



UNIVERSIDAD TÉCNICA ESTATAL DE QUEVEDO

FACULTAD DE CIENCIAS DE LA INGENIERÍA

CARRERA INGENIERÍA AMBIENTAL

Proyecto de investigación previo
a la obtención del título de
Ingeniero Ambiental

Título del proyecto de investigación

MÉTODOS ESTADÍSTICOS BAYESIANOS APLICADOS EN LA MODELACIÓN
DE LA DISTRIBUCIÓN POTENCIAL DE *Fusarium oxysporum* EN ECUADOR
CONTINENTAL

Autor:

Jorge Josué Jacho Saa

Directora del proyecto de investigación

Dra. Yarelys Ferrer Sánchez, PhD

Codirector

Dr. Alexis Herminio Plasencia Vázquez, PhD.

Quevedo - Los Ríos - Ecuador

2023



DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS

Yo, **Jorge Josué Jacho Saa**, declaro que la investigación aquí descrita es de mi autoría; que no ha sido previamente presentada para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

La Universidad Técnica Estatal de Quevedo, puede hacer uso de los derechos correspondientes a este documento, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

Jorge Josue Jacho Saa
C.C. # 0929038628



CERTIFICACIÓN DE CULMINACIÓN DEL PROYECTO DE INVESTIGACIÓN

La suscrita, **Dra. Yarelys Ferrer Sánchez, PhD**, docente de la Universidad Técnica Estatal de Quevedo, certifica que el estudiante **Jorge Josué Jacho Saa**, realizó el Proyecto de Investigación de grado titulado “**MÉTODOS ESTADÍSTICOS BAYESIANOS APLICADOS EN LA MODELACIÓN DE LA DISTRIBUCIÓN POTENCIAL DE *Fusarium oxysporum* EN ECUADOR CONTINENTAL**”, bajo mi dirección, habiendo cumplido con las disposiciones reglamentarias establecidas para el efecto.

.....
Dra. Yarelys Ferrer Sanchez, PhD.
DIRECTORA DEL PROYECTO DE INVESTIGACIÓN

.....
Dr. Alexis Herminio Plasencia Vázquez
CODIRECTOR DEL PROYECTO DE INVESTIGACIÓN



CERTIFICADO DEL REPORTE DE LA HERRAMIENTA DE PREVENCIÓN DE COINCIDENCIA Y/O PLAGIO ACADÉMICO

La suscrita **Dra. Yarelys Ferrer Sánchez, PhD**, docente de la Universidad Técnica Estatal de Quevedo certifica que el estudiante **Jorge Josué Jacho Saa**, realizó el proyecto de investigación con el tema de: **“MÉTODOS ESTADÍSTICOS BAYESIANOS APLICADOS EN LA MODELACIÓN DE LA DISTRIBUCIÓN POTENCIAL DE *Fusarium oxysporum* EN ECUADOR CONTINENTAL**”, fue ingresado al sistema URKUND y presentó el 5% de similitud, considerando el Reglamento e Instructivos de Proyecto de Investigación de la Universidad Técnica Estatal de Quevedo.


by Turnitin

Document Information

Analyzed document	JACHO TESIS (1).docx (D158394379)
Submitted	2023-02-11 04:36:00
Submitted by	
Submitter email	yferrer@uteq.edu.ec
Similarity	5%
Analysis address	fabasolo.uteq@analysis.urkund.com

Sources included in the report

SA

UNIVERSIDAD TÉCNICA ESTATAL DE QUEVEDO / TESIS-FUSARIUM OXYSPORUM.docx
Document TESIS-FUSARIUM OXYSPORUM.docx (D98327756)
Submitted by: yferrer@uteq.edu.ec

 3

.....
Dra. Yarelys Ferrer Sánchez, PhD

DIRECTORA DEL PROYECTO DE INVESTIGACIÓN

.....
Dr. Alexis Herminio Plasencia Vázquez

CODIRECTOR DEL PROYECTO DE INVESTIGACIÓN



UNIVERSIDAD TÉCNICA ESTATAL DE QUEVEDO
FACULTAD DE CIENCIAS DE LA INGENIERÍA
CARRERA INGENIERÍA AMBIENTAL

PROYECTO DE INVESTIGACIÓN

Título:

“Métodos estadísticos bayesianos aplicados en la modelación de la distribución potencial de *Fusarium oxysporum* en Ecuador continental”

Presentado al Consejo Directivo como requisito previo a la obtención del título de Ingeniería Ambiental.

Aprobado por:

PRESIDENTE DEL TRIBUNAL

Blgo. Juan Pablo Urdánigo MSc



Firmado electrónicamente por:
**JULIO CESAR PAZMINO
RODRIGUEZ**

MIEMBRO DEL TRIBUNAL

Ing. Julio Cesar Pazmiño MSc



Firmado electrónicamente por:
**YOMBER JOSE
MONTILLA LOPEZ**

MIEMBRO DEL TRIBUNAL

Ing. Yomber Montilla MSc

QUEVEDO – LOS RÍOS – ECUADOR

2023

AGRADECIMIENTOS

A mis padres y hermano, por su apoyo y motivación a lo largo de toda mi etapa académica para lograr mi objetivo de titularme.

A la Universidad Técnica Estatal de Quevedo, que me ha permitido formar parte de tan prestigiosa institución para culminar mis estudios académicos.

Agradezco a la Dra. Yarelys Ferrer Sánchez y al Dr. Alexis Herminio Plasencia Vázquez, mis tutores de proyecto de investigación, quienes han brindado su apoyo incondicional para la realización con éxito de mi tesis de grado.

A mis docentes y compañeros, quienes han participado constantemente en el desarrollo de mis actitudes académicas.

Jorge Josué Jacho Saa.

DEDICATORIA

A mi madre, Zao Saa, a quien le debo todo lo que he logrado en mi vida y por formarme
como persona.

A mi hermano, Wilmer Jacho, que siempre me ha apoyado en lograr culminar mis
estudios.

A mi familia, quienes me han ayudado en todo momento.

Jorge Josué Jacho Saa.

RESUMEN

Evaluar la eficacia de los modelos de distribución de especies es crucial para predecir la distribución futura de una especie. En esta investigación, se evaluó la eficiencia del modelo bayesiano, Bayesian Additive Regression Trees (BART), frente a modelos estadísticos: Generalized Linear Model (GLM), Generalized Additive Model (GAM) y algoritmos de aprendizaje automáticos: Random Forest (RF), Maximum Entropy (MAXENT) y Gradient Boosting Machine (GBM) para modelar la distribución potencial de *Fusarium oxysporum* en Ecuador continental. Se utilizaron registros de presencia de la especie (GBIF) y variables bioclimáticas (WorldClim), para construir los modelos en RStudio a una resolución espacial de 10 x 10 km². Los modelos se evaluaron mediante AUC, línea de calibración de Miller y clasificación basada en umbrales. Los mapas de probabilidad de presencia de cada modelo se exportaron de RStudio a ArcGis para generar mapas binarios y calcular el área de presencia y ausencia de la especie en Ecuador. Los resultados indican que el modelo RF tuvo el mejor rendimiento en la mayoría de las métricas, con umbral de prevalencia: Especificidad (0.93), estadística de habilidad verdadera (TSS) (0.98); con umbral óptimo: Especificidad (0.98), TSS (0.99); para AUC (1). Sin embargo, GLM presentó buenos resultados en la línea de calibración Miller (1), y la mayor área de presencia del fitopatógeno (127200 km²) y la menor RF (47100 km²). Aunque los modelos bayesianos fueron eficientes en predecir la distribución de *F. oxysporum*, al igual que los demás métodos, se deben realizar investigaciones futuras con más registros de presencia para mejorar la precisión.

Palabras claves: Algoritmos de Aprendizaje Automático, enfermedad de Panamá, MAXENT, BART, GAM, GLM, Random Forest.

ABSTRACT

Evaluating the efficiency of species distribution models is crucial for predicting the future distribution of a species. In this research, the efficiency of the Bayesian model, Bayesian Additive Regression Trees (BART), was evaluated against statistical models: Generalized Linear Model (GLM), Generalized Additive Model (GAM) and machine learning algorithms: Random Forest (RF), Maximum Entropy (MAXENT), Gradient Boosting Machine (GBM) to model the potential distribution of *Fusarium oxysporum* in continental Ecuador. Species occurrence records (GBIF) and bioclimatic variables (WorldClim) were used to build the models in RStudio at a spatial resolution of 10 x 10 km². Models were evaluated using AUC, Miller calibration line and threshold-based classification. Presence probability maps from each model were exported from RStudio to ArcGis to generate binary maps and calculate the area of presence and absence of the species in Ecuador. The results indicate that the RF model had the best performance in most metrics, with threshold prevalence: Specificity (0.93), true skill statistic (TSS) (0.98); with optimal threshold: Specificity (0.98), TSS (0.99); for AUC (1). However, GLM presented good results in the Miller calibration line (1), and the largest area of phytopathogen presence (127200 km²) and the smallest RF (47100 km²). Although Bayesian models were efficient in predicting the distribution of *F. oxysporum*, as were the other methods, future research should be conducted with more occurrence records to improve accuracy.

Keywords: Machine Learning Algorithms, Panama Disease, MAXENT, BART, GAM, GLM, Random Forest.

TABLA DE CONTENIDO

CUBIERTA Y PORTADA	i
DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS	ii
CERTIFICACIÓN DE CULMINACIÓN DEL PROYECTO DE INVESTIGACIÓN.....	iii
CERTIFICADO DEL REPORTE DE LA HERRAMIENTA DE PREVENCIÓN DE COINCIDENCIA Y/O PLAGIO ACADÉMICO.....	iv
Certificación de aprobación por tribunal de sustentación	v
AGRADECIMIENTOS.....	vi
DEDICATORIA.....	vii
RESUMEN	viii
ABSTRACT	ix
TABLA DE CONTENIDO	x
ÍNDICE DE FIGURAS	xiv
ÍNDICE DE ANEXOS	xv
CÓDIGO DUBLIN.....	xvi
INTRODUCCIÓN.....	1
CAPÍTULO I.....	3
CONTEXTUALIZACIÓN DE LA INVESTIGACIÓN	3
1.1. Problema de investigación	4
1.1.1. Planteamiento del problema	4
1.1.1.1. Diagnóstico.....	4
1.1.1.2. Pronóstico.....	5
1.1.2. Formulación del problema.....	5
1.1.3. Sistematización del problema.....	5
1.2. Objetivos.....	6
1.2.1. Objetivo General.....	6
1.2.2. Objetivos Específicos	6

1.3. Justificación	6
CAPÍTULO II.....	8
FUNDAMENTACIÓN TEÓRICA DE LA INVESTIGACIÓN	8
2.1. Marco conceptual.....	9
2.1.1. Modelo de distribución de especies.....	9
2.1.1.1. Modelos aprendizaje automático.....	9
2.1.1.2. Modelos estadísticos	9
2.1.1.3. Modelos Lineales Generalizados (GLM).....	9
2.1.1.4. Modelos aditivos generalizados (GAM)	10
2.1.1.5. Bosques Aleatorios (RF)	10
2.1.1.6. Máquina de Mejora por Gradiente (GBM)	10
2.1.1.7. Árboles de Regresión Aditivos Bayesianos (BART).....	10
2.1.1.8. Máxima Entropía (MaxEnt)	11
2.1.2. <i>Fusarium oxysporum</i>	12
2.1.2.1. Reproducción	12
2.1.2.2. Clasificación de <i>Fusarium oxysporum</i>	12
2.1.2.3. Variantes de <i>Fusarium oxysporum</i>	13
2.1.3. Nicho ecológico.....	13
2.1.3.1. Nicho Fundamental	14
2.1.3.2. Nicho realizado	14
2.1.3.3. Diagrama BAM.....	14
2.2. Marco referencial	15
2.2.1. Participación de Organizaciones Internacionales	15
2.2.2. Impacto de FW en la agricultura	16
2.2.3. <i>Fusarium oxysporum</i> f. sp. <i>cubense</i> Tropical Race 4(TR4) en el mundo	17
2.2.4. Presencia de <i>Fusarium oxysporum</i> en Ecuador	17
2.2.5. Hitos relevantes en el desarrollo de modelos de distribución de especies	17

2.2.6.	Problemas de coincidencia entre datos	18
2.2.6.1.	La interacción entre tipos de datos y sesgos en SDM	18
2.2.7.	Construcción de un SDM	19
2.2.7.1.	Datos reales y datos virtuales	19
CAPÍTULO III		22
MÉTODOLOGÍA DE LA INVESTIGACIÓN		22
3.1.	Localización	23
3.2.	Tipo de investigación	24
3.2.1.	Diagnóstica	24
3.2.2.	Explicativa	24
3.3.	Métodos de investigación	24
3.3.1.	Método inductivo	24
3.3.2.	Método analítico	25
3.4.	Fuentes de recopilación de información	25
3.5.	Diseño de investigación	25
3.5.1.	Objetivo 1: Modelar el nicho ecológico de <i>F. oxysporum</i> bajo diferentes enfoques estadísticos	27
3.5.1.1.	Fuentes de obtención para los datos de presencia de la especie	27
3.5.1.2.	Obtención de variables bioclimáticas	27
3.5.1.3.	Procesamiento de datos en RStudio	28
3.5.2.	Objetivo 2: Caracterizar la distribución potencial de <i>F. oxysporum</i> en Ecuador continental según los métodos de modelación empleados	32
3.5.3.	Objetivo 3: Evaluar las métricas de rendimiento de los modelos de <i>F. oxysporum</i> según los enfoques estadísticos empleados	32
3.6.	Instrumentos de investigación	33
3.7.	Tratamiento de los datos	33
3.8.	Recursos humanos y materiales	34
3.8.1.	Recursos humanos	34

3.8.1.1. Esquipo de dirección y coordinación	34
3.8.2. Materiales	35
CAPÍTULO IV	36
RESULTADOS Y DISCUSIÓN	36
4.1. Modelación del nicho ecológico de <i>Fusarium oxysporum</i> bajo diferentes enfoques estadísticos	37
4.1.1. Estadísticos descriptivos de las variables bioclimáticas.....	37
4.1.2. Variables climáticas con mayor influencia.....	38
4.1.3. Distribución potencial y presencia de <i>Fusarium oxysporum</i> en Ecuador continental: Análisis a través de mapas de favorabilidad y presencia	39
4.2. Caracterizar la distribución potencial de <i>F. oxysporum</i> en Ecuador continental según los métodos de modelación empleados.	43
4.2.1. Mapas binarios presencia-ausencia	43
4.3. Evaluación de las métricas de rendimiento de los modelos de <i>F. oxysporum</i> según los enfoques estadísticos empleados	45
4.3.1. Métricas de validación con umbral de prevalencia	45
4.3.2. Métricas de validación con umbral optimo	46
4.3.3. Área bajo la Curva ROC (AUC).....	47
4.3.4. Línea de calibración Miller.....	48
4.4. Discusión.....	49
CAPÍTULO V.....	54
CONCLUSIONES Y RECOMENDACIONES	54
5.1. Conclusiones.....	55
5.2. Recomendaciones	56
CAPÍTULO VI	57
BIBLIOGRAFÍA	57
6.1. Bibliografía.....	58
CAPÍTULO VII.....	66

ANEXOS	66
--------------	----

ÍNDICE DE TABLAS

Tabla 1. Datos Bioclimáticos en WorldClim: Una vista de las variables disponibles	27
Tabla 2. Recursos empleados en el proyecto de investigación.....	35
Tabla 3. Estadísticos descriptivos de las variables bioclimáticas que caracterizan el área de donde se extrajeron los registros de presencia de <i>F. oxysporum</i> (N=183).....	37

ÍNDICE DE FIGURAS

Figura 1. Diagrama BAM.....	15
Figura 2. Interacción entre tipos de datos y sesgos en SDM..	19
Figura 3. Mapa de Ecuador continental.....	23
Figura 4. Metodología en acción: Mapa de Flujo del proyecto de investigación. Elaborado: Autor.....	26
Figura 5. Distribución Global de <i>F. oxysporum</i> : Registros de Presencia.....	29
Figura 6. Presencia de <i>F. oxysporum</i> en Ecuador, Perú y Colombia: Registros Gráficos. 30	
Figura 7. Mapas de Variables Bioclimáticas en Ecuador, Perú y Colombia: Altitud y bio 1 a bio 15	31
Figura 8. Análisis de la influencia de las variables bioclimáticas en la selección del modelo GLM: Una comparación visual de la precipitación mensual (bio 5), temperatura promedio máxima (bio 12) y mínima (bio 7) y precipitación total anual (bio 3).	38
Figura 9. Mapas de favorabilidad y probabilidad de presencia de <i>F. oxysporum</i> : modelo GLM; figura (a) distribución potencial, figura (b) probabilidad de presencia. Modelo GAM: figura (c) distribución potencial, figura (d) probabilidad de presencia. Modelo MAXENT: figura (e) distribución potencial, figura (f) probabilidad de presencia.....	40
Figura 10. Mapas de favorabilidad y probabilidad de presencia de <i>F. oxysporum</i> : modelo RF; figura (a) distribución potencial, figura (b) probabilidad de presencia. Modelo GBM: figura (c) distribución potencial, figura (d) probabilidad de presencia. Modelo BART: figura (e) distribución potencial, figura (f) probabilidad de presencia.	42
Figura 11. Distribución de <i>F. oxysporum</i> en Ecuador Continental: Mapa Binario de	

Presencia-Ausencia generado con el modelo GLM (Figura 11a), modelo GAM (Figura. 11b), modelo MAXENT (Figura. 11c), modelo RF (Figura. 11d), modelo GBM (Figura. 11e), modelo BART (Figura. 11f)	44
Figura 12. Evaluación del rendimiento de seis modelos (GLM, GAM, MX, RF, GBM y BART) a través de métricas con un umbral de prevalencia: Correct Classification Rate (CCR), sensitivity, specificity, precision, recall, True Skill Statistic (TSS) y skappa.	45
Figura 13. Evaluación del rendimiento de seis modelos (GLM, GAM, MX, RF, GBM y BART) a través de métricas con un umbral optimo: Correct Classification Rate (CCR), sensitivity, specificity, precision, recall, True Skill Statistic (TSS) y skappa.	46
Figura 14. Análisis de la eficacia de los modelos (GLM, GAM, MX, RF, GBM y BART) en la predicción de la distribución potencial de <i>F. oxysporum</i> mediante la métrica Área bajo la Curva (AUC).	47
Figura 15. 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 <i>F. oxysporum</i>	48

