta de un conjunto de 19 variables basadas en datos climáticos mensuales interpolados espacialmente con una resolución de 2.5 arc-segundos (Fick y Hijmans 2017).

Dado el amplio rango de distribución de la especie, se establecieron límites del área para obtener los registros de presencia, los cuales corresponden al área que incluye Colombia, Ecuador y Perú (entre los 81° y 68° O, y desde 19° S hasta 12° N). De estos registros, se excluyeron aquellos reportados con estado de ocurrencia como "ausencia" o cuya incertidumbre de coordenadas fue mayor que la resolución espacial de la cuadrícula de análisis (10 km × 10 km). Además, se usó el paquete "scrubr" (Chamberlain, 2020) en R (R Core Team, 2024), para limpiar coordenadas incompletas, imprecisas, imposibles o improbables. La base de datos final incluía 92 registros de presencia de *F. oxysporum* f. sp. *ubense*.

Muchos registros de presencia tenían mucha proximidad; sin embargo, para lograr una estandarización se usó el mismo tamaño de píxel 10 km × 10 km para todo el análisis. De esta forma, cada píxel ocupado contó como una sola presencia, independientemente de la cantidad de registros que tuviese. De esta forma se reduce la influencia del sesgo de muestreo ya que algunas localidades, sobre todo las accesibles, tienen esfuerzos de estudio excesivamente grandes (Staude, Navarro y Pereira, 2020). Se consideró que la autocorrelación espacial que pudiese quedar en los datos, después de este proceso, es un resultado natural de la dispersión de la especie y no debe eliminarse artificialmente (Legendre y Legendre 2012).

Con relación a las variables predictoras, inicialmente se incluyeron las 19 variables bioclimáticas en la modelación. Para identificar las variables que fueran más relevantes para la modelación de la distribución de la especie, se pueden seleccionar dada su relevancia ecológica, y estadísticamente utilizando diferentes métodos. Por ejemplo, se debe evitar el uso de variables muy correlacionadas, a través de un análisis correlativo de pares de variables y eliminando algunas de ellas con base en el conocimiento ecológico. Cuando hay ausencia de un buen conocimiento ecológico, se puede utilizar, por ejemplo, la función "multGLM" para seleccionar variables que no estén correlacionadas por encima de $|r| = 0,8$ y que sean suficientemente informativas según el BIC o Criterio de Información

Bayesiano en un Modelo Lineal Generalizado (GLM). En este caso, el último procedimiento (GLM) fue utilizado para modelar las relaciones lineales entre variables, con el cual se identificaron como más relevantes la isothermalidad (bio 3), la temperatura máxima del mes más cálido (bio 5), el rango de temperatura anual (bio 7) y la precipitación anual (bio 12). Estas variables finalmente fueron seleccionadas para los siguientes pasos de la modelación.

Modelación del Nicho Ecológico

Se siguieron las mejores prácticas para los modelos de distribución de especies (MDE) en cuanto al manejo de las variables de respuesta y predictoras, la construcción de modelos y la evaluación mediante métricas que valoran diferentes facetas del rendimiento del modelo (Sillero y Barbosa, 2021). Se utilizó un enfoque de conjunto con seis métodos de modelación de complejidad variable, incluidos algoritmos de aprendizaje automático basados en regresión y en árboles: modelos lineales generalizados (*Generalized Linear Models*, GLM), modelos aditivos generalizados (GAM), bosques aleatorios (*random forest*, RF), árboles de regresión potenciados (*Boosted Regression Trees*, GBM), árboles de regresión aditivos bayesianos (*Bayesian Additive Regression Trees*, BART) y Máxima entropía (MaxEnt). Los cinco primeros métodos producen probabilidad de presencia, que es importante para una combinación evaluable de predicciones de diferentes métodos. Con la resta de la prevalencia, se pudieron comparar los métodos de probabilidad con el de MaxEnt que indica directamente la favorabilidad ambiental del píxel para la presencia de la especie.

Los algoritmos GLM, GAM, MaxEnt y RF se han utilizado ampliamente en los MDE (Peterson *et al.* 2011), mientras que BART es un método novedoso que produce resultados robustos (Carlson, 2020; Baquero *et al.*, 2021). Todos los modelos se ajustaron en el programa RStudio v2023.06.1 (R Core Team, 2024), y se utilizaron los paquetes "base", "gam" (Hastie, 2015), "randomForest" (Liaw y Wiener 2002), "maxnet" (Phillips, Anderson, Dudík, Schapire y Blair, 2017), "gbm" (Elith y Leathwick, 2017) y "embarcadero" (Carlson, 2020) para la calibración de estos.

Los datos se dividieron en diez bloques de validación cruzada: las celdas de la cuadrícula del área de estudio se dividieron en diez grupos, utilizando el paquete "blockcv" versión 2.1.1 (Valavi, Elith, Lahoz y Guillera, 2018) de R, con una selección sistemática y asignación de bloques espaciales en diez bloques de 10 km × 10 km. La validación cruzada de bloques es más confiable que la validación cruzada aleatoria, la cual subestima el error de predicción y da como resultado una selección errónea del modelo (Valavi *et al.* 2018). La separación espacial de los sets de datos de calibración y validación permitió evaluar si los modelos funcionaban bien tanto en áreas cercanas como en áreas más distantes, lo cual tiene relevancia si los modelos se extrapolan en el espacio o en el tiempo (Valavi *et al.* 2018). Se construyeron diez modelos diferentes para cada algoritmo de modelación, en los que un set se dejó fuera del modelo de forma secuencial y se usó para la validación de las predicciones del modelo (Fielding y Bell 1997).

El rendimiento de la validación cruzada de cada algoritmo se evaluó con el paquete "modeva" versión 2.0 (Barbosa *et al.*, 2013). Para evaluar el desempeño de los modelos, se emplearon tres métricas principales: (a) el AUC, que mide el rendimiento de discriminación general, la capacidad del modelo para distinguir las presencias de las ausencias; (b) el *True Skill Statistic* (TSS), que mide el rendimiento de la clasificación, la proporción de presencias y ausencias clasificadas correctamente (Allouche, Tsoar y Kadmon, 2006), utilizando la prevalencia de la calibración como valor umbral para clasificar las presencias y ausencias previstas; y (c) la pendiente de calibración de Miller, que evalúa la confiabilidad del modelo, la desviación general de las probabilidades predichas de las frecuencias de ocurrencia observadas (Miller, Hui y Tierney, 1991; Pearce y Ferrier 2000). Los valores de estas métricas se promediaron en los diez sets de validación cruzada para cada algoritmo. Para el AUC, que varía de 0 a 1, se fijó un umbral de rendimiento en 0,7, lo que delimita predicciones deficientes (Rapacciuolo *et al.* 2012). Para el TSS que varía entre -1 a 1, el umbral de rendimiento proporcional se fijó en 0.4. Para la pendiente de calibración de Miller, se usó un umbral de rendimiento de 0.5 por encima o por debajo del valor ideal de 1, como recomiendan Baquero *et al.* (2021).

Posteriormente, se calcularon métricas de clasificación basadas en umbrales, tales como CCR (*Correct Classification Rate*), *Sensitivity*, *Specificity*, *Precision*, *Recall* y *Kappa*, para evaluar la precisión de los modelos en la clasificación de las áreas como favorables o no favorables para la presencia de la especie. Con estas métricas, se pudo determinar el desempeño óptimo de cada modelo y comparar los resultados entre ellos. La precisión, especificidad, sensibilidad y *Recall* son métricas complementarias que permitieron evaluar la exactitud de los resultados del modelo. A través de la precisión se midió la proporción de observaciones verdaderas positivas en relación con todas las observaciones, mientras que la especificidad evaluó la proporción de observaciones verdaderas negativas. La sensibilidad evaluó la capacidad del modelo de detectar correctamente la presencia de

una especie, mientras que el *Recall* evaluó la proporción de presencias verdaderas que el modelo ha identificado correctamente. Con el coeficiente *Kappa* se midió la consistencia entre los resultados del modelo y los datos de verificación, ya que permite comparar la precisión de diferentes modelos.