ÍNDICE DE ANEXOS

Anexo 1. Proceso de Limpieza de Datos en RStudio con el Paquete “scrubr”	67
Anexo 2. Proceso de Cálculo y Trazado en RStudio para el Área Bajo la Curva ROC.....	67
Anexo 3. Proceso de Cálculo y Trazado en RStudio para Métricas de Clasificación Basadas en Umbrales.....	67
Anexo 4. Proceso de Cálculo y Trazado en RStudio para la Línea de Calibración de Miller	68
Anexo 5. Código en RStudio: Conversión de probabilidad a favorabilidad	68
Anexo 6. Código en RStudio: mapear la favorabilidad.....	69
Anexo 7. Generación de mapas binarios presencia-ausencia mediante la herramienta "reclassify".....	69

CÓDIGO DUBLIN

Titulo:	Métodos estadísticos bayesianos aplicados en la modelación de distribución potencial de <i>Fusarium oxysporum</i> en Ecuador continental.				
Autor:	Jacho Saa, Jorge Josué				
Palabras clave:	Métodos estadísticos bayesianos	Modelación de distribución potencial	Evaluación de modelos estadísticos	<i>Fusarium oxysporum</i>	Evaluación de métricas
Fecha de publicación:					
Editorial:	UTEQ				
Resumen:	Resumen.- Se evaluó la eficiencia del modelo bayesiano BART, comparándolo con modelos estadísticos (GLM, GAM) y con algoritmos de aprendizaje automáticos (RF, MAXENT, GBM), para predecir la distribución de <i>Fusarium oxysporum</i> en Ecuador continental. La investigación utilizó datos de presencia de la especie (GIBF) y variables bioclimáticas (WorldClim), para desarrollar los modelos con RStudio a una resolución de 10 x 10 Km², los cuales fueron evaluados con métricas de rendimiento. Los resultados mostraron que RF tuvo el mejor rendimiento en la mayoría de las métricas. Los mapas binarios de presencia-ausencia generados con ArcGis mostraron que GLM tuvo la mayor área de presencia, mientras que RF tuvo la mayor área de ausencia. Por lo tanto, se recomienda continuar con estudios usando modelos bayesianos con mayor registro de presencia de la especie. Abstract.- The efficiency of the Bayesian BART model was evaluated, comparing it with statistical models (GLM, GAM) and with automatic learning algorithms (RF, MAXENT, GBM), to predict the distribution of <i>Fusarium oxysporum</i> in continental Ecuador. The research used species occurrence data (GIBF) and bioclimatic variables (WorldClim) to develop the models with RStudio at a resolution of 10 x 10 km², which were evaluated with performance metrics. The results showed that RF had the best performance in most metrics. Binary presence-absence maps generated with ArcGis showed that GLM had the largest area of presence, while RF had the largest area of absence. Therefore, it is recommended to continue with studies using Bayesian models with higher presence record of the species.				
Descripción:	69 Hojas: dimensiones, 29 x 21 cm + CD-ROM 6162				
URI:					

INTRODUCCIÓN

La presencia de enfermedades en grandes monocultivos está influenciada principalmente por las variaciones climáticas y los avances tecnológicos. Para mantener la producción y evitar pérdidas de rendimiento, estos sistemas a menudo dependen de tecnologías que requieren grandes insumos. Bajo estas condiciones, los mapas de riesgos de enfermedades están bien valorados para prevenir brotes a gran escala o problemas relevantes con los cultivos y amenazas a la seguridad alimentaria [1] [2] [3].

Los requerimientos climáticos como temperatura y precipitación definen la predisposición a enfermedades en regiones específicas donde está presente el cultivo huésped. Los cambios climáticos previstos mundialmente afectarán potencialmente la distribución y manejo de enfermedades en numerosos cultivos [4] [5]. Por lo tanto, hay que tener en cuenta estas complejas interacciones que afectan a los sistemas de patógenos, con el fin de estimar su propagación y los riesgos de enfermedad en diferentes escenarios [3].

Particularmente, el género *Fusarium* está ampliamente distribuido y se encuentra en la mayoría de las regiones climáticas del mundo [6]. Una de las especies más conocidas del género es *Fusarium oxysporum*, la cual probablemente causa mayor daño económico a los cultivos agrícolas que cualquier otro patógeno vegetal [7]. Existen varias razas que son agentes de brote de varios cultivos de alto impacto económico como *F. oxysporum* f. sp. cubense raza 1 en banano o raza 4, que destruyó la industria bananera Cavendish de Taiwán en la década de 1970 [8].

Actualmente, un enfoque importante y preciso para predecir áreas potencialmente favorables para los patógenos, consiste en trabajar con conjuntos de modelos de distribución de especies (MDE) [9], utilizando como variables predictoras métricas de eventos climáticos. Los MDE caracterizan las distribuciones potenciales actuales de las especies, es decir, el espacio climático donde ya se han encontrado estos individuos, y proyectan cambios en la distribución del espacio climático según una región de estudio [10].

Ecuador es un país con una biodiversidad rica, pero también está vulnerable a la invasión de patógenos como el *F. oxysporum*. Por esta razón, se considera importante aplicar modelos estadísticos confiables y robustos para estimar su distribución potencial. En este proyecto de investigación, se evaluó la eficacia de los modelos bayesianos, los cuales han mostrado resultados sobresalientes en predicciones similares. La realización de este estudio es clave para comprender la distribución potencial de *F. oxysporum* en el Ecuador continental y tomar medidas eficaces para prevenir su invasión y controlar su propagación.

CAPÍTULO I
CONTEXTUALIZACIÓN DE LA INVESTIGACIÓN

1.1. Problema de investigación

1.1.1. Planteamiento del problema

Ecuador es un país productor de materias primas, en el cual la economía depende principalmente de la extracción de petróleo y de una gran variedad de cultivos agrícolas, entre los que destaca el banano. *Fusarium oxysporum* es una especie de hongo causante de la enfermedad de Panamá en los cultivos de banano, además es causante de enfermedades en una gran cantidad de especies vegetales. Su nivel de peligrosidad radica en su fácil propagación, incentivada por el mismo ser humano, que, mediante el transporte de mercancías, ha permitido su invasión a distintos continentes.

Su distribución descontrolada ha llegado a los países pertenecientes a la comunidad andina, lo cual ha activado las alarmas por el riesgo potencial de su distribución dentro del territorial nacional del Ecuador. El principal impacto se prevé en el sector agrícola, donde las pérdidas pueden ser incalculables, al reducir una de las principales fuentes de ingresos económicos del país. Por este motivo, en conjunto con los demás países se han tomado medidas para controlar la propagación de este hongo, como el establecimiento de cercos epidemiológicos, desinfección de personas y mercancías por vía marítima como terrestre, y la divulgación de información sobre el hongo, para así evitar la propagación de dicha especie [3].

1.1.1.1. Diagnóstico

Ecuador es uno de los países con mayor biodiversidad del planeta, y el primero en cuanto a cantidad de especies presentes por kilómetro cuadrado. Esto se debe a su ubicación geográfica estratégica, que le atribuye características geológicas, topográficas y climáticas que permiten la presencia de gran cantidad de especies de flora y fauna. Impulsada por la alta fertilidad del suelo, la agricultura tiene una importante presencia en la economía del país, a través de cultivos como el banano, cacao, plátano, café, entre otros. Las condiciones climáticas proporcionan ambientes idóneos para que especies invasoras como *F. oxysporum* se distribuyan dentro del territorio ecuatoriano y afecten

tanto a especies vegetales comunes, como a cultivos de gran importancia como el banano, que resulta ser el más afectado por dicho hongo.

1.1.1.2. Pronóstico

La aparición del patógeno *F. oxysporum* Raza RT4 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. Aunque en Ecuador no se han registrado oficialmente casos de la Raza RT4, sí existen otras clases del patógeno presentes desde 1930. La incertidumbre sobre el momento y lugar en que se registrará el primer brote de *F. oxysporum* en el país representa una amenaza para el cultivo de banano y la economía de un gran número de personas. Es por eso que es fundamental realizar estudios para evaluar la eficacia de diferentes modelos estadísticos en la predicción de la distribución potencial de este patógeno y tomar medidas preventivas.

1.1.2. Formulación del problema

¿Qué tan eficaces son los métodos estadísticos bayesianos para predecir la distribución potencial de *F. oxysporum* en Ecuador continental?

1.1.3. Sistematización del problema

¿Cómo varían las características del modelo de nicho ecológico de *F. oxysporum* según los enfoques de diferentes métodos estadísticos?

¿Cuáles son las áreas más favorables para la distribución potencial de *F. oxysporum* en Ecuador continental?

¿Cuáles son las ventajas de los métodos bayesianos para predecir la distribución potencial de *F. oxysporum*?

¿Cuál o cuáles métodos estadísticos tienen mejor rendimiento en la modelación de la distribución potencial de *F. oxysporum*?

1.2. Objetivos

1.2.1. Objetivo General

Evaluar la eficacia de los métodos estadísticos bayesianos en la modelación de la distribución potencial de *F. oxysporum* en Ecuador continental.

1.2.2. Objetivos Específicos

- Modelar el nicho ecológico de *F. oxysporum* bajo diferentes enfoques estadísticos.
- Caracterizar la distribución potencial de *F. oxysporum* en Ecuador continental según los métodos de modelación empleados.
- Evaluar las métricas de rendimiento de los modelos de *F. oxysporum* según los enfoques estadísticos empleados.

1.3. Justificación

En las últimas dos décadas, se han implementado más de dos docenas de métodos estadísticos y de aprendizaje rápido en la modelación de la distribución de especies. De estos, solo un pequeño grupo han logrado prevalecer debido a su fácil aplicación, velocidad computacional y el fuerte rendimiento predictivo en rigurosas predicciones cruzadas, como el de MaxEnt, uno de los más populares gracias a los excelentes resultados proporcionados [11].

Los árboles de regresión aditiva bayesiana (BART), son una nueva alternativa a otros métodos populares de árboles de clasificación. Como en otros enfoques, BART genera un conjunto de árboles de decisión que explican diferentes componentes de la varianza en la variable de resultado. A diferencia de los bosques aleatorios o los BRT, la formulación de BART es bayesiana, con la probabilidad posterior de un modelo conformado por los priores $P(\text{árboles})$ sobre cómo deberían ser los árboles. Los métodos

de GAM y RF se consideran entre los algoritmos de aprendizaje automático más potentes utilizados para predecir la distribución de especies. Sin embargo, su efectividad para proyectar diferentes niveles de sistemas ecológicos, como la densidad y riqueza de especies, no ha sido completamente confirmada [12] [13].

En la actualidad, ante la gran cantidad de herramientas estadísticas disponibles para predecir la distribución de las especies, se ha hecho necesario evaluar su pertenencia 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 [12]. Por este motivo, en el desarrollo de esta investigación se emplearán seis métodos de modelación de la distribución de especies (GAM, GLM, RF, BRT, BART y MAXENT). De esta forma, se evaluará la eficacia en la predicción de la distribución potencial de *F. oxysporum*, bajo las condiciones bioclimáticas presentes en Ecuador, estableciendo escenarios de áreas potenciales de propagación que sirvan de referencia para futuros proyectos que abarquen temas de regulación de especies invasoras y planes de manejo ambiental aplicados a agroecosistemas del Ecuador.

CAPÍTULO II

FUNDAMENTACIÓN TEÓRICA DE LA INVESTIGACIÓN

2.1. Marco conceptual

2.1.1. Modelo de distribución de especies

Los Modelos de Distribución de Especies (MDE) son una colección de técnicas ampliamente utilizadas para estimar la presencia de especies en un área, basándose en modelos cuantitativos o reglas. Existe una amplia cantidad de material de revisión y libros de texto que proporcionan una descripción detallada de cómo se aplican estos modelos. En la actualidad, los MDE son la herramienta de modelado más empleada para predecir los efectos del cambio climático en la biodiversidad. Este aumento en el interés por los MDE probablemente sea debido a la mayor disponibilidad de datos digitales y a la existencia de software fácil de usar, así como a la presencia de guías, manuales y libros de texto completos [14].

2.1.1.1. Modelos aprendizaje automático

Los Modelos de Aprendizaje Automático funcionan a través de un proceso compuesto por entrenamiento y predicción. Durante el entrenamiento, se suministra al modelo una serie de datos que incluyen valores de entrada y salida previamente conocidos. El modelo entonces utiliza un algoritmo para identificar la relación existente entre ellos [15].

2.1.1.2. Modelos estadísticos

Los Modelos Estadísticos son instrumentos empleados para representar y comprender la correlación entre las variables en un grupo de datos. Estos modelos se fundamentan en teoremas matemáticos y estadísticos y se utilizan para predecir y calcular valores para datos no vistos previamente [15].

2.1.1.3. Modelos Lineales Generalizados (GLM)

Los Modelos Lineales Generalizados o Generalized Linear Models (GLM) brindan la capacidad de especificar una distribución adecuada de errores y una función de asociación para datos de conteo, como la riqueza de especies, utilizando una distribución Poisson o

binomial negativo. Una de las ventajas de los GLM y enfoques similares, como GAM y GLMM, frente a métodos de aprendizaje automático estándar, mínimos cuadrados ordinarios y algunos modelos basados en PLS, es la opción de elegir la familia de modelos adecuada para la tarea de modelado específica

2.1.1.4. Modelos aditivos generalizados (GAM)

Un Modelo Aditivo Generalizado o Generalized Additive Model (GAM) es una herramienta útil para ajustar las relaciones no lineales "flexibles" entre variables. Los GAM son una ampliación de los Modelos Lineales Generalizados (GLM) y ofrecen gran parte de las mismas funciones, incluyendo la posibilidad de modelar datos que no sean normales [12].

2.1.1.5. Bosques Aleatorios (RF)

El modelo de bosques aleatorios o Random Forest (RF) es una técnica de aprendizaje automático que combina varios árboles de clasificación o regresión para prevenir el sobreajuste. Este enfoque combinatorio permite estabilizar la incertidumbre de los árboles individuales y producir predicciones confiables [16].

2.1.1.6. Máquina de Mejora por Gradiente (GBM)

El modelo Máquina de Mejora por Gradiente o Gradient Boosting Machine (GBM) es una técnica de aprendizaje automático de tipo ensamblado que combina varios modelos débiles para crear uno más preciso. Utiliza el proceso de boosting para ir ajustando paulatinamente los pesos de los datos, enfocando la atención en aquellos casos en los que el modelo no ha logrado una clasificación o predicción satisfactoria [17].

2.1.1.7. Árboles de Regresión Aditivos Bayesianos (BART)

Los árboles de regresión aditivos Bayesianos o Bayesian Additive Regression Trees (BART) son un enfoque innovador y novedoso en comparación con otros métodos populares de clasificación de árboles. Al igual que en otros métodos, BART genera un

conjunto de árboles de decisión que explican los diferentes componentes de la variabilidad en la variable de respuesta. A diferencia de los bosques aleatorios o los árboles de regresión generalizados (BRT), la formulación de BART es bayesiana, con una probabilidad posterior de un modelo conformada por prioridades sobre cómo deben verse los árboles (es decir, los parámetros utilizados para generar esos árboles) [15]:

Al igual que los BRT, los BART introducen varianza ajustando un conjunto de árboles poco profundos de "aprendices débiles", pero en lugar de ser controlado por una sola distribución previa como en BRT, es controlado explícitamente por tres distribuciones: la probabilidad de que un árbol se detenga en un nodo con una determinada profundidad, la probabilidad de que se seleccione una determinada variable para una regla de división, y la probabilidad de que se divida esa variable en un valor particular [12]. Estas tres distribuciones son generalmente tratadas como uniformes, mientras que la primera se especifica como una ley de potencia negativa para restringir la profundidad del árbol y evitar el sobreajuste [15].

El proceso de generación de los árboles se lleva a cabo a través de un proceso de MCMC, en el que los árboles se actualizan aleatoriamente para expandirse, reorganizarse o podarse. Cada modelo resultante es una suma de árboles, que se ajusta a los residuos de la suma de los árboles restantes, en contraposición a los bosques aleatorios, que promedian las predicciones entre árboles [11].

2.1.1.8. Máxima Entropía (MaxEnt)

El método de Máximo Entropía o Maximum Entropy (MaxEnt) es una herramienta muy conocida y efectiva en el campo de la Modelación de Distribución de Especies (SDM) que se basa en datos de presencia disponibles en amplia escala. Al crear modelos, MaxEnt considera variables ambientales individuales y brinda clasificaciones estables para la importancia de estas variables a través de sets de datos de validación. Gracias a la función integrada de MaxEnt que evalúa la contribución de cada variable ambiental a la eficacia del modelo, las contribuciones de las variables se reportan de manera estandarizada en numerosos estudios publicados [18].

2.1.2. *Fusarium oxysporum*

El complejo de especies de *F. oxysporum* (FOSC) es responsable de causar enfermedades en cultivos valiosos en todo el mundo, conocidas como marchitez vascular. Aunque la reproducción sexual en el FOSC no se comprende, la transferencia de genes de una fuente a otra puede ayudar a explicar la variedad encontrada en las cepas patógenas [7]. Para que *F. oxysporum* cause una enfermedad en un cultivo vulnerable, se requiere que siga una serie de etapas, comenzando con la producción de esporas y terminando con una infección sistémica bien establecida [13] .

En principio, cada etapa de la infección ofrece una posibilidad para mitigar el riesgo de enfermedad. Esto incluye cambios en la comunidad de microorganismos en el suelo, que pueden afectar la capacidad de las cepas del patógeno para sobrevivir, germinar e infectar las raíces de las plantas. Además, muchos aspectos de la planta huésped, como la composición de los exudados de las raíces, la estructura de la corteza de las raíces y la capacidad de detectar y responder eficientemente a la invasión de un patógeno, pueden impedir el avance de *F. oxysporum* [13].

2.1.2.1. Reproducción

Históricamente, se ha identificado a *F. oxysporum* en base a su aspecto asexual, aunque algunos estudios han sugerido la existencia de un ciclo sexual oculto. Esta teoría ha sido apoyada por investigaciones filogenéticas que ubican a *F. oxysporum* dentro de la familia Gibberella. Estas investigaciones también han revelado una compleja subestructura filogenética en *F. oxysporum*, lo que indica la presencia de múltiples especies ocultas dentro de este género [19].

2.1.2.2. Clasificación de *Fusarium oxysporum*

La categorización de *F. oxysporum* en 66 formas específicas se basa en su semejanza en la morfología tanto de los microcomidios como de los macrocomidios, así como en sus medidas. La clasificación en formas especiales de *F. oxysporum* se ha vuelto común debido a la amplia variabilidad morfológica dentro de esta especie [6].

La clasificación subespecífica de *F. oxysporum* se basa en su capacidad para causar enfermedades en diferentes plantas huésped. Por lo tanto, se considera que las cepas que afectan a la misma planta huésped pertenecen a la misma forma especial. Sin embargo, esta práctica puede llevar a suposiciones erróneas al suponer que una forma especial en particular es filogenéticamente homogénea [19].

2.1.2.3. Variantes de *Fusarium oxysporum*

Varios autores han puesto de manifiesto las relaciones polifiléticas dentro de varias formas especiales de *F. oxysporum*. Por ejemplo, *F. oxysporum* f. sp. *batatas*, *F. oxysporum* f. sp. *cubense* y *F. oxysporum* f. sp. *vasinfectum*. Además, varias formas especiales de *F. oxysporum* son capaces de infectar y causar la enfermedad en más de una planta (a veces no relacionada), mientras que otras están muy especializadas en una planta específica [19].

Algunas de las variantes más comunes de *F. oxysporum* incluyen: *Fusarium oxysporum* f. sp. *lycopersici*: Esta forma patogénica afecta a los tomates y es responsable de causar la enfermedad del pie negro. *Fusarium oxysporum* f. sp. *vasinfectum*: Esta forma patogénica afecta a las alfalfas y es responsable de causar la enfermedad del collar amarillo. *Fusarium oxysporum* f. sp. *melonis*: Esta forma patogénica afecta a los melones y es responsable de causar la enfermedad de la raíz de los melones. *Fusarium oxysporum* f. sp. *cubense*: Esta forma patogénica afecta a la caña de azúcar y es responsable de causar la enfermedad del mal del topo [6].

.

2.1.3. Nicho ecológico

El término "nicho ecológico" es crucial para entender la ecología y evolución de las especies, aunque su significado y alcance sigue siendo objeto de discusión. La primera definición del nicho ecológico fue propuesta por G. Evelyn Hutchinson en 1957 y la describió como un "espacio multidimensional" en el que cada punto representa un estado ambiental que permite que una especie sobreviva. Esta definición ha sido muy influyente y ha motivado a los científicos a utilizar modelos de nicho ecológico o de distribución de especies (SDM), que describen el espacio de nicho en términos de la presencia de especies y las variables ambientales relevantes [20].

2.1.3.1.Nicho Fundamental

La capacidad de una especie para vivir en un ambiente específico sin la interferencia de otros organismos se conoce como su nicho fundamental [21]. Este nicho se determina por las condiciones ambientales que una especie puede soportar [21]. Las áreas del planeta que tienen estas condiciones ambientales ideales podrían ser el hogar de esta especie. Sin embargo, la presencia de competidores y depredadores, así como las limitaciones de dispersión de la especie, restringen su rango geográfico ocupado a solo una porción de su nicho fundamental disponible [21].

2.1.3.2.Nicho realizado

La definición de "nicho realizado" de Hutchinson (1957) [22] se centra en la situación actual de una especie en términos de su distribución y presencia en un ecosistema específico, considerando tanto las interacciones positivas como negativas con otras especies presentes en el mismo ecosistema. Este enfoque brinda una comprensión más profunda de cómo las relaciones con otras especies impactan en la presencia y sobrevivencia de una especie en un ecosistema determinado [23].

2.1.3.3.Diagrama BAM

El diagrama BAM (mostrado en la Figura 1) es un gráfico que permite visualizar la ubicación de las especies en relación a factores abióticos (A; como la temperatura y la precipitación), factores bióticos (B; como las interacciones con otros seres vivos), y la accesibilidad a diferentes áreas a través de la dispersión (M). Este diagrama es una herramienta útil para representar la distribución geográfica de las especies [24].

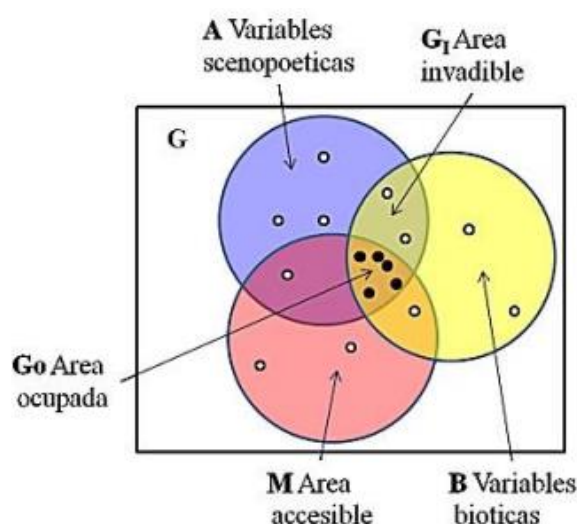


Figura 1. Diagrama BAM. Fuente: Sobaron (2017).

2.2. Marco referencial

2.2.1. Participación de Organizaciones Internacionales

CGIAR es una asociación de investigación global y la red de investigación agrícola más grande del mundo. Colabora con más de 1000 socios en todo el mundo a través de 15 programas de investigación temáticos [25]. El Programa de Investigación en Raíces, Tubérculos y Banano (RTB) es uno de ellos y está compuesto por 5 centros de investigación: Bioversity International, CIAT, IITA y CIRAD [26]. Trabajan con más de 360 socios. RTB comenzó en 2012 y solicitó una evaluación prioritaria, llevando a cabo un estudio riguroso y armonizado cuantitativo para los 5 cultivos principales: plátano, yuca, papa, camote y ñame. [25].

Se han hecho llamados frecuentes para aumentar la inversión global para reducir el impacto de FWB y la Organización de las Naciones Unidas para la Agricultura y la Alimentación lanzó recientemente una iniciativa global que busca donantes para invertir US \$ 98 millones para una respuesta concertada [25]. Sin embargo, si bien existe cierto consenso de que la amenaza de Foc TR4 es real y debe abordarse, acordar una estrategia de inversión global es un desafío. A menudo, el colapso de la producción bananera se presenta como inminente a través de palabras como extinción del banano, apocalipsis y el fin del banano [25].

La IPPC está a la vanguardia en la lucha contra *Fusarium* TR4, y un equipo recientemente formado sobre *Fusarium* TR4, reúne a expertos cualificados de todo el mundo. En 2021, el equipo comenzó a preparar actividades de sensibilización sobre la vigilancia, el diagnóstico y las directrices para aplicar medidas fitosanitarias, incluida la vigilancia previa y posterior a la detección (para delimitar eventuales brotes de TR4). Ejercicios de simulación sobre la vigilancia y los brotes de plagas se están considerando para fortalecer las capacidades nacionales como una actividad esencial para los acuerdos de preparación [27].

2.2.2. Impacto de FW en la agricultura

La propagación de la enfermedad en China ha resultado en la pérdida de más de 100,000 hectáreas de tierra de producción en los últimos dos o tres décadas [16]. Esto ha sucedido en China y otros cuatro países en la subregión del Mekong, debido a que los productores de Cavendish por contrato buscan tierras sin contaminación. Sin embargo, la magnitud de la amenaza para los sistemas de pequeños agricultores con una amplia diversidad de cultivos y prácticas agrícolas es incierta [16].

El algodón es un cultivo vital desde un punto de vista económico, ya que es la fibra más importante y un cultivo importante de semillas oleaginosas [28]. Se cultiva en más de 80 países, ocupando el 2,5% de la superficie de tierra cultivable mundial y generando un valor de producción de 50 mil millones de dólares estadounidenses [28]. Sin embargo, esta producción se ve amenazada por la Marchitez por *Fusarium*, causada por un patógeno fúngico llamado *Fusarium oxysporum* f. sp. *vasinfectum* (FOV), transmitido a través del suelo y las semillas [28].

La enfermedad ha causado pérdidas significativas de rendimiento en muchos países productores de algodón. Por ejemplo, en los Estados Unidos, las pérdidas económicas debido a FOV oscilaron entre 0,19 y 1,36%. Además, en Darling Downs, Australia, FOV causó pérdidas económicas de algodón de 57 millones de dólares australianos durante la temporada de cultivo 2009-2010 [28].

2.2.3. *Fusarium oxysporum* f. sp. *cubense* Tropical Race 4 (TR4) en el mundo

Fusarium oxysporum f. sp. *cubense* Tropical Race 4 (TR4) ha sido una enfermedad restringida en Asia desde 1973 [26]. Sin embargo, en los últimos diez años, TR4 ha cruzado fronteras e invadido África en 2013 y América del Sur en 2019, representando una amenaza para países donde TR4 aún no ha llegado y donde los cultivos de plátano tienen una importancia fundamental en términos sociales y económicos [26]. El objetivo es prevenir la expansión de TR4 a los países vecinos, y se deben tomar medidas preventivas para mantener una vigilancia adecuada. [27].

2.2.4. Presencia de *Fusarium oxysporum* en Ecuador

La presencia de Fusariosis en el cultivo de banano en Ecuador fue registrada por primera vez en la provincia de Guayas en 1936 [29]. Se cree que la enfermedad se propagó a otras partes del país a través del suelo o material infectado. De acuerdo con investigaciones realizadas por ESPOL, las cepas de Foc en Ecuador provendrían de un solo linaje clonal, asociado al Grupo de Compatibilidad Vegetativa (GCV) 0120 y categorizado como la raza 1 [29].

2.2.5. Hitos relevantes en el desarrollo de modelos de distribución de especies

Guisan y Thuiller [30] indican que uno de los primeros intentos de modelado de distribución de especies encontrado en la literatura son las predicciones basadas en nichos de especies de cultivos realizadas por Henry Nix y sus colaboradores [31].

El modelado de distribución de especies ha sido un área de investigación activa desde la década de 1970. Aquí hay una línea de tiempo de los eventos importantes en la historia del modelado de distribución de especies: 1970-1980: Se desarrollaron los primeros modelos de distribución de especies basados en factores climáticos y geográficos [10]. 1980-1990: Se comenzaron a utilizar modelos basados en la biología de las especies para predecir la distribución [10]. 1990-2000: Los modelos de distribución de especies se incorporaron en sistemas de información geográfica (SIG) y se utilizaron para identificar áreas de conservación importantes. 2000-2010: Se desarrollaron nuevos modelos basados

en la biología de las especies y en la información genética, incluyendo la utilización de modelos basados en árboles filogenéticos [10]. 2010-actualidad: Se han desarrollado nuevas técnicas de modelado de distribución de especies, incluyendo la incorporación de datos de presencia y ausencia, la utilización de modelos basados en Bayes y la incorporación de información sobre el uso del hábitat [10].

2.2.6. Problemas de coincidencia entre datos

Los Modelos de Distribución de Especies (MDE) pueden ser contruidos con diferentes tipos de datos de especies, y estos tienen un impacto importante en la cantidad que se estima [30]. Por lo tanto, los MDE están particularmente sujetos a problemas debido a la falta de correspondencia entre los datos y el objetivo previsto [30]. Algunas aplicaciones de MDE requieren conocer la probabilidad de que una especie esté en un lugar o la extensión total que ocupa [30]. Otros solo requieren conocer la aptitud relativa del lugar para identificar los mejores y peores. Aunque existen estudios recientes que abordaron esta variación en la información necesaria, todavía falta una evaluación integral de cómo los diferentes procesos de muestreo, tipos de datos y enfoques de modelado afectan la utilidad de los MDE en todas las aplicaciones [32].

2.2.6.1. La interacción entre tipos de datos y sesgos en SDM

Hay tres aspectos importantes a tener en cuenta cuando se utiliza información sobre especies para construir un MDE: (1) si se puede calcular la frecuencia de ocurrencia de una especie; (2) el impacto de una detección imprecisa; y (3) el efecto de un sesgo en la muestra ambiental (Figura 2) [30]. Estos factores, que se discutirán con más detalle a continuación, dictan los tipos de aplicaciones que un MDE puede ofrecer de manera confiable, ya que determinan si el modelo resultante es capaz de: (1) estimar la probabilidad de que una especie se encuentre en un lugar; (2) proporcionar una estimación relativa de la probabilidad de ocurrencia de una especie, que es proporcional a la probabilidad real; (3) clasificar correctamente los lugares en términos de probabilidad de ocurrencia de una especie; o (4) brindar una estimación sesgada de la presencia de especies en la que ni siquiera se logra clasificar correctamente los lugares en términos de la probabilidad real de ocurrencia [32].

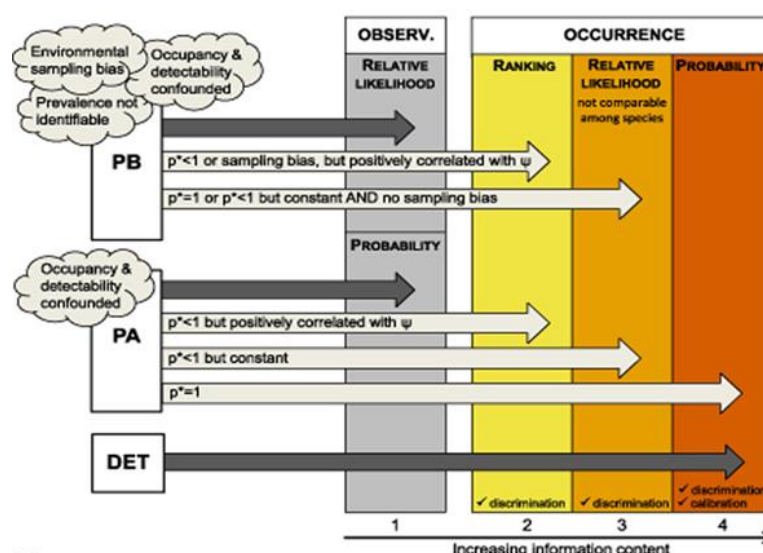


Figura 2. Interacción entre tipos de datos y sesgos en MDE. Elaborado: Guillera (2020).

2.2.7. Construcción de un SDM

Durante el proceso de construcción de un SDM, hay una serie de decisiones metodológicas que deben ser tomadas y estas pueden introducir incertidumbre en el modelo y sus predicciones [33]. Es importante comprender estas fuentes de incertidumbre para poder evaluar la confiabilidad de las predicciones [33]. Algunos de estos factores se pueden controlar parcialmente (por ejemplo, el tamaño de la muestra, el diseño de muestreo, la técnica de modelado), pero otros son más difíciles de manejar (por ejemplo, los patrones de distribución y dispersión de especies) [33]. La mayoría de los estudios previos han evaluado los efectos de estos factores de manera individual, lo que impide una evaluación de sus efectos relativos. La mayoría de estos estudios han utilizado datos reales, pero algunos han utilizado datos simulados o virtuales [33].

2.2.7.1. Datos reales y datos virtuales

Los estudios que utilizan datos reales en su mayoría han procedido degradando las condiciones iniciales individualmente, por ejemplo, agregando autocorrelación espacial, engrosando el grano o disminuyendo el tamaño de la muestra, y luego midiendo el aumento resultante en el error de predicción y la disminución en el poder predictivo. Incluso si se ha aprendido mucho de ellos, estos análisis de un solo factor no pueden permitirnos medir los efectos relativos de diferentes factores.

Otros estudios han evaluado cómo la variación en los factores puede afectar las predicciones, pero no son informativos sobre los tamaños del efecto relativo porque se desconoce la verdad y solo se puede cuantificar la incertidumbre. Evaluar los efectos relativos de varios factores en la construcción de modelos requiere que se conozca la verdad. El uso de un enfoque de 'ecólogo virtual' con datos de distribución de especies virtuales simulados y observadores virtuales ofrece una perspectiva de prueba de la verdad y permite un control total de cómo se combinan los factores [33].

2.2.8. Modelos estadísticos bayesianos

Los árboles de regresión aditiva bayesianos combinan la familiaridad conceptual y las fortalezas de los métodos de clasificación de árboles con un componente bayesiano sencillo que gestiona de manera intuitiva la incertidumbre en el modelo [11]. Esto los convierte en una opción atractiva, no solo frente a los enfoques bayesianos existentes, sino también frente a los métodos populares de clasificación de árboles, como los BRT [11]. BRT cuenta con varias implementaciones disponibles y eficaces para la inferencia ecológica, y ha demostrado un rendimiento consistente en pruebas rigurosas de modelos de distribución de especies [15]. No obstante, también presenta algunas debilidades, como la tendencia a sobreajustarse y la falta de transparencia en los procedimientos de ajuste [15]. Muchos estudios simplemente siguen prácticas anecdóticas y eligen los hiperparámetros basándose en los valores predeterminados del software, en lugar de seleccionarlos mediante validación cruzada formal, tal como se recomienda en la literatura temprana [15]

2.2.8.1. Fortalezas y debilidades de los modelos estadísticos bayesianos

Las fortalezas de los modelos estadísticos bayesianos aplicados a la modelación de la distribución de especies incluyen: Flexibilidad: Los modelos bayesianos pueden modelar una amplia variedad de distribuciones de especies, incluyendo aquellas con patrones de distribución complejos o heterogéneos [11]. Incorporación de incertidumbre: Los modelos bayesianos permiten la incorporación de la incertidumbre en la estimación de la distribución de especies, lo que resulta en una evaluación más precisa de la probabilidad de presencia de una especie en una determinada área [11]. Inclusión de datos previos: Los modelos bayesianos permiten la incorporación de información previa, como datos históricos o datos de otros estudios, lo que puede ser útil en la modelación de especies

raras o en peligro de extinción. Interpretabilidad: Los modelos bayesianos son fáciles de interpretar y permiten la obtención de información adicional, como la probabilidad de que una especie esté presente en una determinada área [15].

Sin embargo, también existen algunas debilidades en el uso de modelos bayesianos para modelar la distribución de especies, incluyendo: Complejidad: Los modelos bayesianos pueden ser computacionalmente costosos y requieren una comprensión sólida de estadística y programación para su implementación y análisis [11] . Supuestos implícitos: Al igual que cualquier modelo estadístico, los modelos bayesianos están basados en ciertos supuestos, como la normalidad de los datos o la independencia de las observaciones [11]. Si estos supuestos no se cumplen, los resultados del modelo pueden ser incorrectos. Requisito de datos: Los modelos bayesianos requieren una cantidad significativa de datos para producir estimaciones precisas, lo que puede ser un desafío en el caso de especies raras o de baja densidad [15].

CAPÍTULO III
MÉTODOLOGÍA DE LA INVESTIGACIÓN

3.1. Localización

La República de Ecuador está ubicada en la costa noroeste de América del Sur ($01^{\circ}30' N$ y $03^{\circ}23,5'$ paralelos, $75^{\circ}12' W$ y $81^{\circ}00' W$ meridianos), limitando al norte con Colombia, al sur y este con Perú, y el Océano Pacífico al oeste (Figura 3). 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: Páramo (tundra alpina), manglares, bosque nuboso y selva tropical; incluyendo tres biomas: la Costa, los Andes y la Amazonía [34].