Después de la validación cruzada, los modelos seleccionados se ajustaron nuevamente con el conjunto de datos completo, para evitar la sensibilidad a la eliminación de puntos (Araújo et al., 2019). Los valores de probabilidad predichos se convirtieron en favorabilidad, que es una función de la probabilidad de presencia de las especies y de la relación presencia-ausencia en la muestra modelada (Real, Barbosa y Vargas, 2006). Así se elimina el efecto de la prevalencia de las especies sobre los valores predichos, lo que hace que las predicciones sean comparables entre modelos (Real, Barbosa y Bull, 2017). Para resumir las predicciones de favorabilidad (10/algoritmo) de los diferentes modelos se obtuvo un promedio por cada algoritmo. De esta manera, las predicciones finales pudieron compararse entre algoritmos de modelación.

RESULTADOS Y DISCUSIÓN

En Ecuador, se han realizado varios estudios referentes a la modelación de distribución de especies invasoras (Ron, 2005; Ferrer-Sánchez et al., 2021). Sin embargo, no se encuentran registros de estudios acerca del uso de modelos estadísticos bayesianos centrados en Ecuador continental. Debido a esta situación, resulta relevante el desarrollo de investigaciones como la que se presenta, que sirvan como línea base para futuros proyectos respecto a patógenos como *F. oxysporum* f. sp. *cubense* en Ecuador, donde se compruebe la eficacia de los modelos bayesianos. A su vez, estos proyectos en el futuro deben complementarse con metodologías aplicadas en países vecinos como Colombia y Perú, donde se han realizado investigaciones relacionadas con este patógeno (Acurio, Rafael y Dangles, 2010).

Fusarium oxysporum f. sp. *cubense* tiene una alta capacidad de adaptación, lo que le permite distribuirse en distintas condiciones bioclimáticas en el espacio geográfico (Cuadro 1). Hubo registros de la especie en sitios con temperaturas máximas tan bajas como 10 °C y extremas de 37 °C, promediando los 26 °C (Cuadro 1). De igual forma, se evidenció esta capacidad adaptativa con 16 °C de rango anual de la temperatura, en sitios con bajas precipitaciones anuales (139 mm) y en otros con precipitaciones extremas (3782 mm, Cuadro 1).

Todos los modelos coinciden en una alta favorabilidad para *F. oxysporum* en la Sierra. La distribución basada en el modelo GLM, es heterogénea (Figura 1). El mapa de favorabilidad (Figura 1) indica una presencia superior al 0.90 en la región Sierra y una disminución progresiva en la región Costa y en el Oriente. Por otro lado, el mapa de probabilidad de presencia (Figura 1), refleja una baja presencia de la especie en prácticamente todo el territorio ecuatoriano, con excepción de la Sierra, donde se encuentra una concentración mayor en las provincias de Cotopaxi, Tungurahua y Chimborazo, con un nivel de alrededor del 0.050. Para el modelo GAM, el mapa de favorabilidad muestra que la presencia de la especie es alta en la región Sierra, con un nivel superior a 0.90. En la región Costa, aunque hay una disminución, la especie se puede encontrar en las provincias Manabí, Guayas y parte de El Oro, con valores alrededor de 0.70 (Figura 1).

El modelo obtenido a través de MaxEnt también mostró una distribución heterogénea de *F. oxysporum* f. sp. *cubense*. El mapa de favorabilidad identificó una presencia significativa de más de 0.90 en la región Sierra. En la región Costa solo se localizó en las provincias Manabí, Guayas, Santa Elena y El Oro, con un valor cercano a 0.50. Por su parte, el mapa de probabilidad de presencia (Figura 1) es similar al de favorabilidad. Los resultados del modelo RF muestran la menor representación de *F. oxysporum* f. sp. *cubense* en territorio (Figura 2). En el mapa de probabilidad de presencia (Figura 2), solo aparece una cantidad limitada de puntos en las provincias Pichincha y Cotopaxi.

Cuadro 1. Estadísticos descriptivos de las variables bioclimáticas que se utilizaron para modelar el nicho climático de *Fusarium oxysporum* f. sp. *cubense*.

Table 1. Descriptive statistics of the bioclimatic variables that were used to model the climatic niche of *Fusarium oxysporum* f. sp. *cubense*.

VARIABLES	Media ± DE	Min	Max	Error estándar	Coef. Var.
Isotermalidad	0.74 ± 0.11	0.46	0.93	0.01	14.21
Temperatura máxima del mes más cálido (°C)	25.89 ± 6.06	10.40	37.00	0.45	23.41
Rango anual de temperatura (°C)	16.25 ± 4.92	9.20	28.80	0.36	30.27
Precipitación anual (mm)	1131.90 ± 528.60	139.0	3782.0	39.08	46.70

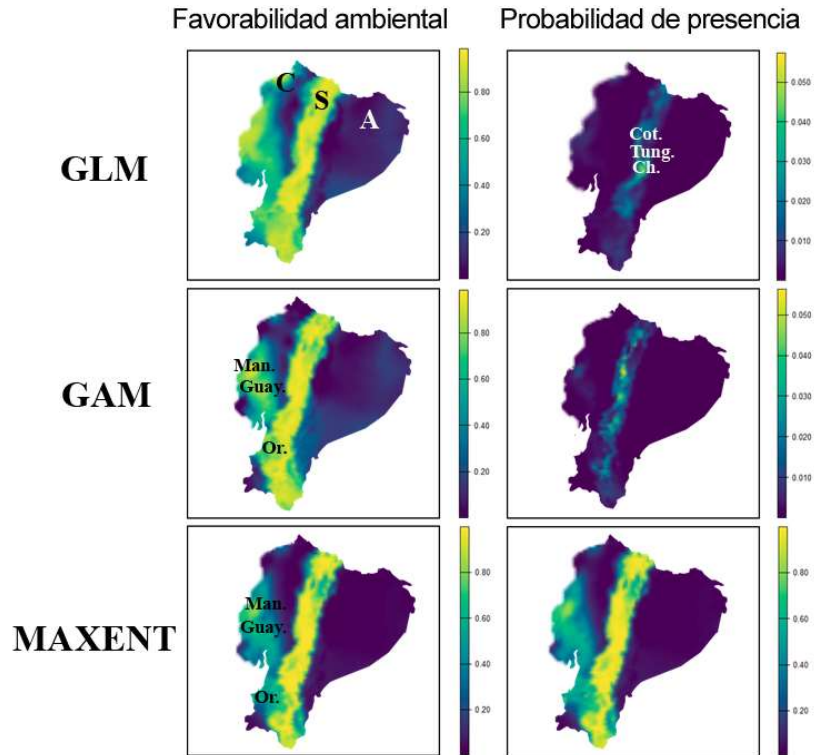


Figura 1. Mapas de favorabilidad y probabilidad de presencia de *Fusarium oxysporum* f. sp. *cubense* en Ecuador continental; GLM = modelo lineal generalizado; GAM = modelo aditivo generalizado; MAXENT = máxima entropía; C = costa; S = sierra; A = amazonía; Man. = Manabí; Guay. = Guayas; Or. = El Oro.
Figure 1. Maps of favorability and probability of presence of *Fusarium oxysporum* f. sp. *cubense* in continental Ecuador; GLM = generalized linear model; GAM = generalized additive model; MAXENT = maximum entropy; C = coast; S = highlands; A = amazon; Man. = Manabí; Guay. = Guayas; Or. = El Oro.