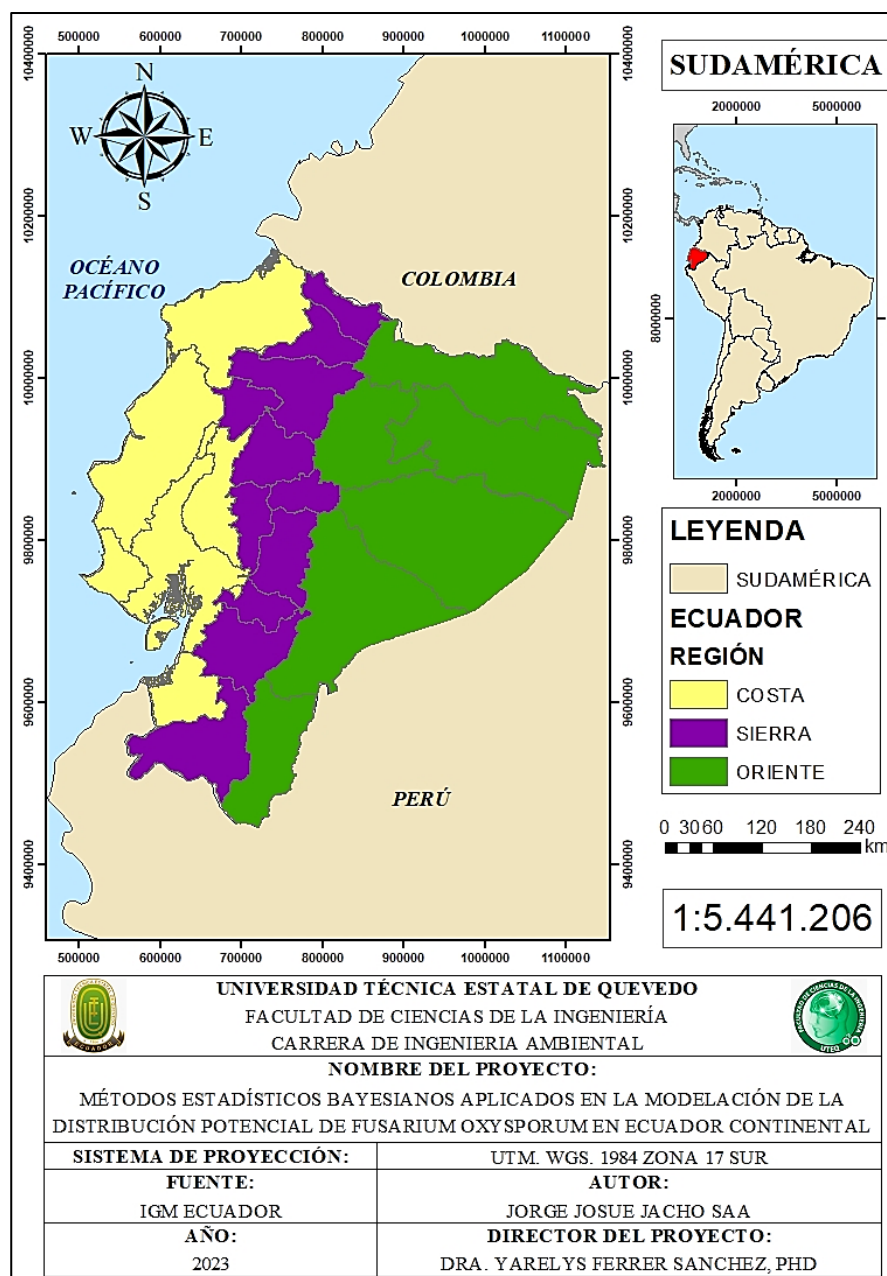


Figura 3. Mapa de Ecuador continental.

3.2. Tipo de investigación

3.2.1. Diagnóstica

El objetivo de la investigación es evaluar la eficiencia de MDE (Generalized Linear Models, Modelos Aditivos Generalizados, Bosques Aleatorios, Arboles de Regresión Potenciados, Arboles de Regresión Aditivos Bayesianos) frente al de Máxima Entropía (Maxent), para estimar la distribución potencial de *F. oxysporum* en Ecuador continental.

3.2.2. Explicativa

Esta investigación tiene un enfoque explicativo y busca determinar si el hongo *F. oxysporum* se adapta a la zona continental de Ecuador a través de la aplicación de métodos de Modelado de Distribución de Especies. Esto es relevante debido a la experiencia en otros países, donde la introducción de esta especie ha resultado en graves consecuencias para el sector agrícola. Si se confirma la adaptación de *F. oxysporum* en Ecuador, podría tener consecuencias similares a las observadas en otros países.

3.3. Métodos de investigación

3.3.1. Método inductivo

La investigación se enfocará en el uso del método inductivo a través de la aplicación de modelos de distribución de especies para determinar el nicho ecológico de la especie *F. oxysporum* en el Ecuador continental. Se utilizarán los registros individuales de presencia de la especie para lograr este objetivo y entender su potencial adaptación a la zona.

3.3.2. Método analítico

El método analítico es utilizado para el estudio de fenómenos, mediante un análisis de todos sus componentes, por esta razón, se empleó este método en la presente investigación, ya que, se evaluará el nicho ecológico de la especie *F. oxysporum*, para determinar las afectaciones de la distribución potencial en el área de estudio.

3.4. Fuentes de recopilación de información

Para el proyecto de investigación la información fue obtenida por fuentes secundarias como son:

- Literatura científica
- Revistas científicas
- Base de datos con registro de especies a nivel internacional
- Base de datos meteorológicos y climáticos globales de alta resolución espacial

3.5. Diseño de investigación

Para el presente proyecto de investigación, en la figura 4 se representa el proceso metodológico realizado para modelar la distribución potencial de *F. oxysporum* en Ecuador continental.

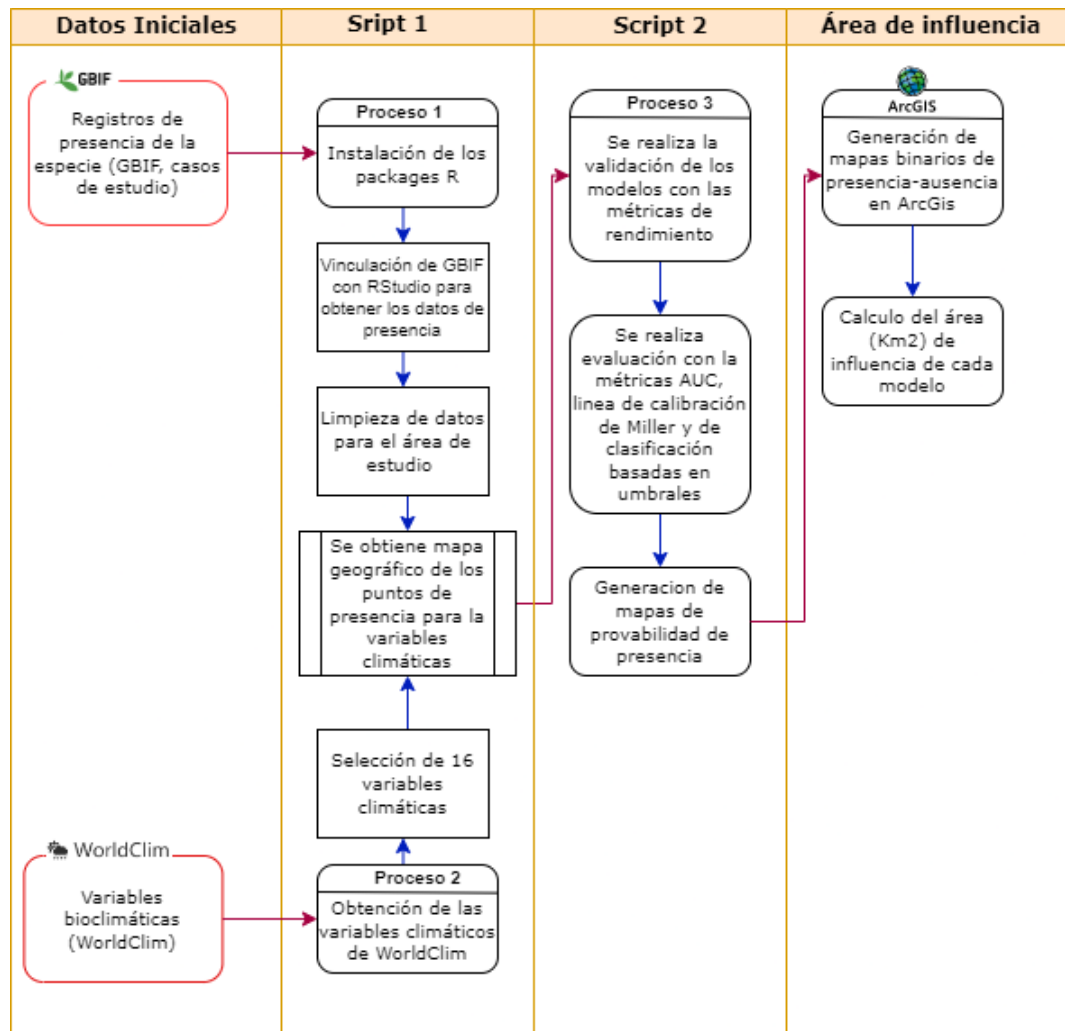


Figura 4. Metodología en acción: Mapa de Flujo del proyecto de investigación. Elaborado: Autor.

3.5.1. Objetivo 1: Modelar el nicho ecológico de *F. oxysporum* bajo diferentes enfoques estadísticos

3.5.1.1. Fuentes de obtención para los datos de presencia de la especie

Se accedió a los datos de presencia de la especie *F. oxysporum* a través de la base de datos Global Biodiversity Information Facility (GBIF) [35]. GBIF es una red internacional financiada por varios gobiernos que brinda acceso libre a información sobre todos los tipos de vida en la Tierra. Se complementó la base de datos generada a partir de los puntos de presencia obtenidos en GBIF con información de casos de estudios de la especie en Ecuador, proveniente de la literatura científica de diferentes instituciones gubernamentales [36].

3.5.1.2. Obtención de variables bioclimáticas

Las variables climáticas se obtuvieron de la base de datos WorldClim (WorldClim) [37], la cual reúne datos meteorológicos y climáticos globales de alta resolución espacial. En su versión más actual, consta de un conjunto de 19 variables (tabla 1) basadas en datos climáticos mensuales interpolados espacialmente con una resolución espacial de 2.5 arc-segundos [38].

Tabla 1. Datos Bioclimáticos en WorldClim: variables disponibles.

SIGLAS	VARIABLES
BIO1	Temperatura media anual
BIO2	Rango diurno medio (Media de la temperatura mensual (temperatura-temperatura mínima))
BIO3	Isotermalidad (BIO2/BIO7)
BIO4	Estacionalidad de la temperatura (desviación estándar x 100)
BIO5	Temperatura máxima del mes más cálido
BIO6	Temperatura mínima del mes más frío
BIO7	Rango de temperatura anual (BIO5-BIO6)
BIO8	Temperatura media del cuarto más

	húmedo
BIO9	Temperatura media del trimestre más seco
BIO10	Temperatura media del cuatrimestre más cálido
BIO11	Temperatura media del cuatrimestre más frío
BIO12	Precipitación anual
BIO13	Precipitación del mes más lluvioso
BIO14	Precipitación del mes más seco
BIO15	Estacionalidad de la precipitación (coeficiente de variación)
BIO16	Precipitación del cuatrimestre más húmedo más húmedo
BIO17	Precipitación del trimestre más seco
BIO18	Precipitación del cuatrimestre más cálido
BIO19	Precipitación del cuatrimestre más frío

3.5.1.3. Procesamiento de datos en RStudio

Para el desarrollo de la investigación se usó RStudio, el cual es un entorno de desarrollo integrado (IDE) para el lenguaje de programación R. Se descargaron varios paquetes (packages: terra, raster, rgbif, scrubr, sdmpredictors, fuzzysim, gam, maxnet, randomForest, gbm, modeva, automap, sf), necesarios para cumplir distintas funciones durante el proceso de elaboración de los modelos de distribución.

En RStudio, se procedió a crear un nuevo script, donde se usó la línea de código "setwd(dirname(rstudioapi::getActiveDocumentContext())\$path)" que establece el directorio de trabajo que contiene el script actual. Se cargaron los paquetes previamente instalados desde la librería mediante la función "library (nombre del paquete)" y así para cada uno (rgbif, terra, geodata, scrubr, sdmpredictors, fuzzysim). Además, se creó la carpeta para archivos de salida con el código "dir.create("../nombre")".

Para obtener los datos de presencia de la especie, se empleó el código “myspecies <- "Fusarium oxysporum" el cual permite asignar una cadena de caracteres en el lenguaje de programación en RStudio. Y se usa la línea de código “gbif_data <- occ_data(scientificName = myspecies, hasCoordinate = TRUE, limit = 5000)”, para obtener los datos de la base de datos de GBIF, donde se estableció un límite máximo de 5000 registros de presencia para reducir el tiempo del proceso. Para obtener un mapa de presencia mundial de la especie (Figura 5).

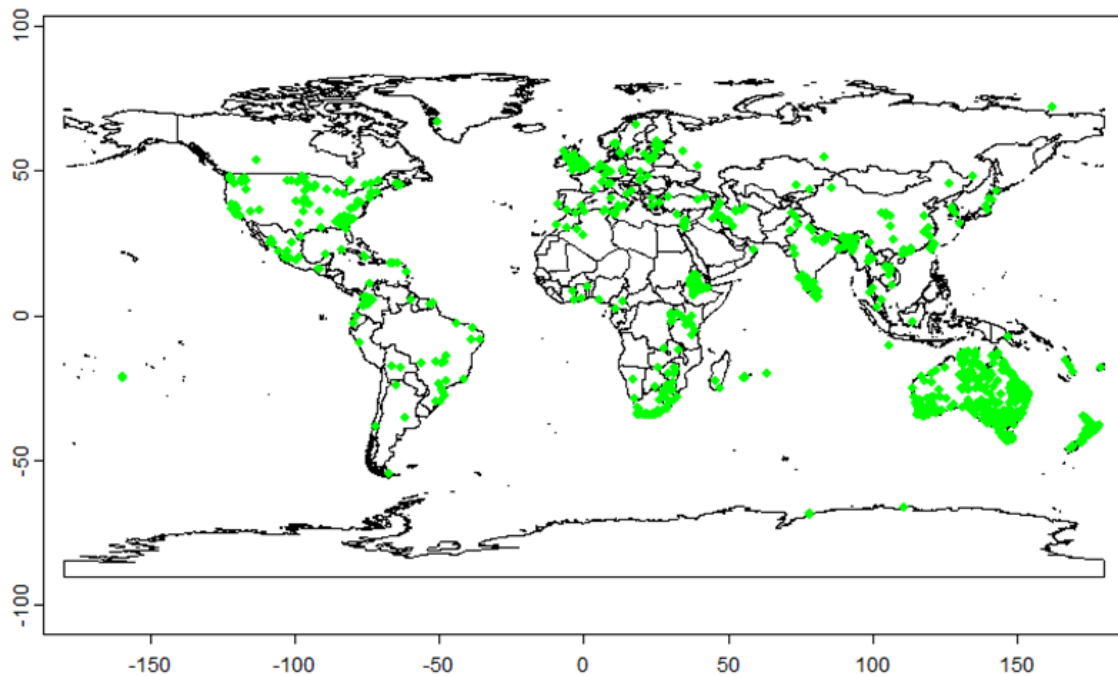


Figura 5. Distribución global de *F. oxysporum*: registros de presencia.

Dada a la capacidad de la especie de contar con un amplio rango de distribución, se establecieron límites del área para obtener los registros de presencia, los cuales corresponden a los países desde Colombia, Ecuador y hasta Perú. Mediante la línea de código “mywindow <- ext(c(-81, -68, -19, 12))”, donde los límites de la ventana se extienden desde los -81 grados de longitud hasta los -68 grados de longitud (en el eje x), y desde -19 grados de latitud hasta 12 grados de latitud (en el eje y).

Para reducir los registros de presencia para el área de estudio se empleó el código “plot(countries, xlim = range(gbif_data\$data\$decimalLongitude), ylim = range(gbif_data\$data\$decimalLatitude), border = "darkgrey)”. Debido a que los datos en línea alojados en la base de datos GBIF pueden contener errores, es importante una

limpieza de los datos. Primero se hizo una eliminación de los registros de ausencia y para ser aún más riguroso con la limpieza de datos se usó la función del paquete “scrubr”, que permite marcar coordenadas incompletas como imposibles para así limpiar los datos y validarlos. Después para tratar la incertidumbre en las coordenadas geográficas, se usó la función "presences <- coord_uncertain(presences, coorduncertaintyLimit = 7071)" donde el valor 7071 representa la distancia del centroide a la esquina de una celda de 10x10 km². De esta manera, se obtuvo el mapa con los registros de presencia del área seleccionada (Figura 6).

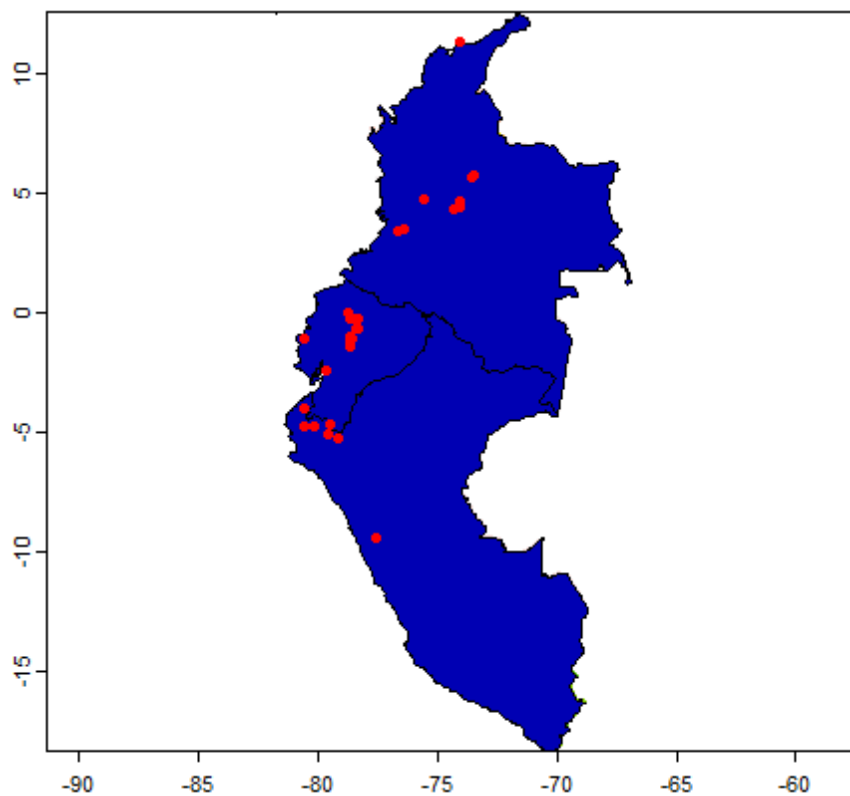


Figura 6. Presencia de *F. oxysporum* en Ecuador, Perú y Colombia: Registros Gráficos.