Los mapas basados en el modelo GBM (Figura 2), también muestran una distribución heterogénea. Si embargo, hacia la región del Oriente se observa una favorabilidad baja en la mayoría de su territorio. Con el modelo BART, en contraste, la región del Oriente u Amazónica casi en su totalidad de superficie se observan valores cercanos a 0.30. En el mapa de probabilidad de presencia (Figura 2) se observa una representación reducida de presencia en la región Sierra, con valores cercanos a 0.020 y una presencia muy reducida en la región Costa del país.

Los resultados del mapa binario según el modelo GLM (Figura 3a), muestran que *F. oxysporum* está ausente en una superficie de 163 700 km² en el territorio continental de Ecuador (Cuadro 2). Esta ausencia es predominantemente en la región del Oriente y parte de las provincias de Los Ríos, Santo Domingo y Esmeraldas (Figura 3). Por otro lado, se ha identificado la presencia de condiciones adecuadas para el hongo en una superficie de 12 7200 km², siendo predominantemente en la región Sierra y parte de las provincias del Guayas, Manabí, El Oro y parte del perfil costero de Esmeraldas (Figura 3). Para el modelo GAM, la especie tiene una distribución potencial limitada, con condiciones adecuadas en un área de 122 100 km² en la región Sierra y parte de las provincias del Guayas, Manabí y El Oro (Cuadro 2; Figura 3).

Con MaxEnt, *F. oxysporum* tuvo la mayor extensión de área idónea, con 106 800 km² principalmente en la Sierra y las provincias de Guayas, Santa Elena y El Oro (Figura 3). De acuerdo con el modelo Random Forest (RF) (Figura 3), está presente potencialmente en una superficie de solo 47 100 km², mayormente en la región Sierra y en las provincias de Santa Elena, Guayas y Manabí (Figura 3).

Para el modelo GBM (Figura 3), la especie abarca un área de distribución potencial de 110 520 km², en particular en la región Sierra y parte de las provincias del Guayas, Manabí y El Oro (Figura 3). Según el modelo BART (Figura 3), la especie está potencialmente presente en la región Sierra y en una proporción muy baja en la Costa, con un área total de 109 400 km² (Figura 3).

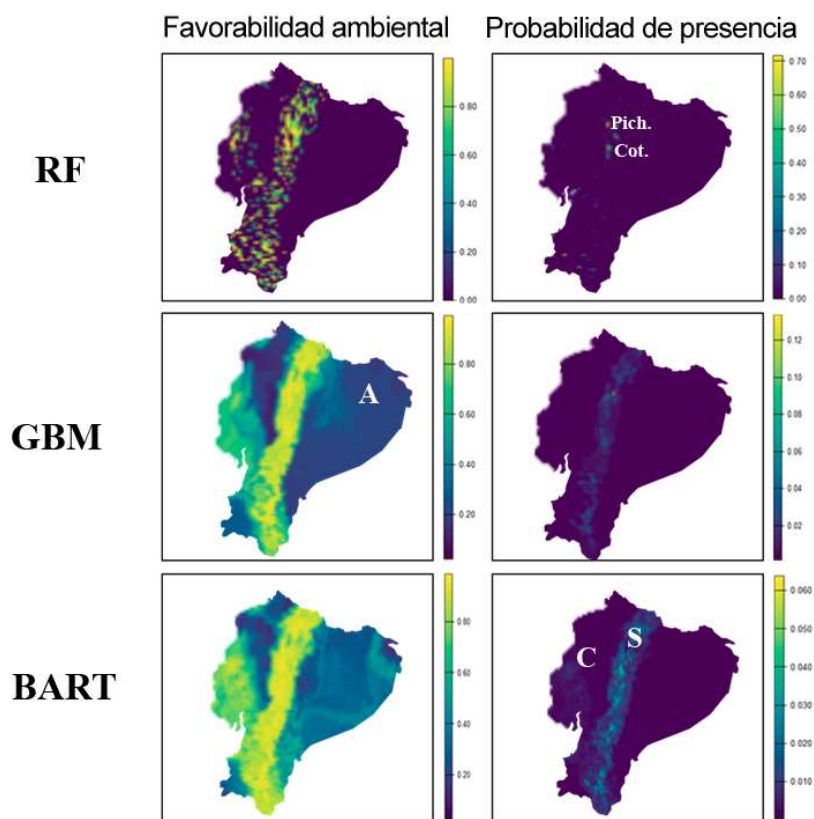


Figura 2. Mapas de favorabilidad y probabilidad de presencia de *Fusarium oxysporum* f. sp. *cubense* en Ecuador continental; RF = bosque aleatorio; GBM = árboles de regresión potenciados; BART = árboles de regresión aditiva bayesianos; Pich. = Pichincha; Cot. = Cotopaxi; A = amazonía; C = costa; S = sierra. Figure 2. Maps of favorability and probability of presence of *Fusarium oxysporum* f. sp. *cubense* in continental Ecuador; RF = random forest; GBM = Boosted Regression Trees; BART = bayesian additive regression trees; Pich. = Pichincha; Cot. = Cotopaxi; A = amazonian; C = costa; S = highlands.

Cuadro 2. Extensión de área de presencia de *Fusarium oxysporum* f. sp. *cubense* en Ecuador continental según diferentes algoritmos de modelación de su nicho climático.

Table 2. Extent of *Fusarium oxysporum* f. sp. *cubense* presence in mainland Ecuador according to different algorithms for modeling its climatic niche.

Modelos	Área de presencia	Área de ausencia	Porcentaje
	----- km ² -----		%
GLM	127 200	163 700	43.73
GAM	122 100	168 800	41.97
MAXENT	184 100	106 800	63.29
Random Forest	47 100	243 800	16.19
GBM	110 520	180 380	37.99
BART	109 400	181 500	37.61

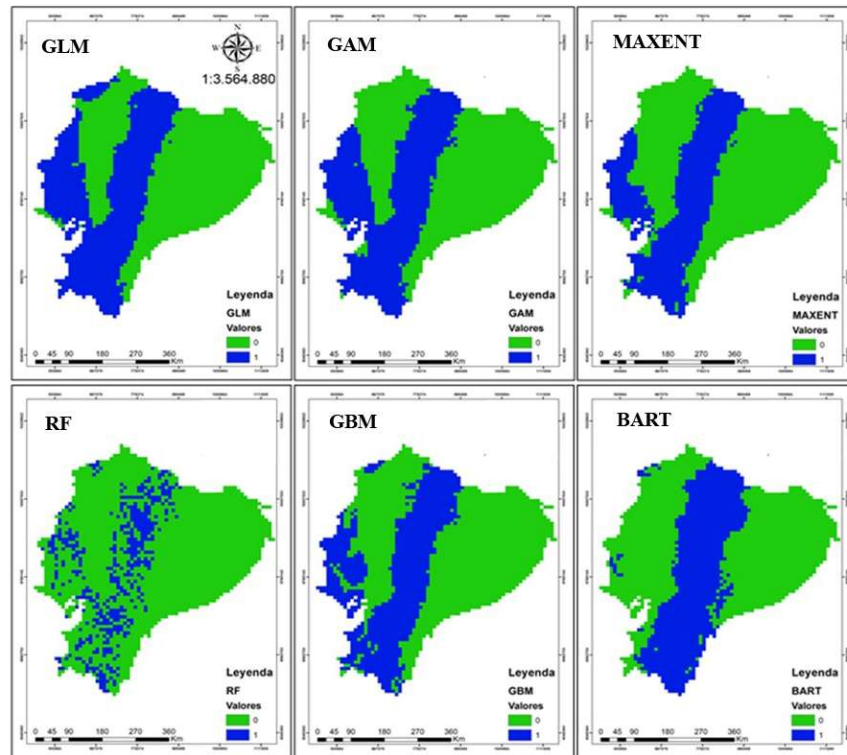


Figura 3. Mapas binarios (1 = presencia; 0 = ausencia) de distribución potencial de *Fusarium oxysporum* f. sp. *cubense* en Ecuador continental, obtenidos con diferentes modelos; GLM = modelo lineal generalizado; GAM = modelo aditivo generalizado; MAXENT = máxima entropía; RF = bosques aleatorios; GBM = árboles de regresión potenciados; BART = árboles de regresión aditiva bayesianos.