En el estudio realizado se seleccionaron variables bioclimáticas que se consideraron relevantes para modelar la distribución de la especie *F. oxysporum*. Se decidió inicialmente incluir la variable de altitud porque esta es una variable que puede influir en la distribución de la especie. Posteriormente, se decidió considerar solo las variables bioclimáticas desde la bio 1 hasta la 15, sin incluir las variables bioclimáticas a partir de la bio 16, porque la inclusión de estas no proporcionaría información adicional relevante para el modelado de la distribución de la especie, debido que son índices que se calculan a partir de las primeras 15 variables.

Para seleccionar las variables bioclimáticas más relevantes para modelar la distribución de la especie, se utilizó la función de GLM en Rstudio. La función de GLM se utiliza para modelar relaciones lineales entre variables, y en este caso se utilizó para identificar las variables bioclimáticas más importantes para la distribución de la especie.

Las variables bioclimáticas bio 3, 5, 7 y 12 fueron seleccionadas como las más relevantes para modelar la distribución de la especie *F. oxysporum*. La variable bio 3 representa la precipitación total anual, la variable bio 5 representa la precipitación mensual, la variable bio 7 representa la temperatura promedio mínima mensual y la variable bio 12 representa la temperatura promedio máxima mensual. Estas variables son importantes para modelar la distribución de la especie debido a su capacidad de influir en los factores ecológicos que pueden afectar la distribución de la especie. Posteriormente, se procedió a obtener las variables climáticas recortadas para el área de estudio seleccionada (Figura. 7).

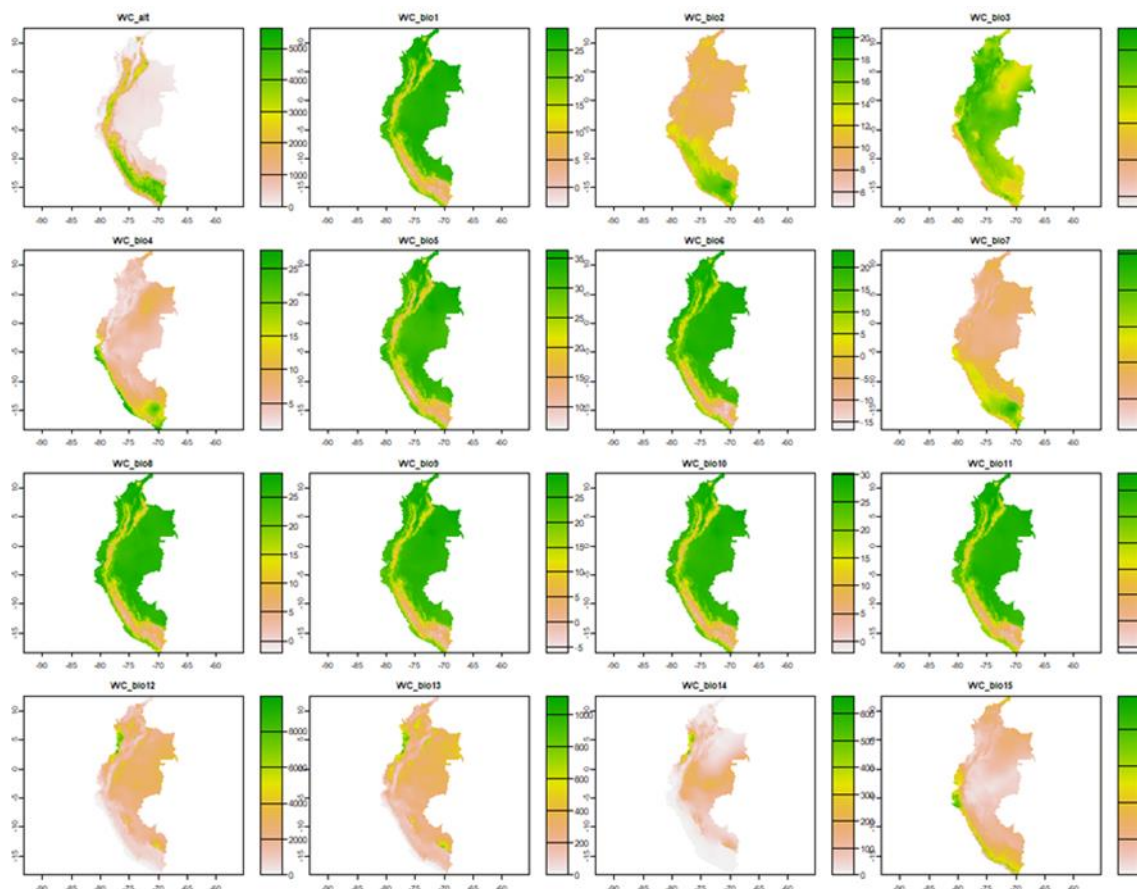


Figura 7. Mapas de variables bioclimáticas en Ecuador, Perú y Colombia: altitud y bio 1 a bio 15

Después de tener los datos de presencia de la especie *F. oxysporum* y las variables bioclimáticas que se determinaron como las más relevantes, se procedió a aplicar diversos modelos de distribución de especies en Rstudio. Estos modelos incluyen GLM (Generalized Linear Model), GAM (Generalized Additive Model), MAXENT (Maximum Entropy), RF (Random Forest), GBM (Gradient Boosting Machine) y BART (Bayesian Additive Regression Trees). Estos modelos permiten tener una idea de cómo la especie se distribuye en el ambiente, basándose en los datos de presencia y en las variables bioclimáticas seleccionadas. Al utilizar diferentes métodos, se pueden comparar los resultados y determinar cuál es el modelo más apropiado para predecir la distribución de la especie en cuestión.

3.5.2. Objetivo 2: Caracterizar la distribución potencial de *F. oxysporum* en Ecuador continental según los métodos de modelación empleados

Se utilizaron los seis modelos de distribución potencial de la especie que se obtuvieron en Rstudio (GLM, GAM, MAXENT, RF, GBM y BART). Estos mapas, una vez generados, fueron importados al software ArcGIS. En ArcGIS, se utilizó la herramienta "classify", la cual permite reclasificar los rangos de valor que presentan los polígonos. Se redujo la cantidad de rangos a solo dos, para obtener los mapas binarios presencia-ausencia, usando el valor del umbral obtenido para cada modelo. Una vez obtenidos los modelos en formato de mapas binarios, se procedió a multiplicar el número de polígonos de presencia como ausencia con el valor correspondiente a la resolución espacial seleccionada (10x10 km).

3.5.3. Objetivo 3: Evaluar las métricas de rendimiento de los modelos de *F. oxysporum* según los enfoques estadísticos empleados

Se utilizó RStudio para calcular diferentes métricas de rendimiento de los modelos generados. En primer lugar, se calculó la AUC (Área bajo la curva ROC) como una métrica general de rendimiento del modelo. La AUC mide la habilidad del modelo para distinguir entre las áreas donde la especie está presente y las que están ausentes. A continuación, se utilizó la línea de calibración de Miller para evaluar el desempeño de los

modelos en términos de su capacidad para predecir la presencia o ausencia de la especie. Posteriormente, se calcularon métricas de clasificación basadas en umbrales, tales como CCR (Correct Classification Rate), Sensitivity, Specificity, Precision, Recall, TSS (True Skill Statistic) 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. Por último, se utilizaron métricas de validación con umbral óptimo, incluyendo CCR, Sensitivity, Specificity, Precision, Recall, TSS y Kappa, para evaluar la precisión de los modelos bajo diferentes escenarios. Con estas métricas, se pudo determinar el desempeño óptimo de cada modelo y comparar los resultados entre ellos.

3.6. Instrumentos de investigación

- La herramienta RStudio v4.2.2 fue utilizada para el procesamiento de los datos de presencia de la especie y las variables climáticas, con el fin de generar los modelos de distribución de la especie.
- Se emplearon diversos paquetes de R para procesar y obtener los registros de presencia, incluyendo terra, raster, rgeos, rgdal, rgeos, sdm, fuzzy, gam, maxnet, randomForest, gbm, ModEVA, automap y sf.
- Se utilizó ArcGIS 10.8 para identificar las áreas favorables en el país y calcular la distancia en kilómetros cuadrados que presentan condiciones ideales para la presencia del patógeno

3.7. Tratamiento de los datos

Para el análisis estadístico, se evaluaron los modelos de distribución de especies mediante el software RStudio, mediante el uso de varias métricas de rendimiento, que se pueden utilizar para medir su efectividad. El AUC permitió medir la capacidad de los modelos en clasificar correctamente un objeto como perteneciente o no a una determinada especie. La línea de calibración de Miller evaluó la capacidad del modelo de predecir la presencia o ausencia de una especie en una ubicación dada. Por su parte, con el STT se analizó la

relación entre la sensibilidad y la especificidad del modelo, mientras que la CCR midió la precisión general del modelo en la clasificación de presencia/ausencia de la especie.

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.

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. En resumen, la combinación de diferentes métricas de rendimiento permitió evaluar de manera más completa la efectividad de los modelos de distribución de especies y seleccionar el mejor para una aplicación determinada.

3.8. Recursos humanos y materiales

3.8.1. Recursos humanos

3.8.1.1. Esquipo de dirección y coordinación

- Dra. Yarelys Ferrer Sanchez, PhD. Directora de Tesis
- Dr. Alexis Herminio Plasencia Vázquez, PhD. Codirector de Tesis

3.8.2. Materiales

Tabla 2. Recursos empleados en el proyecto de investigación.

Material bibliográfico	Softwares especializados	Registro de especies	Paquetes R
Artículos científicos	ArcGis v10.8	Global Biodiversity Information Facility (GBIF)	terra raster rgbif scrubr sdmpredictors fuzzysim gam maxnet randomForest gbm ModEVA automap sf
Registros de presencia (<i>F. oxysporum</i>)	RStudio R v4.2.2 Microsoft office		

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

4.1. Modelación del nicho ecológico de *Fusarium oxysporum* bajo diferentes enfoques estadísticos

4.1.1. Estadísticos descriptivos de las variables bioclimáticas

A través de los estadísticos descriptivos de las variables ambientales (tabla 3), se demuestra la alta capacidad de adaptación que posee la especie para poder distribuirse en distintas condiciones bioclimáticas en el espacio geográfico.

Tabla 3. Estadísticos descriptivos de las variables bioclimáticas que caracterizan el área de donde se extrajeron los registros de presencia de *F. oxysporum* (N=183).

Variables	Media	Desviación estándar	Mínimo	Máximo	Coefficiente de Variación
altitud	1630,20	1193,30	3,00	3998,00	73,19
bio1	17,99	5,47	5,10	27,00	30,42
bio2	11,64	2,54	6,50	17,10	21,84
bio3	0,736	0,10	0,46	0,93	14,21
bio4	11,98	10,63	1,57	52,51	88,77
bio5	25,89	6,06	10,40	37,00	23,41
bio6	9,64	5,75	-0,80	22,80	59,70
bio7	16,25	4,91	9,20	28,80	30,26
bio8	18,80	6,04	3,70	30,00	32,15
bio9	17,07	5,33	4,30	27,60	31,20
bio10	19,33	5,98	5,40	30,00	30,96
bio11	16,33	5,13	3,30	26,50	31,44
bio12	1131,89	528,60	139,00	3782,00	46,70
bio13	188,97	84,69	69,00	580,00	44,81
bio14	28,47	28,11	0,00	135,00	98,74
bio15	62,94	32,36	10,00	187,00	51,41

4.1.2. Variables climáticas con mayor influencia

Influencia de las variables bioclimáticas en la selección por GLM: la precipitación mensual (bio 5) es la más influyente con un aporte superior a 30. La temperatura promedio mínima mensual (bio 7) tuvo el menor aporte con un valor de 14. Revisar Figura 8 para más detalles.

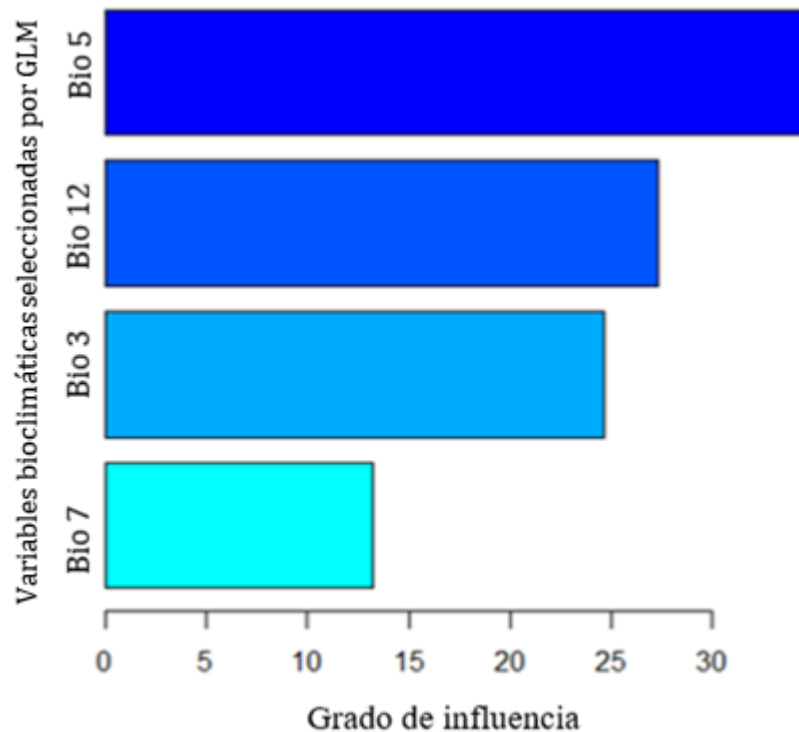


Figura 8. Análisis de la influencia de las variables bioclimáticas en la selección del modelo GLM: Una comparación visual de la precipitación mensual (bio 5), temperatura promedio máxima (bio 12) y mínima (bio 7) y precipitación total anual (bio 3).

4.1.3. Distribución potencial y presencia de *Fusarium oxysporum* en Ecuador continental: Análisis a través de mapas de favorabilidad y presencia

El mapa de presencia de *F. oxysporum* en Ecuador continental, basado en el modelo GLM, muestra una distribución heterogénea de la especie en el territorio. El mapa de favorabilidad (Figura 9a), 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 9b), 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 y Tungurahua, con un nivel alrededor del 0.050.

Para el modelo GAM, muestran una distribución no homogénea de la especie en el territorio. 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 pueden encontrar en las provincias de Manabí, Guayas y parte de El Oro, con valores alrededor de 0.70. En cambio, en la región del oriente, la presencia es muy baja (Figura 9c). Por el contrario, el mapa de probabilidad de presencia muestra que para la mayoría del territorio hay una baja presencia de la especie, con la excepción de la sierra, donde aparece con un nivel aproximado de 0.020 (Figura 9d).

El modelo obtenido a través de MAXENT, muestra una distribución heterogénea de *F. oxysporum* en territorio continental ecuatoriano. El mapa de favorabilidad muestra una presencia significativa de más de 0.90 en la región Sierra. En la región Costa solo se localiza en las provincias de Manabí, Guayas, Santa Elena y El Oro, con un valor cercano a 0.50. En la región del oriente la presencia es nula (Figura 9e). Por su parte, el mapa de probabilidad de presencia (Figura 11b) es similar al de favorabilidad, en cuanto a la distribución de la especie.

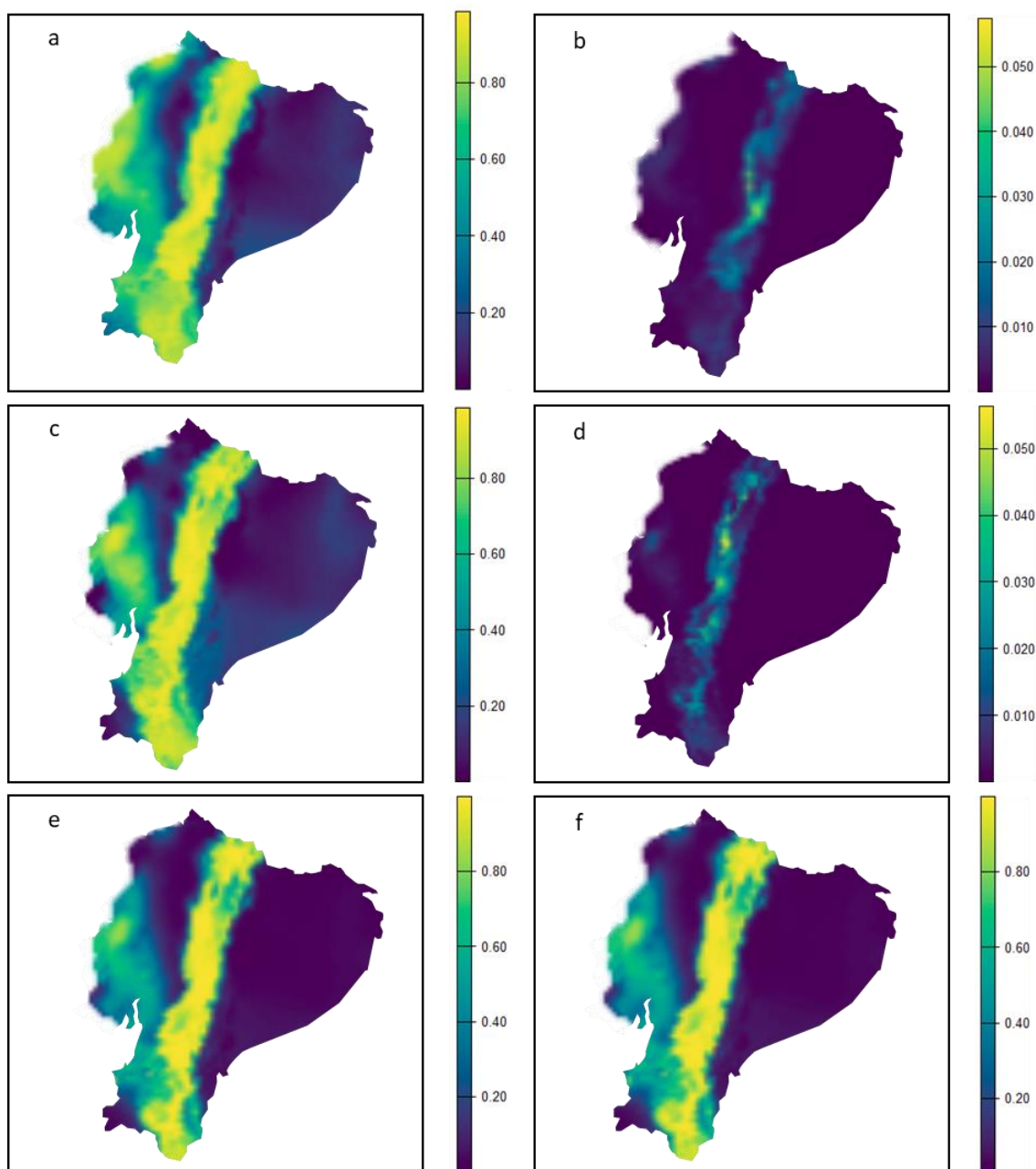


Figura 9. Mapas de favorabilidad y probabilidad de presencia de *F. oxysporum*: modelo GLM; figura (a) distribución potencial, figura (b) probabilidad de presencia. Modelo GAM: figura (c) distribución potencial, figura (d) probabilidad de presencia. Modelo MAXENT: figura (e) distribución potencial, figura (f) probabilidad de presencia. Leyenda: La escala de color, desde el color morado más oscuro hasta el color amarillo más claro, representa la influencia de la variable en el nivel de presencia de la especie

Los resultados del modelo RF muestran la menor representación de *F. oxysporum* en territorio ecuatoriano continental. El mapa de favorabilidad destaca una mayor presencia de la especie en la región Sierra y una disminución de esta en las provincias de Manabí, Guayas, Esmeraldas y El Oro; en la región Costa. En la región del oriente una vez más se observa una presencia nula (Figura 10a). En el mapa de presencia real (Figura 10b), solo aparecen una cantidad limitada de puntos en las provincias de Pichincha y Cotopaxi.