Figure 3. Binary maps (1 = presence; 0 = absence) of *Fusarium oxysporum* f. sp. *cubense* distribution in continental Ecuador, obtained with different models; GLM = generalized linear model; GAM = generalized additive model; MAXENT = maximum entropy; RF = random forests; GBM = Boosted Regression Trees; BART = bayesian additive regression trees.

Aunque los MDE pueden ser herramientas valiosas para entender la presencia de las especies en un área, resulta importante evaluar su confiabilidad antes de utilizarlos para tomar decisiones de conservación y manejo. La calidad de los datos de entrada y las opciones metodológicas influyen en la precisión de estos modelos (Sofaer *et al.* 2019). Debido a esto, en la literatura se recomienda utilizar múltiples modelos para evaluar la efectividad y precisión, seleccionando siempre el que muestre los mejores resultados (Norberg *et al.*, 2019; Hao *et al.*, 2020). Además, estos deben ser complementados con distintas métricas para evaluar el rendimiento como pueden ser el AUC, la línea de calibración de Miller, *Kappa* entre otras (Thakuri *et al.*, 2019; Hao *et al.*, 2020).

Con las métricas con un umbral de prevalencia, el mejor modelo en términos de rendimiento fue el *Random Forest* (RF), con un *Correct Classification Rate* (CCR) de 0.96, una sensibilidad de 1, una especificidad de 0.96, una precisión de 0.02, un recall de 1, un *True Skill Statistic* (sTSS) de 0.98 y una *Kappa* de 0.52 (Figura 4). El modelo con el peor rendimiento fue MAXENT (MX), con un CCR de 0.02 y una precisión de 0.00 (Figura 4).

Con las métricas de validación con umbral óptimo, el modelo *Random Forest* (RF) tuvo el mejor rendimiento, con un CCR de 0.98, sensibilidad de 1, especificidad de 0.98, precisión de 0.02, *Recall* de 1, TSS de 0.99 y una *Kappa* de 0.46. Los modelos MX y RF tuvieron una sensibilidad de 1, mientras que con GLM, GBM y BART el valor de *Kappa* fue de 0.46.

Los resultados de la métrica de evaluación AUC para los seis modelos, muestran que con RF se obtuvo el valor máximo de 1, seguido de BART con 0.942, GBM 0.935, MX 0.929, GAM 0.93 y GLM 0.913 (Figura 5). Estos resultados sugieren que RF tuvo la mejor habilidad para distinguir entre las dos clases. En el caso de la línea de calibración de Miller, GLM tuvo una pendiente de 1 y una intercepción de 0, lo que indica un excelente rendimiento (Figura 6). GAM siguió con una pendiente de 1.09 y una intercepción de 0.45 (Figura 6).

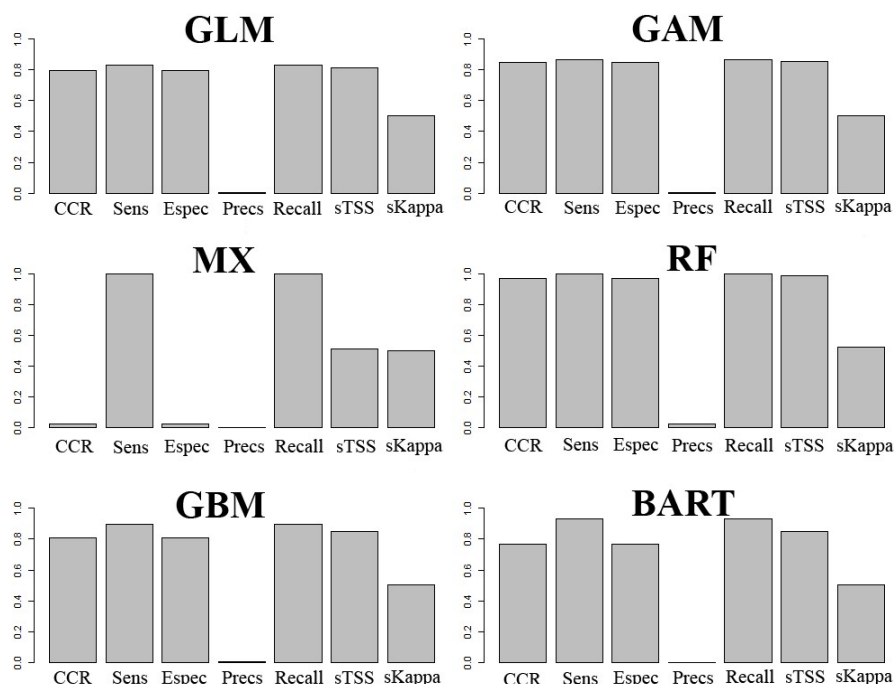


Figura 4. Evaluación del rendimiento de seis modelos (GLM, GAM, MX, RF, GBM y BART) en la predicción de la distribución potencial de *Fusarium oxysporum* f. sp. *cubense* en Ecuador continental, a través de métricas con un umbral de prevalencia (0.000895); CCR = correct classification rate; Sens = sensitivity; Espec = specificity; Precs = precision; recall = recall; sTSS = true skill statistic; sKappa = sKappa.

Figure 4. Evaluation of the performance of six models (GLM, GAM, MX, RF, GBM and BART) in predicting the potential distribution of *Fusarium oxysporum* f. sp. *cubense* in continental Ecuador, through metrics with a prevalence threshold (0.000895); CCR = correct classification rate; Sens = sensitivity; Espec = specificity; Precs = precision; recall = recall; sTSS = true skill statistic; sKappa = sKappa.

GBM tuvo una pendiente de 1.34 y una intercepción de 1.84, mientras que BART tenía una pendiente de 1.54 y una intercepción de 2.61 (Figura 6). El modelo con MX tuvo una pendiente de 0.51 y una intercepción de -6.42, y RF tuvo una pendiente de 14.74 y una intercepción de 8.18 (Figura 6).

Si bien RF alcanza los valores más altos en métricas de discriminación como AUC (1.0) y TSS (0.99), sus resultados de calibración son extremadamente pobres. La Figura 6 muestra que RF tiene una pendiente de 14.74, lo cual indica una pobre correspondencia entre las probabilidades predichas y la tasa real de presencia. En contraste, el modelo GLM muestra una calibración perfecta (pendiente=1, intercepto=0). Se ha comprobado que este modelo sobresale cuando es comparado con otros (Williams *et al.*, 2009). No obstante, es importante tener en cuenta que con GLM también se obtuvieron buenos resultados en la línea de calibración Miller, donde fue el único con el que se obtuvo valor de 1, indicando que este modelo fue muy preciso. Es recomendable tener presente que los MDE muestran resultados diferentes aplicados en distintas especies. Random Forest ha presentado resultados pobres en comparación con Maxent y GLM aplicado a especies en Australia (Shabani, Kumar y Ahmadi, 2016), lo cual puede ser debido a un conjunto de datos específico y a la ecología de la especie en cuestión. Es posible que en algunas regiones geográficas o con ciertas especies, un modelo determinado funcione mejor que otros debido a las características únicas de los datos (Shabani *et al.*, 2016). En otro estudio comparativo entre los modelos MaxEnt, GLM, GBM y RF, se encontró que los modelos ajustados presentaron un alto rendimiento en la predicción de datos de entrenamiento (Ahmadi, Hemamii, Mahmoud y Shabani, 2023). Sin embargo, se observó una variabilidad en la clasificación de datos con los modelos GBM y RF, mostrando una mayor sensibilidad a los datos de entrenamiento y el modelo GLM mostró una especificidad más baja, mientras que MaxEnt fue el más efectivo en la identificación de datos de prueba en procedimientos aleatorios y de ponderación de fondo (Ahmadi *et al.*, 2023). Esto se puede relacionar con la capacidad del modelo para manejar eficazmente datos categóricos y continuos y su capacidad para incorporar información geográfica y ambiental de manera efectiva.

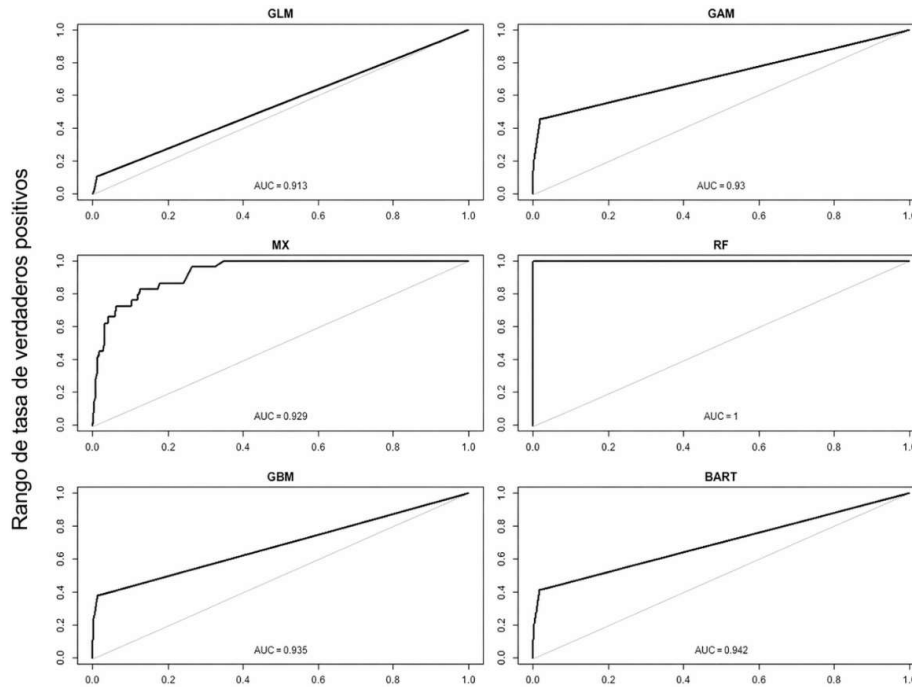


Figura 5. Análisis de la eficacia de los modelos (GLM, GAM, MX, RF, GBM y BART) en la predicción de la distribución potencial de *Fusarium oxysporum* f. sp. *cupense* en Ecuador continental mediante la métrica área bajo la curva (AUC); la línea gris representa la clasificación aleatoria (AUC = 0.5).

Figure 5. Analysis of the effectiveness of the models (GLM, GAM, MX, RF, GBM and BART) in predicting the potential distribution of *Fusarium oxysporum* f. sp. *cupense* in continental Ecuador using the area under the curve (AUC) metric; the gray line represents random classification (AUC = 0.5).

Además, las técnicas de regularización utilizadas por MaxEnt pueden ayudar a evitar el sobreajuste del modelo y mejorar la capacidad de generalización de este (Ab Lah, Yusop, Hashim y Mohd, 2021).

La variación entre los modelos estadísticos y los algoritmos de aprendizaje automático al predecir la distribución potencial de un fitopatógeno, como *Fusarium oxysporum* f. sp. *cupense*, es significativa y se manifiesta tanto en el área geográfica predicha como en el rendimiento y la confiabilidad de las predicciones. Los estudios comparativos, que incluyeron modelos estadísticos (GLM, GAM) y algoritmos de aprendizaje automático (RF, MaxEnt, GBM, y el modelo bayesiano BART), han demostrado que no hay un algoritmo que sea superior a otro de manera universal, ya que el rendimiento predictivo está influenciado por la calidad de los registros de las especies y las variables ambientales utilizadas.

Aunque todos los modelos coinciden en que la región Sierra es la más favorable para el patógeno, los rangos de distribución potencial varían dramáticamente, con MaxEnt prediciendo más del doble del área de presencia que Random Forest (106 800 km² vs. 47 100 km²). La variación en los resultados demostró una compensación entre la discriminación y la calibración. Mientras que los algoritmos de aprendizaje automático basados en árboles (RF, GBM) a menudo sobresalen en la velocidad computacional y en métricas de discriminación general debido a su capacidad para manejar relaciones no lineales, los modelos estadísticos simples como GLM obtuvieron una mayor precisión en la estimación de probabilidades reales, lo cual es fundamental para la toma de decisiones si se requiere que el valor predictivo sea confiable. Esta variación recalca la necesidad de utilizar un enfoque de conjunto (evaluación de múltiples modelos) para evaluar la pertinencia de las herramientas estadísticas y mitigar las limitaciones o sesgos de cualquier modelo individual antes de tomar decisiones de manejo fitosanitario.

Es importante destacar que los modelos utilizados pueden estar sujetos a limitaciones y sesgos, y que los resultados deben ser interpretados con precaución. Por ejemplo, la calidad y la disponibilidad de los datos de entrada, la configuración de los modelos y la elección de las variables predictoras pueden afectar los resultados finales. La configuración de los modelos es muy importante, ya que cada modelo tiene diferentes hiperparámetros que deben ajustarse para optimizar su rendimiento. Si estos hiperparámetros no se ajustan adecuadamente,

puede haber un sesgo hacia ciertos tipos de variables predictoras, lo que puede afectar la precisión de la predicción (Guillera-Arroita *et al.*, 2015). Otro factor importante que puede influir en los resultados es la elección de las variables predictoras. Es importante elegir cuidadosamente las variables que se utilizarán para entrenar el modelo, ya que variables irrelevantes o redundantes pueden reducir su precisión. Por otro lado, variables importantes pero ausentes pueden dejar de capturar información relevante para el modelo, lo que también puede afectar negativamente la precisión de la predicción (Zurell *et al.*, 2020).

Estos modelos requieren una cantidad suficiente de datos de presencia para funcionar adecuadamente (Wunderlich, Lin, Anthony y Petway, 2019). Si la cantidad de datos es insuficiente, pueden tener dificultades para estimar la distribución de la especie de manera precisa y tener una incertidumbre elevada en sus predicciones. En estos casos, puede ser necesario recopilar más datos como se requiere en este estudio, o considerar la implementación de enfoques alternativos, como modelos no bayesianos (Carlson, 2020). En este estudio, aunque el modelo BART no tuvo buenos resultados en las métricas evaluadas, esto no significa que no sea adecuado para la modelación de especies. Es importante tener en cuenta las circunstancias y características específicas de la investigación. En un estudio realizado en Bosnia y Herzegovina (Konowalik y Nosol, 2021), se demostró que BART es efectivo en la modelación de especies con un número reducido de registros de presencia, superando incluso a otros modelos como MaxEnt y RF. El problema radica en que modelos de alta complejidad y flexibilidad como RF y GBM son conocidos por requerir muchos datos y pueden sobreajustarse fácilmente a muestras pequeñas, lo que podría estar ocurriendo aquí. El perfecto AUC de 1.0 para RF es una señal de alerta de posible sobreajuste. Por otro lado, modelos más simples como GLM son más robustos con tamaños de muestra limitados.