Para el mapa basado en el modelo GBM, muestran también una distribución heterogénea de la especie en el territorio. El mapa de favorabilidad (Figura 10c), muestra una presencia superior a 0.90 en la región Sierra. En la región Costa, en las provincias de Manabí, Guayas, Santa Elena, Esmeraldas Los Ríos y El Oro muestra una presencia con un valor cercano a 0.60, mientras que en la región del Oriente se observa una presencia cercana al 0.30 en la mayoría de sus provincias. Para el mapa de probabilidad de presencia (Figura 10 d), está ligeramente representada en toda la región de la Sierra.

Con el modelo BART, se puede observar que la distribución de *F. oxysporum* en el territorio del Ecuador Continental es también heterogénea. El mapa de favorabilidad (Figura 14e), muestra una alta probabilidad de presencia en la región Sierra, con valores superiores a 0.90. En las provincias de Manabí, Guayas, Esmeraldas, Los Ríos, Santo Domingo y El Oro, de la región Costa, los valores son cercanos a 0.80. En contraste, en la región del Oriente, casi en su totalidad de superficie se observan valores cercanos a 0.30. En el mapa de probabilidad de presencia (Figura 10f), 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.

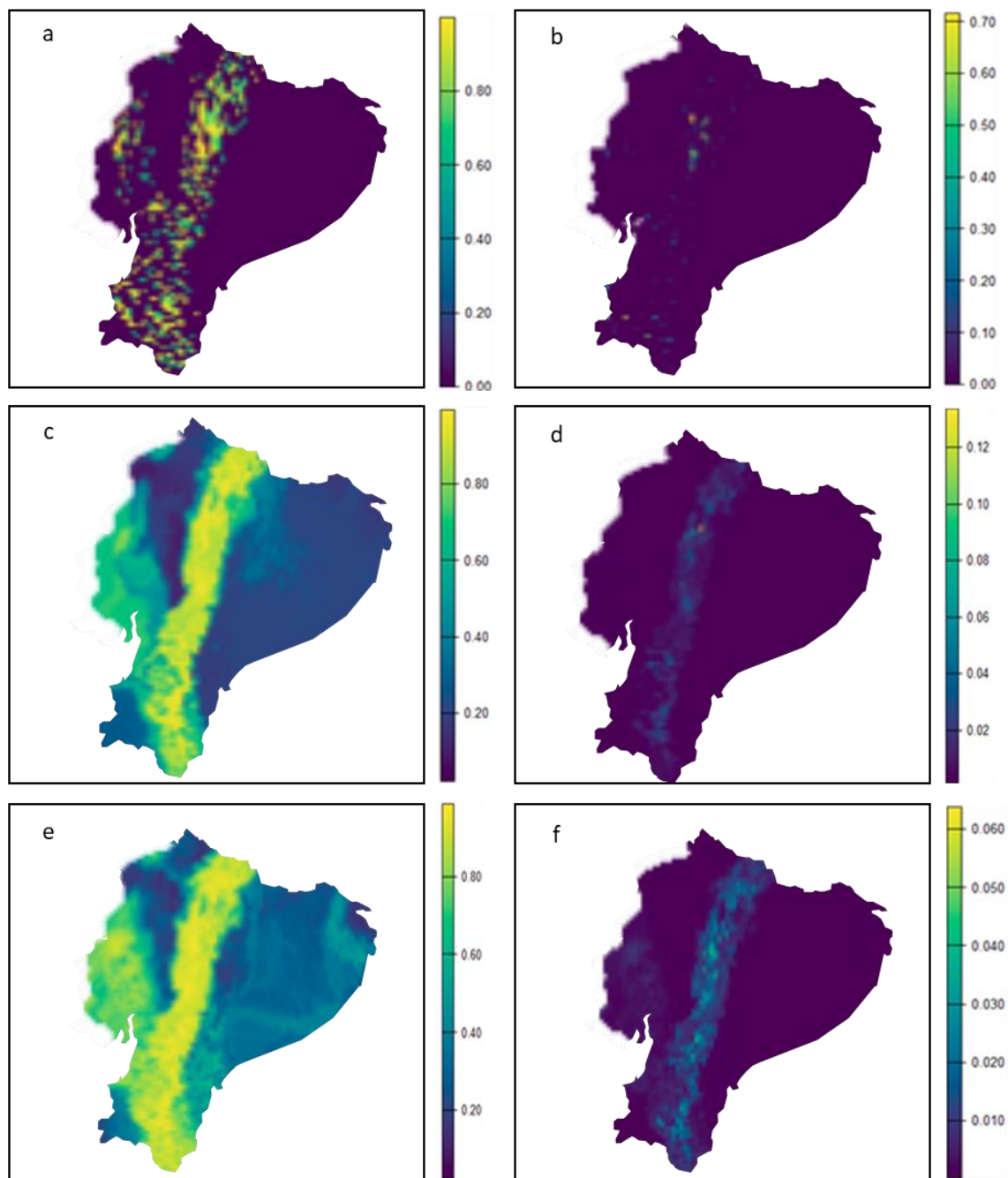


Figura 10. Mapas de favorabilidad y probabilidad de presencia de *F. oxysporum*: modelo RF; figura (a) distribución potencial, figura (b) probabilidad de presencia. Modelo GBM: figura (c) distribución potencial, figura (d) probabilidad de presencia. Modelo BART: figura (e) distribución potencial, figura (f) probabilidad de presencia. Leyenda: La escala de color, desde el color morado más oscuro hasta el color amarillo más claro, representa la influencia de la variable en el nivel de presencia de la especie

4.2. Caracterizar la distribución potencial de *F. oxysporum* en Ecuador continental según los métodos de modelación empleados.

4.2.1. Mapas binarios presencia-ausencia

Los resultados del mapa binario según el modelo GLM (Figura 11a), el hongo *F. oxysporum* está ausente en una superficie de 163700 km² en el territorio continental de Ecuador. Esta ausencia es predominantemente en la región del Oriente y parte de las provincias de Los Ríos, Santo Domingo y Esmeraldas (Figura 11a). Por otro lado, se ha identificado la presencia del hongo en una superficie de 127200 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 11a).

Para el modelo GAM (Figura 11b), la especie tiene una distribución potencial limitada en el territorio continental de Ecuador. La especie está ausente en un área de 168800 km² en la región del Oriente y parte de las provincias de Los Ríos, Santo Domingo, Esmeraldas y Santa Elena (Figura 11b). Por otro lado, se encontró presente en un área de 122100 km² en la región Sierra y parte de las provincias del Guayas, Manabí y El Oro (Figura 11b).

Con el modelo MAXENT, el hongo *F. oxysporum* se estimó su presencia en un total de 106800 km², correspondiendo principalmente a la región Sierra y a las provincias de Guayas, Santa Elena, El Oro (Figura 11c). En cuanto a la ausencia ocupan un área de 184100 km² (Figura 11c).

De acuerdo con el modelo de Random Forest (RF) (Figura 11d), *F. oxysporum* está ausente en una porción significativa del territorio ecuatoriano continental, en una superficie de 243800 km², concentrándose principalmente en la región del Oriente y en las provincias de Los Ríos, Santo Domingo y Esmeraldas (Figura 11d). La especie está presente en una superficie de solo 47100 km², con una concentración más alta en la región Sierra y en las provincias de Santa Elena, Guayas y Manabí (Figura 11d).

Para el modelo GBM (Figura 11e), indica que un área total de 180500 km² la especie está ausente, predominantemente en la región del Oriente y parte de las provincias de Los Ríos, Santo Domingo y Esmeraldas. La especie está presente en un área de 110600 km², en particular en la región Sierra y parte de las provincias del Guayas, Manabí y El Oro (Figura 11e).

Según el modelo BART (Figura 11f), la especie está ausente en la región del Oriente, excepto en la provincia de Zamora Chinchipe y en las provincias de Los Ríos, Santo Domingo, Santa Elena, Guayas y Esmeraldas, abarcando un área total de 181500 km². Está presente mayormente en la región Sierra y en una proporción muy baja en la Costa, con un área total de 109400 km² (Figura 11f)

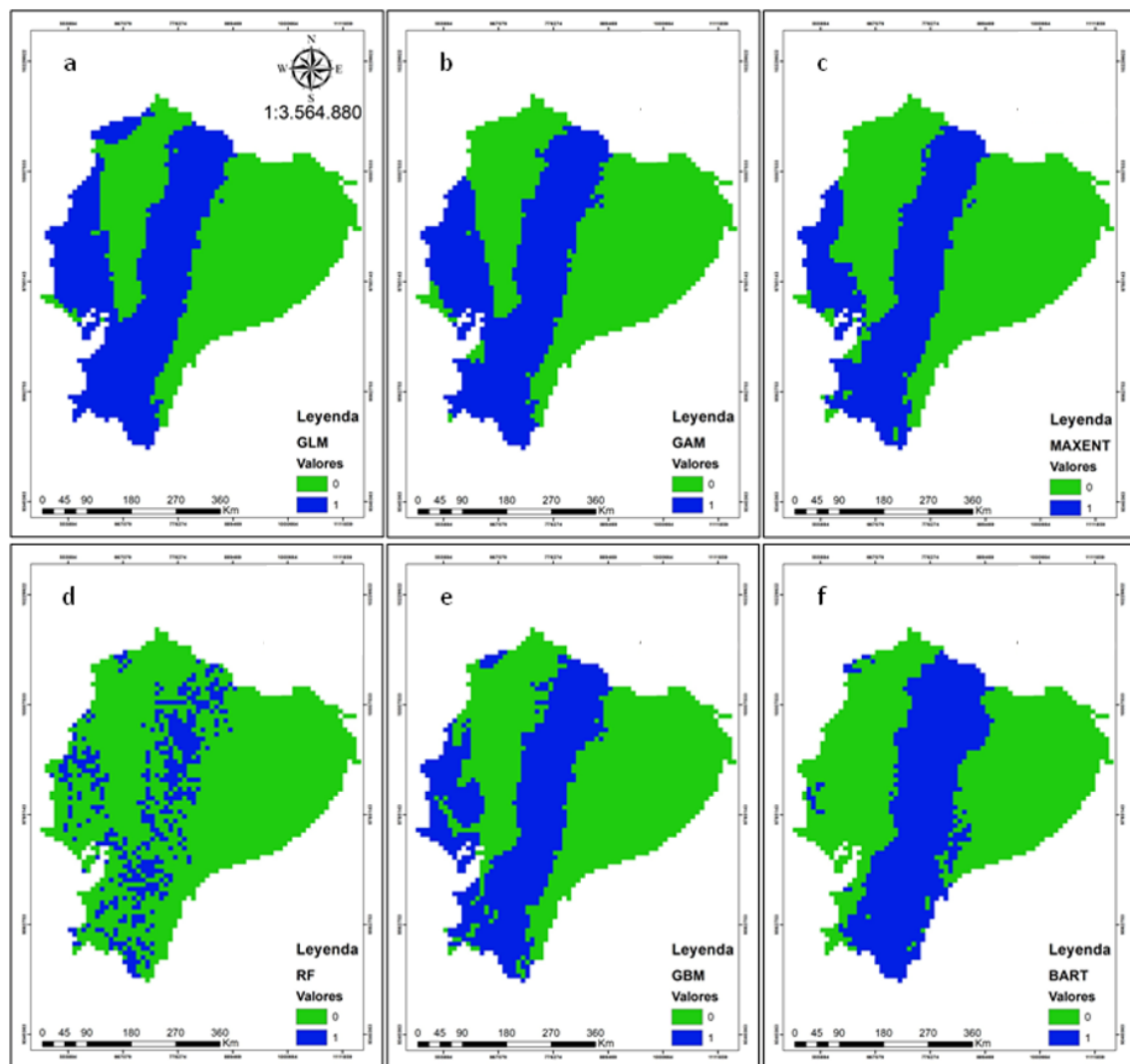


Figura 11. Distribución de *F. oxysporum* en Ecuador Continental: Mapa Binario de Presencia-Ausencia (1=presencia; 0= ausencia) generado con el modelo GLM (Figura 11a), modelo GAM (Figura. 11b), modelo MAXENT (Figura. 11c), modelo RF (Figura. 11d), modelo GBM (Figura. 11e), modelo BART (Figura. 11f)

4.3. Evaluación de las métricas de rendimiento de los modelos de *F. oxysporum* según los enfoques estadísticos empleados

4.3.1. Métricas de validación con umbral de prevalencia

El mejor modelo en términos de rendimiento según las métricas evaluadas es 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 skappa de 0.52. El modelo con el peor rendimiento fue MAXENT (MX) con un CCR de 0.02 y una precisión de 0.00. Se sugiere revisar la figura 12 para una visualización completa de los resultados

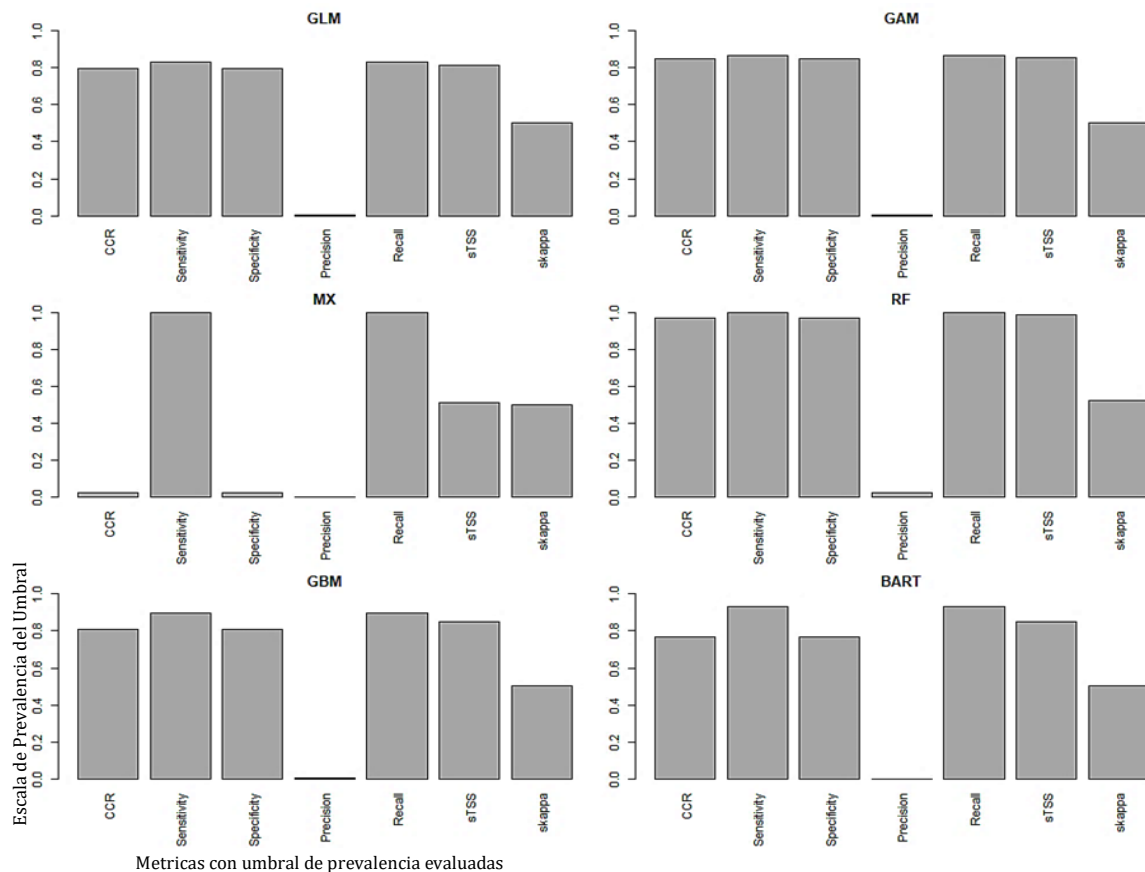


Figura 12. Evaluación del rendimiento de seis modelos (GLM, GAM, MX, RF, GBM y BART) a través de métricas con un umbral de prevalencia: Correct Classification Rate (CCR), sensitivity, specificity, precision, recall, True Skill Statistic (TSS) y skappa. Leyenda: Eje y: Valores con umbral de prevalencia con un rango desde 0.0 hasta 1.0. Eje x: Nombres de las métricas de rendimiento evaluadas

4.3.2. Métricas de validación con umbral optimo

El modelo Random Forest (RF) obtuvo el mejor rendimiento en la métrica Correct Classification Rate (CCR) con 0.98, la mejor sensibilidad y especificidad con valores de 1 y 0.98 respectivamente, la mejor precisión con 0.02, la mejor métrica Recall con 1, el mejor True Skill Statistic (TSS) con 0.99 y una skappa de 0.46. Se sugiere revisar la figura 13 para una visualización completa de los resultados. MX y RF tuvieron la mejor sensibilidad con 1. GLM, GBM y BART también tuvieron una skappa de 0.46.