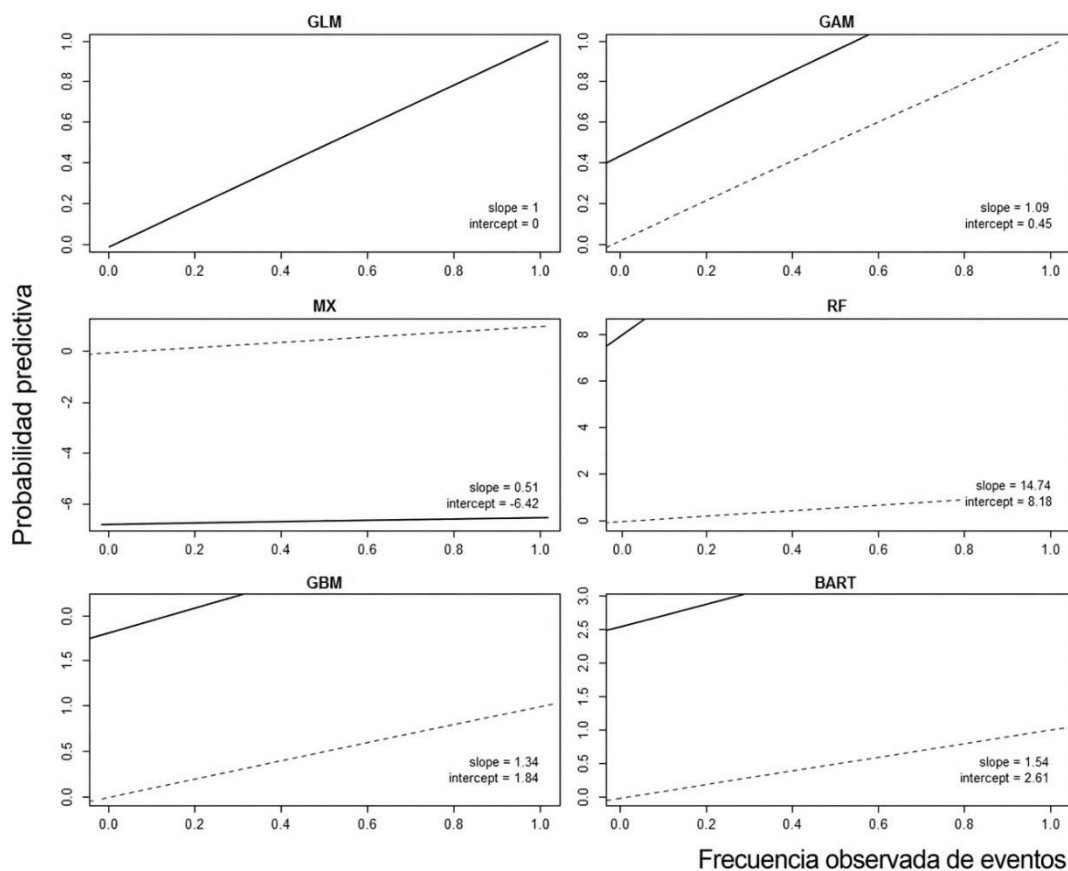


Figura 6. Calibración de los modelos (GLM, GAM, MX, RF, GBM y BART) utilizando la línea de calibración de Miller para medir la precisión de la distribución potencial de *Fusarium oxysporum* f. sp. *cupense* en Ecuador continental.

Figure 6. Calibration of the models (GLM, GAM, MX, RF, GBM and BART) using the Miller calibration line to measure the precision of the potential distribution of *Fusarium oxysporum* f. sp. *cupense* in continental Ecuador.

Por estas razones, es importante evaluar diferentes modelos de distribución de especies en circunstancias específicas, como las relacionadas con actividades productivas. En el caso de Ecuador, por ser un país productor agrícola, es muy vulnerable al impacto de especies invasoras sobre diferentes cultivos y sus ecosistemas. Estas invasiones biológicas representan una de las principales causas de pérdida de biodiversidad y extinción de especies (Linders *et al.*, 2019). También, son responsables de pérdidas económicas directas en la agricultura, y la silvicultura (Courtois, Figuières, Mulier y Weill, 2018). En Ecuador, el 20% de la población se dedica al sector agropecuario y se estima que un 54% desempeña actividades agrícolas (INEC, 2021).

Los resultados sugieren que la presencia o ausencia de este hongo puede estar asociada con las características geográficas, climáticas y ecológicas de cada región. La región de la Sierra, presenta características favorables para el desarrollo y expansión del hongo, que pueden estar asociadas a la temperatura y humedad, así como el suelo y disponibilidad de hospedantes. Por otro lado, la ausencia de la especie en la región Amazónica puede estar relacionada con factores que inhiben su crecimiento y desarrollo, como pueden ser las condiciones climáticas y temperaturas más altas que no son favorables para este hongo, o la ausencia de hospedantes adecuados.

Para las personas que viven en estas regiones con mayor probabilidad de presencia, esta situación representa un problema grave, ya que muchas dependen económicamente de las actividades agrícolas y del éxito de varios cultivos que pueden ser afectados por este fitopatógeno como el plátano y banano. En el periodo 2020, el área destinada a cultivos permanentes fue de 1.4 millones de hectáreas. La región Costa concentra la mayor superficie con un 71.8%, seguida de la Sierra con 17.5% y la Amazonía con solo el 10.6% (INEC, 2021).

En particular, los diferentes modelos coinciden en una baja probabilidad de presencia del patógeno en las provincias de la Amazonía, como Sucumbíos, Orellana, Pastaza y Morona Santiago, lo que sugiere que estas pueden ser idóneas para la agricultura. Sin embargo, debido a que parte de estas áreas son consideradas zonas protegidas, no se pueden establecer grandes cultivos que puedan afectar los ecosistemas de la región. Por lo tanto, se deben buscar alternativas en otras regiones, y en este caso, se sugiere la costa ecuatoriana, que cuenta con la mayor área agrícola activa del país. Aunque los resultados de los diferentes modelos mostraron que varias provincias de la región Costa podrían verse afectadas por la presencia del patógeno, se identificaron ciertos cantones en la provincia de Los Ríos, como Quevedo, Mocache, Ventanas, Quimsaloma, Valencia y Buena Fe, donde la presencia de *F. oxysporum* f. sp. *cubense* tendría una baja probabilidad. Estas áreas pueden ser consideradas como las más adecuadas dentro de la región Costa para el desarrollo de cultivos agrícolas. Es necesario mencionar, además, que este patógeno está fuertemente influenciado no solo por el clima, sino también por las propiedades del suelo (pH, textura, materia orgánica) y, fundamentalmente, por la distribución de sus plantas hospedadoras (banano, tomate, etc.), por ello la omisión de variables predictoras como estas es una limitación ecológica. Los modelos resultantes han descrito un nicho climático potencial, más que un nicho realizado o una distribución real que está restringida por la disponibilidad de suelo y hospedadores adecuados. Esto podría explicar algunas de las distribuciones tan amplias predichas, especialmente en la Sierra, donde la presencia del patógeno estaría ligada a zonas agrícolas específicas. Las áreas de riesgo climático predichas solo serán relevantes si existen los cultivos hospedadores y las condiciones edáficas adecuadas.

CONCLUSIONES

La eficacia de los modelos varía significativamente según la métrica utilizada ya que el rendimiento predictivo está influenciado por la calidad de los registros de la especie y las variables ambientales. A pesar de ello, el modelo Random Forest tuvo buen rendimiento en la mayoría de las métricas empleadas, principalmente en discriminación; sin embargo, el Modelo Lineal Generalizado presentó mejor precisión y confiabilidad, con una calibración excelente en la correspondencia entre las probabilidades predichas y las frecuencias de ocurrencia observadas. Los seis modelos utilizados (BART, GLM, GAM, RF, MaxEnt y GBM) demostraron su eficacia al predecir la distribución potencial de *F. oxysporum* f. sp. *cubense*.

Los modelos identificaron las áreas con mayor y menor riesgo de presencia de *F. oxysporum* f. sp. *cubense* en el territorio ecuatoriano. El patógeno tiene altas probabilidades de extenderse en toda la región Sierra y parte de las provincias de la Costa. Todos los modelos mostraron consistentemente una favorabilidad ambiental superior a 0.90 en la región Sierra. Esta presencia se asocia a las características geográficas, climáticas y ecológicas de la Sierra, tales como la combinación de temperatura y humedad, además de la disponibilidad de hospederos.

La identificación de las regiones con alta y baja probabilidad de presencia de esta especie es de gran relevancia para el monitoreo y control del patógeno en Ecuador continental, lo que permite tomar medidas preventivas para minimizar su impacto en la agricultura y el ambiente. *Fusarium oxysporum* f. sp. *cubense* posee una alta capacidad de adaptación que le permite distribirse en diversas condiciones bioclimáticas dentro del

espacio geográfico ecuatoriano (temperatura máxima 10 °C a 37 °C y precipitación anual 139 mm a 3782 mm). A pesar de los resultados obtenidos, se sugiere continuar la investigación con un mayor número de registros de presencia de la especie para mejorar la precisión y reducir la incertidumbre en las predicciones. El estudio es relevante como línea base, especialmente al explorar modelos bayesianos (BART).

DECLARACIÓN DE ÉTICA

No aplicable.

CONSENTIMIENTO PARA PUBLICACIÓN

No aplicable.

DISPONIBILIDAD DE DATOS

Los conjuntos de datos utilizados o analizados durante el estudio actual están disponibles por el autor correspondiente a solicitud razonable.

CONFLICTO DE INTERESES

Los autores declaramos que no tenemos intereses en competencia.

FINANCIACIÓN

No aplicable.

CONTRIBUCIÓN DE LOS AUTORES

Conceptualización: A.L.B.M., Y.F.S., A.H.P.V., J.J.J.S. y A.S.R.; Metodología: A.L.B.M., Y.F.S., A.H.P.V., J.J.J.S. y A.S.R.; Investigación: A.L.B.M., Y.F.S., A.H.P.V., J.J.J.S. y A.S.R.; Escritura - preparación del borrador original: A.L.B.M., Y.F.S., A.H.P.V., J.J.J.S. y A.S.R.; Escritura - revisión y edición: A.L.B.M., Y.F.S., A.H.P.V. y A.S.R.; Administración del proyecto: Y.F.S.

AGRADECIMIENTOS

Se agradece a la Universidad Técnica Estatal de Quevedo por el financiamiento del proyecto FOCICYT 8va Convocatoria. A los revisores por las sugerencias dadas.

LITERATURA CITADA

- Ab Lah, N. Z., Yusop, Z., Hashim, M., & Mohd, J. (2021). Predicting the habitat suitability of *Melaleuca cajuputi* based on the MaxEnt species distribution model. *Forests*, 12(11), 1449. <https://doi.org/10.3390/f12111449>
- Acurio, A., Rafael, V., & Dangles, O. (2010). Biological invasions in the Amazonian Tropical Rain Forest: the case of Drosophilidae (Insecta, Diptera) in Ecuador, South America. *Biotropica*, 42(6), 717-723.
- Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 43(6), 1223-1232. <https://doi.org/10.1111/j.1365-2664.2006.01214.x>
- Ahmadi, M., Hemamii, M., Mahmoud, M., & Shabani, F. (2023). MaxEnt brings comparable results when the input data are. *Ecology and Evolution*, 13(2), 1-13. <https://doi.org/10.1002/ece3.9827>
- Araújo, M. B., & New, M. (2007). Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, 22(1), 42-47. <https://doi.org/10.1016/j.tree.2006.09.010>
- Araújo, M. B., Anderson, R. P., Márcia Barbosa, A., Beale, C. M., Dormann, C. F., Early, R., ... & Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. *Science Advances*, 5(1), eaat4858.
- Baquero, R. A., Barbosa, A. M., Ayllón, D., Guerra, C., Sánchez, E., Araújo, M. B., & Nicola, G. G. (2021). Potential distributions of invasive vertebrates in the Iberian Peninsula under projected changes in climate extreme events. *Diversity and Distributions*, 27(11), 2262-2276. <https://doi.org/10.1111/ddi.13401>