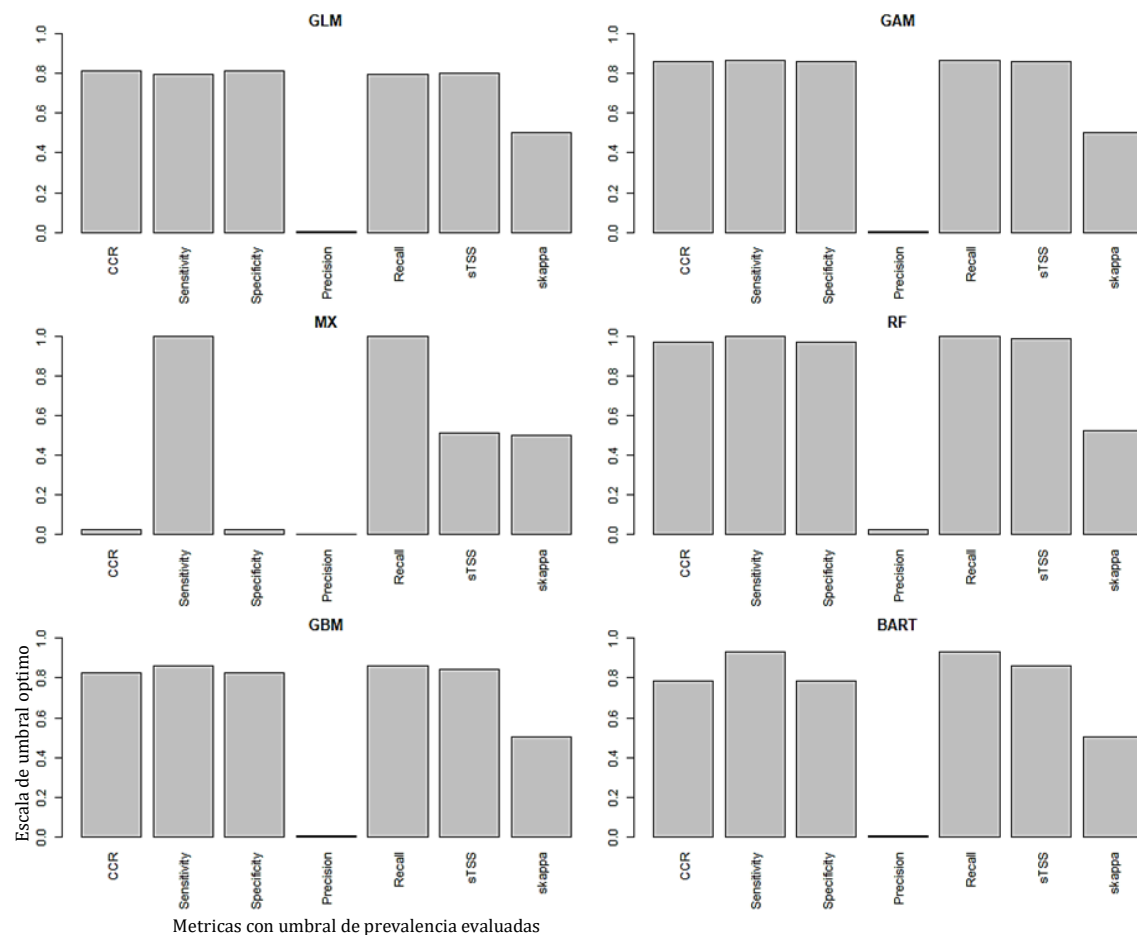


Figura 13. Evaluación del rendimiento de seis modelos (GLM, GAM, MX, RF, GBM y BART) a través de métricas con un umbral optimo: Correct Classification Rate (CCR), sensitivity, specificity, precision, recall, True Skill Statistic (TSS) y skappa. Leyenda: Eje y: Valores de umbral optimo con un rango desde 0.0 hasta 1.0. Eje x: Nombres de las métricas de rendimiento evaluadas

4.3.3. Área bajo la Curva ROC (AUC)

Los resultados de la métrica de evaluación AUC para los seis modelos (figura 14) fueron los siguientes: RF obtuvo la máxima AUC de 1, seguido de BART con 0.942, GBM con 0.935, MX con 0.929, GAM con 0.93 y GLM con 0.913. Estos resultados sugieren que RF tuvo la mejor habilidad para distinguir entre las dos clases, seguido por BART, GBM, MX, GAM y GLM en ese orden.

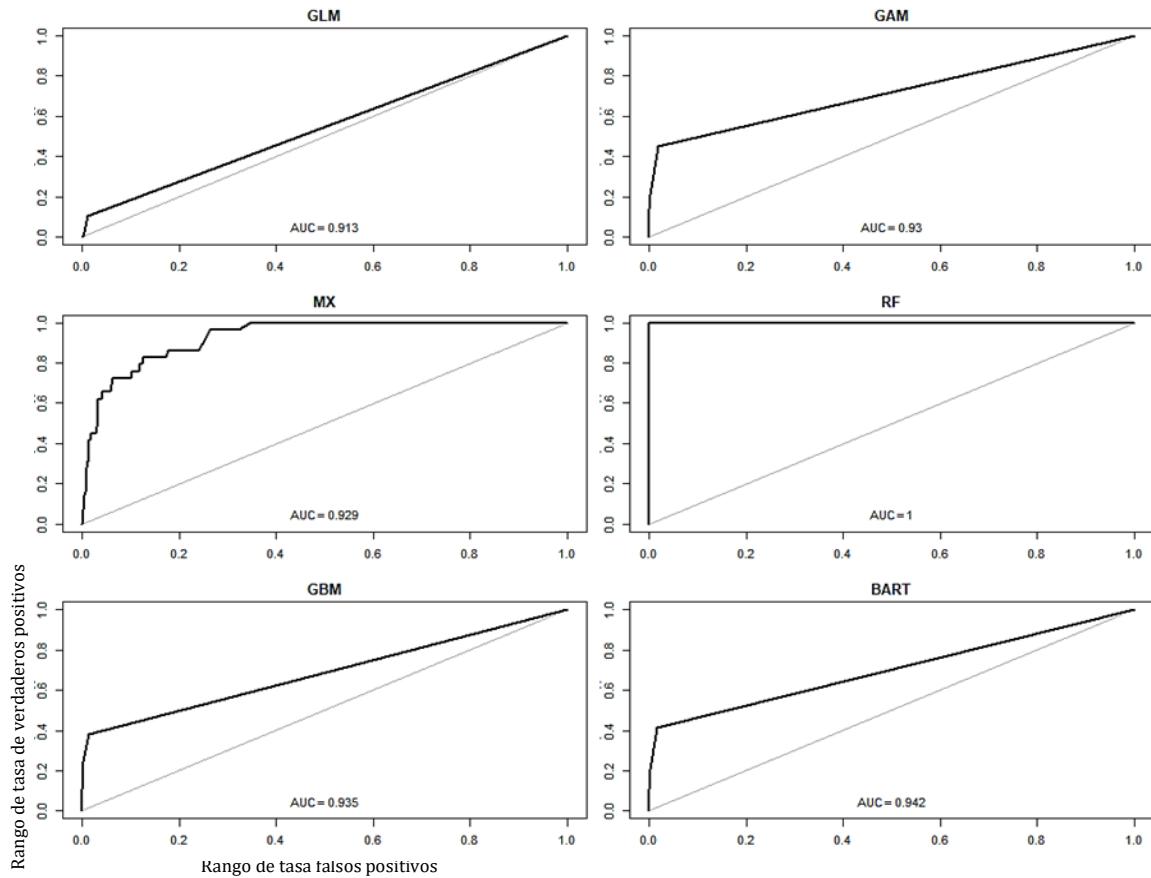


Figura 14. Análisis de la eficacia de los modelos (GLM, GAM, MX, RF, GBM y BART) en la predicción de la distribución potencial de *F. oxysporum* mediante la métrica Área bajo la Curva (AUC). Leyenda: El eje “x” representa la tasa de falsos positivos (1-especificidad) y el eje “y” representa la tasa de verdaderos positivos (sensibilidad)

4.3.4. Línea de calibración Miller

En los resultados de la línea de calibración de Miller, GLM tuvo una pendiente de 1 y una intercepción de 0, lo que lo coloca en el primer lugar en términos de rendimiento (Figura 15). GAM siguió con una pendiente de 1.09 y una intercepción de 0.45 (Figura 15). 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 15). Por último, MX tuvo una pendiente de 0.51 y una intercepción de -6.42, y RF tenía una pendiente de 14.74 y una intercepción de 8.18 (Figura 15).

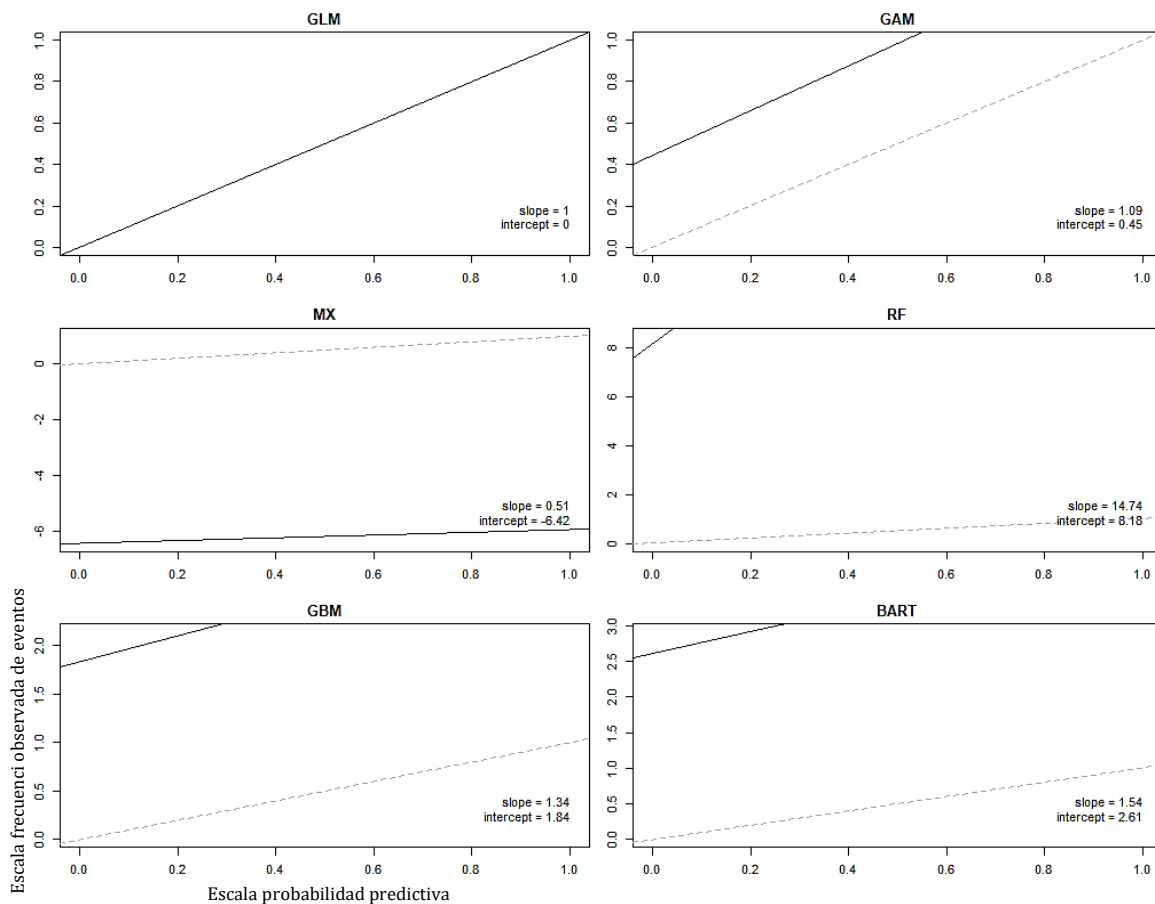


Figura 15. Calibración de los modelos (GLM, GAM, MX, RF, GBM y BART) utilizando la línea de calibración de Miller para medir la precisión de la distribución potencial de *F. oxysporum*. Leyenda: eje “x” representa la probabilidad predictiva, eje “y” representa la frecuencia observada de eventos

4.4. Discusión

En la actualidad, dada la importancia de la seguridad alimentaria de la población, y la pérdida de biodiversidad en todo el mundo, resulta importante contar con mecanismos que permitan tener una mejor planificación en caso de proliferación de especies invasoras, que puedan alterar los sistemas agrícolas y ambientales [39]. Desde este punto de vista, los modelos de distribución de especies resultan ser herramientas valiosas en la planificación y prevención de escenarios peligrosos, brindando estimaciones de las regiones más propensas a la propagación de patógenos.

En Ecuador, se han realizado varios estudios referentes a la modelación de distribución de especies invasoras [40] [41]. 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 estudios como el que se presenta, que sirvan como línea base para futuras investigaciones respecto a patógenos como *F. oxysporum* en Ecuador, donde se compruebe la eficacia de los modelos bayesianos, complementándolo con metodologías aplicadas en países vecinos como Colombia y Perú, donde si se han realizado investigaciones relacionadas con este patógeno [42].

Los MDE pese 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 [43]. Debido a esto en la literatura [44] [45], recomiendan utilizar múltiples modelos para evaluar la efectividad y precisión, seleccionando siempre el que muestre los mejores resultados. 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 [45] [46].

De los seis modelos que se tuvieron en cuenta para este estudio (GLM, GAM, MAXENT, RF, GBM y BART), se compararon entre sí, mediante el uso de diferentes métricas de evaluación sugeridas en la literatura [47]. Gracias a esta comparación, se pudo seleccionar el modelo más apropiado para la especie en Ecuador continental. Se demostró, que Random Forest presento los mejores resultados en las métricas de precisión, sensibilidad,

especificidad, veracidad, precisión y Kappa, lo cual también se reflejó en la máxima tasa de clasificación correcta y en el AUC. Este resultado coincide con lo mostrado en otros estudios donde Random Forest destacó al compararlo con otros [48]. No obstante, es importante tener en cuenta que con el modelo GLM también se obtuvieron buenos resultados en la línea de calibración Miller, donde fue el único que obtuvo valor de 1, indicando que el modelo fue muy preciso.

Es recomendable tener presente que los MDE muestran resultados diferentes aplicados en distintas especies. Por ejemplo, Random Forest ha presentado resultados pobres en comparación con Maxent y GLM aplicado a especies en Australia, esto puede ser debido a un conjunto de datos específico y la especie en cuestión. Es posible que en ciertas especies o regiones geográficas, un modelo determinado funcione mejor que otros debido a las características únicas de los datos. [49]. 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 ($AUC > 0,9$ y $TSS > 0,5$). Sin embargo, se observó una variabilidad en la clasificación de datos fuera de bolsa, con los modelos GBM y RF mostrando una mayor sensibilidad a los datos de entrenamiento y el modelo GLM mostrando una especificidad más baja. El modelo MaxEnt fue el más efectivo en la identificación de datos de prueba en procedimientos aleatorios y de ponderación de fondo [50]. Esto se puede deber a la capacidad del modelo para manejar eficazmente datos categóricos y continuos y su capacidad para incorporar información geográfica y ambiental en el modelo de manera efectiva. Además, las técnicas de regularización utilizadas por MaxEnt pueden ayudar a evitar el sobreajuste del modelo y mejorar la capacidad de generalización del modelo [51].

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 puede influir en los resultados obtenidos. Cada modelo tiene diferentes hiperparámetros que deben ajustarse para optimizar el rendimiento del modelo. Si estos hiperparámetros no se ajustan adecuadamente, puede haber un sesgo en el modelo hacia ciertos tipos de variables predictoras, lo que puede afectar la precisión de

la predicción [32]. 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 la precisión del modelo. 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 [14].

Estos modelos requieren una cantidad suficiente de datos de presencia para funcionar adecuadamente [52]. Si la cantidad de datos es insuficiente, el modelo puede 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 o considerar la implementación de enfoques alternativos, como modelos no bayesianos [11].

En los resultados de la investigación, aunque el modelo BART no presenta una destacada mejora 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 previo realizado en Bosnia y Herzegovina [53], se demostró que el modelo 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.

Considerando lo expuesto en los resultados de la investigación anteriormente mencionada, es importante destacar la relevancia de evaluar diferentes modelos de modelación de especies en circunstancias específicas, como las relacionadas con la producción agrícola. En el caso de Ecuador como un país productor agrícola se ve altamente vulnerable a especies invasoras que alteren los ecosistemas. Estas invasiones biológicas representan una de las principales causas de pérdida de biodiversidad y extinción de especies [54] [55]. Como también responsables de pérdidas económicas (por ejemplo, pérdidas directas de agricultura, silvicultura, ecosistemas, etc.) [56] [57]. En Ecuador el 20% de la población se dedica al sector agropecuario y se estima que un 54% desempeña actividades agrícolas (Instituto Nacional de Estadística y Censos, 2020).

De acuerdo con Random Forest, que presentó los mejores resultados, se observa que el patógeno tiene altas probabilidades de extenderse en toda la región Sierra y parte de las provincias de la Costa. Estos resultados sugieren que la presencia o ausencia del hongo puede estar asociada con las características geográficas, climáticas y ecológicas de cada región. La región de la Sierra, por ejemplo, puede tener características que sean favorables para el desarrollo y expansión del hongo, esto podría deberse a una temperatura y humedad adecuadas, así como características del suelo y disponibilidad de hospedantes. Por otro lado, la ausencia de la especie en la región del Oriente puede estar relacionada con factores que inhiben su crecimiento y desarrollo, como pueden ser las condiciones climáticas más secas y calurosas que no son favorables para el hongo, o la ausencia de hospedantes adecuados en la región.

Para las personas que viven en estas regiones con mayor probabilidad de invasión, 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. 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% [58].

Los resultados obtenidos en este estudio son de gran relevancia para el monitoreo y control de la distribución del patógeno *F. oxysporum* en el Ecuador. La identificación de las regiones con alta presencia y baja probabilidad de esta especie invasora permitirá tomar medidas preventivas para minimizar su impacto en la agricultura y el medio ambiente. Es importante destacar que diversos estudios han demostrado la efectividad de modelos como los utilizados en este estudio para predecir la aparición de especies invasoras y, así, implementar medidas preventivas para reducir su impacto o controlarlas [39].