- Barbosa, M. A., Real, R., Muñoz, A. R., & Brown, J. A. (2013). New measures for assessing model equilibrium and prediction mismatch in species distribution models. *Diversity and Distributions*, 19(10), 1333-1338. <https://doi.org/10.1111/ddi.12100>
- Burgess, L. W. (1981). General ecology of the Fusaria. In P. E. Nelson, T. A. Toussoun, & R. J. Cook (Eds.). *Fusarium: Diseases, biology, and taxonomy* (pp. 225-235). University Park, PA, USA: Pennsylvania State University Press.
- Cabrera, M., Garmendia, G., Rufo, C., Pereyra, S., & Vero, S. (2020). Trichoderma atroviride como controlador biológico de fusariosis de espiga de trigo mediante la reducción del inóculo primario en rastrojo. *Terra Latinoamericana*, 38(3), 629-651. <https://doi.org/10.28940/terra.v38i3.664>
- Carlson, C. J. (2020). embarcadero: Species distribution modelling with Bayesian additive regression trees in r. *Methods in Ecology and Evolution*, 11(7), 850-858. <https://doi.org/10.1111/2041-210X.13389>
- Chamberlain, S. (2020). *scrubr: Clean biological occurrence records. R package version 0.3.2*. London, UK: Scott Chamberlain.
- Correll, J. C. (1991). The relationship between formae speciales, races, and vegetative compatibility groups in *Fusarium oxysporum*. *Phytopathology*, 81(9), 1061-1064.
- Courtois, P., Figuières, C., Mulier, C., & Weill, J. (2018). A cost-benefit approach for prioritizing invasive species. *Ecological Economics*, 146, 607-620. <https://doi.org/10.1016/j.ecolecon.2017.11.037>
- Elith, J., & Leathwick, J. (2017). *Boosted regression trees for ecological modeling. R documentation*. Canberra, Australia: Commonwealth Scientific and Industrial Research Organisation (CSIRO).
- Ferrer-Sanchez, Y., Jacho-Saa, W. R., Urdánigo Zambrano, J. P., Abasolo-Pacheco, F., Plasencia-Vázquez, A. H., Zambrano-Mero, G. J., ... & Estrella Bravo, G. V. (2021). Invasiones biológicas en agroecosistemas de Ecuador Continental: nicho ecológico de especies exóticas y cultivos agrícolas bajo riesgo. *Acta Biológica Colombiana*, 26(3), 352-364. <https://doi.org/10.15446/abc.v26n3.81765>
- Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12), 4302-4315. <https://doi.org/10.1002/joc.5086>
- Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24(1), 38-49. <https://doi.org/10.1017/S0376892997000088>
- GBIF (Global Biodiversity Information Facility). (2024). GBIF: The Global Biodiversity Information Facility. Copenhagen, Denmark: GBIF Secretariat. Consultada el 13 de marzo, 2024, desde <https://www.gbif.org>
- Guillera-Arroita, G., Lahoz-Monfort, J., Elith, J., Gordon, A., Kujala, H., Lentini, P., ... & Wintle, B. (2015). Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography*, 24(3), 276-292. <https://doi.org/10.1111/geb.12268>
- Hao, T., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2020). Testing whether ensemble modelling is advantageous for maximising predictive performance of species distribution models. *Ecography*, 43(4), 549-558. <https://doi.org/10.1111/ecog.04890>
- Hastie, T. (2015). *gam: Generalized additive models. R package version 1.20*. Palo Alto, CA, USA: Trevor Hastie.
- INEC (Instituto Nacional de Estadística y Censos). (2021). Boletín técnico: Encuesta de superficie y producción agropecuaria continua (ESPAC) 2020. Quito, Ecuador: INEC.
- Kleemann, J., Koo, H., Hensen, I., Mendieta-Leiva, G., Kahnt, B., Kurze, C., ... & Fürst, C. (2022). Priorities of action and research for the protection of biodiversity and ecosystem services in continental Ecuador. *Biological Conservation*, 265, 109404. <https://doi.org/10.1016/j.biocon.2021.109404>
- Konowalik, K., & Nosol, A. (2021). Evaluation metrics and validation of presence-only species distribution models based on distributional maps with varying coverage. *Scientific Reports*, 11(1), 1482. <https://doi.org/10.1038/s41598-020-80062-1>
- Kosicki, J. Z. (2020). Generalised Additive Models and Random Forest Approach as effective methods for predictive species density and functional species richness. *Environmental and Ecological Statistics*, 24, 273-292. <https://doi.org/10.1007/s10651-020-00445-5>
- Lake, T. A., Briscoe, N. J., & Wintle, B. A. (2020). Evaluation of species distribution models using cross-validation: The importance of threshold selection and performance metrics. *Ecological Modelling*, 431, 109164. <https://doi.org/10.1016/j.ecolmodel.2020.109164>
- Legendre, P., & Legendre, L. (2012). *Numerical ecology* (3rd ed.). Amsterdam, Netherlands: Elsevier.
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3), 18-22.
- Linders, T. E. W., Schaffner, U., Eschen, R., Abebe, A., Choge, S. K., Nigatu, L., ... & Allan, E. (2019). Direct and indirect effects of invasive species: Biodiversity loss is a major mechanism by which an invasive tree affects ecosystem functioning. *Journal of Ecology*, 107(6), 2660-2672. <https://doi.org/10.1111/1365-2745.13268>
- Miller, M. E., Hui, S. L., & Tierney, W. M. (1991). Validation techniques for logistic regression models. *Statistics in Medicine*, 10(8), 1213-1226. <https://doi.org/10.1002/sim.4780100805>
- Norberg, A., Abrego, N., Blanchet, F. G., Adler, F. R., Anderson, B. J., Anttila, J., ... & Ovaskainen, O. (2019). A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. *Ecological Monographs*, 89(3), e01370.
- Pearce, J., & Ferrier, S. (2000). Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, 133(3), 225-245. [https://doi.org/10.1016/S0304-3800\(00\)00322-7](https://doi.org/10.1016/S0304-3800(00)00322-7)
- Pearson, R. G., Thuiller, W., Araújo, M. B., Martínez-Meyer, E., Brotons, L., McClean, C., ... & Lees, D. C. (2006). Model-based uncertainty in species range prediction. *Journal of Biogeography*, 33(10), 1704-1711. <https://doi.org/10.1111/j.1365-2699.2006.01460.x>
- Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M., & Araújo, M. B. (2011). *Ecological niches and geographic distributions*. Princeton, NJ, USA: Princeton University Press.
- Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., & Blair, M. E. (2017). Opening the black box: an open-source release of Maxent. *Ecography* 40(7), 887-893. <https://doi.org/10.1111/ecog.03049>
- Rapacciuolo, G., Roy, D. B., Gillings, S., Fox, R., Walker, K., & Purvis, A. (2012). Climatic associations of British species distributions show good transferability in time but low predictive accuracy for range change. *PLoS One*, 7(7), e40212. <https://doi.org/10.1371/journal.pone.0040212>
- Real, R., Barbosa, A. M., & Vargas, J. M. (2006). Obtaining environmental favourability functions from logistic regression. *Environmental and Ecological Statistics*, 13(2), 237-245. <https://doi.org/10.1007/s10651-005-0003-3>
- Real, R., Barbosa, A. M., & Bull, J. W. (2017). Species distributions, quantum theory, and the enhancement of biodiversity measures. *Systematic Biology*, 66(3), 453-462. <https://doi.org/10.1093/sysbio/syw072>
- Ron, S. R. (2005). Predicting the distribution of the amphibian pathogen Batrachochytrium dendrobatidis in the New World. *Biotropica*, 37(2), 209-221.
- R Core Team (2024). *R: A language and environment for statistical computing* (version 4.4.1). Vienna, Austria: R Foundation for Statistical Computing.
- Smith, S. N. (2007). An overview of ecological and habitat aspects in the genus *Fusarium* with special emphasis on the soil-borne pathogenic forms. *Plant Pathology Bulletin*, 16(3), 97-120.
- Shabani, F., Kumar, L., & Ahmadi, M. (2016). A comparison of absolute performance of different correlative and mechanistic species distribution models in an independent area. *Ecology and Evolution*, 6(16), 5973-5986. <https://doi.org/10.1002/ece3.2332>

- Sillero, N., & Barbosa, A. M. (2021). Common mistakes in ecological niche models. *International Journal of Geographical Information Science*, 35(2), 213-226. <https://doi.org/10.1080/13658816.2020.1798968>
- Sofaer, H. R., Jarnevich, C. S., Pearse, I. S., Smyth, R. L., Auer, S., Cook, G. L., ... & Hamilton, H. (2019). Development and delivery of species distribution models to inform decision-making. *BioScience*, 69(7), 544-557. <https://doi.org/10.1093/biosci/biz045>
- Staudte, I. R., Navarro, L. M., & Pereira, H. M. (2020). Range size predicts the risk of local extinction from habitat loss. *Global Ecology and Biogeography*, 29(1), 16-25. <https://doi.org/10.1111/geb.13003>
- Thakuri, S., Shrestha, P., Deuba, M., Shah, P., Bhandari, O., & Shrestha, S. (2019). Potential habitat modeling of water hyacinth in lakes of nepal using maxent algorithm. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Science*, 4(2), 103-110. <https://doi.org/10.5194/isprs-annals-IV-5-W2-103-2019>
- Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Aroita, G. (2018). blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*, 10(2), 225-232. <https://doi.org/10.1111/2041-210X.13107>
- Williams, J. N., Seo, C., Thorne, J., Nelson, J. K., Erwin, S., O'Brien, J. M., & Schwartz, M. W. (2009). Using species distribution models to predict new occurrences for rare plants. *Diversity and Distributions*, 15(4), 565-576. <https://doi.org/10.1111/j.1472-4642.2009.00567.x>
- Wunderlich, R., Lin, Y. P., Anthony, J., & Petway, J. (2019). Two alternative evaluation metrics to replace the true skill statistic in the assessment of species distribution models. *Nature Conservation*, 35, 7-116. <https://doi.org/10.3897/natureconservation.35.33918>
- Zurell, D., Franklin, J., König, C., Bouchet, P. J., Dormann, C. F., Elith, J., ... & Merow, C. (2020). A standard protocol for reporting species distribution models. *Ecography*, 43(9), 1261-1277. <https://doi.org/10.1111/ecog.04960>