En particular, se observó una ausencia común de la presencia del patógeno en las provincias del Oriente, como Sucumbíos, Orellana, Pastaza y Morona Santiago, lo que sugiere que estas regiones pueden ser idóneas para la agricultura. Sin embargo, debido a que 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.

Si bien 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 Rios, como Quevedo, Mocache, Ventanas, Quimsaloma, Valencia y Buena Fe, donde la presencia de *F. oxysporum* tendría una baja probabilidad. Estas áreas pueden ser consideradas como las más adecuadas dentro de la región costa para la producción de cultivos agrícolas.

CAPÍTULO V
CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

Las variables ambientales BIO 3, 5, 7 y 12 han sido identificadas como las más influyentes en la distribución de *F. oxysporum* en la región continental de Ecuador. La temperatura y la precipitación son dos de los factores más importantes que influyen en la germinación, el crecimiento y la supervivencia de los hongos, y juegan un papel clave en determinar la distribución de *F. oxysporum* en una región específica. Sin embargo, es importante tener en cuenta que otros factores, como la presencia de hospedadores y la acción humana, también pueden influir en la diseminación del hongo y deben ser considerados en cualquier análisis de distribución de especies.

Los resultados obtenidos en la evaluación de las métricas de los modelos indican que el algoritmo Random Forest (RF) es el más adecuado para modelar la distribución potencial de *F. oxysporum* en el territorio continental de Ecuador. Este modelo obtuvo los mejores resultados en las métricas de precisión, sensibilidad, especificidad, TSS y precisión, lo que se reflejó en la máxima tasa de clasificación correcta y en la máxima área bajo la curva ROC. No obstante, es importante tener en cuenta que con el modelo GLM se obtuvieron buenos resultados en la línea de calibración Miller y que los modelos bayesianos también tuvieron un buen rendimiento.

En general, todos los modelos pueden verse afectados por la cantidad insuficiente de registros para la especie en cuestión. Sin embargo, según los resultados obtenidos en el estudio, el modelo Random Forest (RF) tuvo los mejores resultados en las métricas de precisión, especificidad, sensibilidad, CCR y TSS. Esto sugiere que, RF puede ser el modelo más robusto en casos de baja cantidad de datos de presencia de la especie para modelar la distribución potencial de *F. oxysporum*. Sin embargo, estos resultados deben interpretarse con precaución, ya que la selección del mejor modelo depende de varios factores, incluyendo la complejidad del problema y la calidad de los datos disponibles.

5.2. Recomendaciones

A partir de estos resultados, se pueden recomendar medidas de control y prevención de la presencia de *F. oxysporum* en aquellas áreas donde se observa una mayor presencia de la especie, especialmente en la región Sierra. Además, es importante seguir realizando investigaciones para determinar las causas y factores que influyen en la distribución de la especie en el territorio, con el fin de desarrollar medidas más efectivas de prevención y control.

Para superar las limitaciones en la cantidad de registros de presencia de la especie *F. oxysporum* en el Ecuador continental y mejorar el rendimiento del modelo BART, se recomienda utilizar técnicas de extrapolación. Por ejemplo, la interpolación y la extrapolación cúbica pueden ser útiles para estimar la presencia de la especie en áreas sin datos. La implementación de estas técnicas permitirá una comprensión más precisa y completa del rango potencial de distribución de la especie, lo que a su vez aumentará la precisión y la fiabilidad de los modelos BART.

Se sugiere realizar estudios comparativos adicionales entre los modelos (GLM, GAM, MAXENT, RF, GBM y BART) y con diferentes especies que tengan condiciones climáticas de adaptación variadas. Esto permitirá comparar la efectividad de los modelos en diferentes situaciones y evaluar la capacidad de los modelos para ser aplicables en distintos contextos. Además, proporcionará información valiosa sobre la conexión entre el clima y la distribución de las especies, lo cual puede tener consecuencias significativas para la gestión y preservación de la biodiversidad.

CAPÍTULO VI
BIBLIOGRAFÍA

6.1. Bibliografía

- [1] J. R. Lamichhane, M. Barzman, K. Booij, P. Boonekamp, N. Desneux, L. Huber, P. Kudsk, S. Langrell, A. Ratnadass, P. Ricci, J.-L. Sarah y A. Messéan, «Robust cropping systems to tackle pests under climate change. A review,» *Agronomy for Sustainable Development*, vol. 35, pp. 443-459, December 2014.
- [2] S. Chakraborty y A. Newton, «Climate change, plant diseases and food security: an overview,» *Plant Pathology*, vol. 60, pp. 2-14, February 2011.
- [3] R. Macedo, L. P. Sales, F. Yoshida, L. L. Silva-Abud y J. Lobo M, «Potential worldwide distribution of Fusarium dry root rot in common beans based on the optimal environment for disease occurrence,» *PLOS one*, vol. 12, nº 11, November 2017.
- [4] D. P. Bebber, T. Holmes y S. J. Gurr, «The global spread of crop pests and pathogens,» *Global Ecology and Biogeography*, vol. 23, pp. 1398-1407, August 2014.
- [5] S. Chakraborty, «Migrate or evolve: options for plant pathogens under climate change,» *Global Change Biology*, vol. 19, nº 7, March 2013.
- [6] L. Burgess, «General Ecology of the Fusaria,» pp. 225-235, 1981.
- [7] J. C. Correll, «The Relationship Between Formae Speciales, Races, and Vegetative Compatibility Groups in *Fusarium oxysporum*,» *Phytopathology*, vol. 81, nº 9, pp. 1061-1064, 4 February 1991.
- [8] S. N. Smith, «An Overview of Ecological and Habitat Aspects in the Genus *Fusarium* with Special Emphasis on the Soil-Borne Pathogenic Forms,» *Plant Pathology Bulletin*, vol. 16, pp. 97-120, 1 August 2007.
- [9] M. B. Araújo y M. New, «Ensemble forecasting of species distributions. Trends in Ecology & Evolution,» *Trends in Ecology & Evolution*, vol. 22, nº 1, pp. 42-47, January 2007.
- [10] A. Townsend Peterson, J. Soberón, R. G. Pearson, R. P. Anderson, E. Martínez-Meyer, M. Nakamura y M. Bastos Araújo, *Ecological Niches and Geographic Distributions (MPB-49)*, Princeton University Press, 2011, pp. 1-328.
- [11] C. J. Carlson, «embarcadero: Species distribution modelling with Bayesian additive regression trees in r,» *Methods in Ecology and Evolution*, vol. 11, nº 7, pp. 850-858,

14 March 2020.

- [12] J. Z. Kosicki, «Generalised Additive Models and Random Forest Approach as effective methods for predictive species density and functional species richness,» *Environmental and Ecological Statistics*, vol. 24, pp. 273-292, 29 April 2020.
- [13] T. R. Gordon, «Fusarium oxysporum and the Fusarium Wilt Syndrome,» *Annual Review of Phytopathology*, vol. 55, pp. 23-39, 10 May 2017.
- [14] D. Zurell, J. Franklin, C. König, P. J. Bouchet, C. F. Dormann, J. Elith, G. Fandos y X. Feng, «A standard protocol for reporting species distribution models,» *Ecography*, vol. 43, nº 9, pp. 1261-1277, 1 June 2020.
- [15] S. Brodie, J. T. Thorson, G. Carroll, E. L. Hazen, S. Bograd, M. A. Haltuch, K. K. Holsman, S. Kotwicki, J. F. Samhoury, E. Willis-Norton y R. L. SeldenS, «Trade-offs in covariate selection for species distribution models: a methodological comparison,» *Ecography*, vol. 43, nº 1, pp. 11-24, 2 October 2019.
- [16] J. Luan, C. Zhang, B. Xu, Y. Xue y Y. Ren, «The predictive performances of random forest models with limited sample size and different species traits,» *Fisheries Research*, vol. 227, July 2020.
- [17] S. Brodie, J. Thorson, . G. Carroll, . E. Hazen, S. Bograd, M. Haltuch, K. Holsman y S. Kotwicki, «Compensaciones en la selección de covariables para modelos de distribución de especies: una comparación metodológica,» *Ecography*, vol. 43, nº 1, pp. 11-24, January 2020.
- [18] J. Bradie y B. Leung, «A quantitative synthesis of the importance of variables used in MaxEnt species distribution models,» *Journal of Biogeography*, vol. 44, nº 6, pp. 1344-1361, June 2017.
- [19] L. Lombard, M. Sandoval-Denis, S. Lamprecht y P. Crous, «Epitypification of Fusarium oxysporum – clearing the taxonomic chaos,» *Persoonia - Molecular Phylogeny and Evolution of Fungi*, vol. 43, December 2019.
- [20] S. Pironon, J. Villellas, W. Thuiller, V. Eckhart, M. Geber, D. Moeller y M. García, «The 'Hutchinsonian niche' as an assemblage of demographic niches: Implications for species geographic ranges,» *Ecography*, vol. 41, nº 7, 2017.
- [21] J. Soberón y B. Arroyo-Peña, «Are fundamental niches larger than the realized? Testing a 50-year-old prediction by Hutchinson,» *PLoS ONE* , vol. 12, nº 4, April 2017.

- [22] G. Hutchinson, «Concluding remarks,» *Cold Spring Harb Sym* 22, p. 415–427, 1957.
- [23] . T. Booth, «Assessing species climatic requirements beyond the realized niche: some lessons mainly from tree species distribution modelling,» *Climatic Change*, vol. 145, p. 259–271, 31 October 2017.
- [24] A. T. Peterson, M. E. Cobos y D. Jimenez-Garcia, «Major challenges for correlational ecological niche model projections to future climate conditions,» *Annals of the New York Academy of Sciences*, vol. 1429, n° 1, October 2018.
- [25] C. Staver, D. E. Pems, L. Scheerer, L. Perez Vicente y M. Dita, «Ex Ante Assessment of Returns on Research Investments to Address the Impact of Fusarium Wilt Tropical Race 4 on Global Banana Production,» *Frontier in Plant Science*, vol. 11, n° 844, pp. 1-14, 6 July 2020.
- [26] R. Orr y P. N. Nelson, «Impacts of soil abiotic attributes on Fusarium wilt, focusing on bananas,» *Applied Soil Ecology*, vol. 132, pp. 20-33, December 2018.
- [27] IPPC Secretariat, «2021 IPPC Annual Report - Protecting the world's plant resources from pests.,» Rome, 2022.
- [28] J. Zhang, A. Abdelraheem, Y. Zhu, T. A. Wheeler, J. K. Dever, R. Nichols y T. Wedegaertner, «Importance of temperature in evaluating cotton for resistance to Fusarium wilt caused by Fusarium oxysporum f. sp. vasinfectum race 4,» *Crop Science*, pp. 1-14, 5 January 2021.
- [29] F. Magdama, «Fusarium oxysporum - El hongo má temido en la industria del banano,» *Revista Científica Ecuador es Calidad*, vol. 6, n° 1, 19-22 agosto 2019.
- [30] A. Guisan y W. Thuiller, «Predicting species distribution: offering more than simple habitat models,» *Ecology Letters*, vol. 8, n° 993–1009, 2005.
- [31] H. Nix, J. McMahon y D. Mackenzie, «Potential areas of production and the future of pigeon pea and other grain legumes in Australia. In: The potential for pigeon pea in Australia. Proceedings of Pigeon Pea (Cajanus cajan (L.) Millsp.) Field Day (eds Wallis, E.S. & Whiteman, P.C.),» *University of Queensland*, pp. 1- 5, 1977.
- [32] G. Guillera-Aroita, J. Lahoz-Monfort, J. Elith, A. Gordon, H. Kujala, P. Lentini, M. McCarthy, R. Tingley y B. Wintle, «Is my species distribution model fit for purpose? Matching data and models to applications,» *Global Ecology and Biogeography*, vol. 24, n° 3, p. 276–292, 08 January 2015.

- [33] E. Thibaud, . B. Petitpierre, O. Broennimann, A. Davison y A. Guisan, «Measuring the relative effect of factors affecting speciesdistribution model predictions,» *Methods in Ecology and Evolution*, vol. 5, n° 9, pp. 947-955, 05 May 2014.
- [34] J. Kleemann, H. Koo, I. Hensen, G. Mendieta-Leiva, B. Kahnt, C. Kurze, D. J. Inclan, P. Cuenca, J. K. Noh, M. H. Hoffmann, A. Factos, M. Lehnert, P. Lozano y C. Furst, «Priorities of action and research for the protection of biodiversity and ecosystem services in continental Ecuador,» *Biological Conservation*, vol. 265, January 2022.
- [35] GBIF, «GBIF | Global Biodiversity Information Facility,» [En línea]. Available: <https://www.gbif.org/>.
- [36] M. Arellano, «Detección de Fusarium oxysporum en cultivos de uvilla (Physalis peruviana L.) en la Sierra norte y centro del Ecuador,» 2018.
- [37] WorlClim, «WorlClim,» [En línea]. Available: <https://worldclim.org/data/index.html>.
- [38] S. E. Fick y R. J. Hijmans, «WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas,» *International Journal of Climatology*, vol. 37, n° 12, pp. 4302-4315, 2017.
- [39] M. Barbet-Massin, Q. Rome, C. Villemant y F. Courchamp, «Can species distribution models really predict the expansion of invasive species?,» *PLoS ONE*, vol. 13, n° 3, pp. 1-14, 2018.
- [40] Y. Ferrer-Sánchez, W. Jacho-Saa, J. Urdánigo y F. Abasolo, «Invasiones biológicas en agroecosistemas de Ecuador Continental: nicho ecológico de especies exóticas y cultivos agrícolas bajo riesgo,» *Acta Biológica Colombiana*, vol. 26, n° 3, pp. 352-364, 2022.
- [41] S. R. R. Ron , «Predicting the Distribution of the Amphibian Pathogen Batrachochytrium dendrobatidis in the New World,» *bioTROPICA*, vol. 37, n° 2, pp. 209-221, 2005.
- [42] A. Acurio, V. Rafael y O. Dangles, «Biological Invasions in the Amazonian Tropical Rain Forest: The Case of Drosophilidae (Insecta, Diptera) in Ecuador, South America,» *bioTROPICA*, vol. 42, n° 6, pp. 717-723, 2010.
- [43] H. Sofaer, C. Jarnevich, I. Pearse, R. Smyth y S. Auer, «Development and Delivery of Species Distribution Models to Inform Decision-Making,» *BioScience*, vol. 69,

- n° 7, p. 544–557, 2019.
- [44] A. Norberg, N. Abrego, G. Blanchet, F. Adler y B. Anderson, «A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels,» vol. 89, n° 3, pp. 1-24, 2019.
- [45] T. Hao , J. Elith, J. Lahoz y G. Guillera, «Testing whether ensemble modelling is advantageous for maximising predictive performance of species distribution models,» *Ecography*, vol. 43, n° 4, pp. 549-558, 2020.
- [46] S. Thakuri, P. Shrestha, M. Deuba, P. Shah, O. Bhandari y S. Shrestha, «Potential habitat modeling of water hyacinth in lakes of nepal using maxent algorithm,» *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Science*, vol. 4, n° 2, pp. 103-110, 2019.
- [47] F. Shabani y L. Kumar, «Risk Levels of Invasive *Fusarium oxysporum* f. sp. in Areas Suitable for Date Palm (*Phoenix dactylifera*) Cultivation under Various Climate Change Projections,» *PLOs ONE*, vol. 8, n° 12, pp. 1-13, 2013.
- [48] J. Williams, C. Seo, J. Thorne, J. Nelson y S. Erwin, «Using species distribution models to predict new occurrences for rare plants,» *Diversity and Distributions*, vol. 15, n° 4, pp. 565-576, 2009.
- [49] F. Shabani, L. Kumar y M. Ahmadi, «A comparison of absolute performance of differentcorrelative and mechanistic species distribution models inan independent area,» *Ecology and Evolution*, vol. 6, n° 16, pp. 5973-5986, 2016.
- [50] M. Ahmadi, M. Hemamii, M. Mahmoud y F. Shabani, «MaxEnt brings comparable results when the input data are MaxEnt brings comparable results when the input data are distribution models,» *Ecology and Evolution*, vol. 13, n° 2, pp. 1-13, 2023.
- [51] N. Z. Ab Lah, Z. Yusop, M. Hashim y J. Mohd, «Predicting the Habitat Suitability of *Melaleuca cajuputi* Based on the MaxEnt Species Distribution Model,» *Forests*, vol. 12, n° 1449, pp. 1-18, 2021.
- [52] R. Wunderlich, Y.-P. Lin, J. Anthony y J. Petway , «Two alternative evaluation metrics to replace the true skill statistic in the assessment of species distribution models,» *Nature Conservation*, vol. 35, p. 97–116, 2019.
- [53] K. Konowalik y A. Nosol, «Evaluation metrics and validation of presence-only species distribution models based on distributional maps with varying coverage,» *Scientific Reports*, vol. 11, n° 1482 , pp. 1-15, 2021.

- [54] T. Linders, U. Schaffner, R. Eschen, A. Abebe, C. Simon y L. Nigatu, «Direct and indirect effects of invasive species: Biodiversity loss is a major mechanism by which an invasive tree affects ecosystem functioning,» vol. 107, nº 6, pp. 2660-2672, 2019.
- [55] G. Mollet , J. Pantel y T. Romanuk, «The Effects of Invasive Species on the Decline in Species Richness: A Global Meta-Analysis,» *Advances in Ecological Research*, vol. 56, pp. 61-83, 2017.
- [56] C. J. A. Bradshaw, A. J. Hoskins, P. J. Haubrock, R. N. Cuthbert y C. Diagne, «Detailed assessment of the reported economic costs of invasive species in Australia,» *NeoBiota*, vol. 67, pp. 511 - 550, 2021.
- [57] P. Courtois, C. Figueires, C. Mulier y J. Weill, «A Cost–Benefit Approach for Prioritizing Invasive Species,» *Ecological Economics*, vol. 146, pp. 607-620, 2018.
- [58] Instituto Nacional de Estadística y Censos, «Boletín Técnico: Encuesta de Superficie y Producción Agropecuaria Continua, 2020,» Quito, 2021.
- [59] R. J. Hijmans, S. E. Cameron, J. L. Paraa, P. G. Jones y A. Jarvis, «Very high resolution interpolated climate surfaces for global land areas,» *International Journal of Climatology*, vol. 25, nº 15, 30 November 2005.
- [60] S. Ranjitkar , N. M. Sujakhu , Y. Lu , Q. Wang , M. Wang y J. He , «Climate modelling for agroforestry species selection in Yunnan Province, China,» *Environmental Modelling & Software*, vol. 75, pp. 263-272, 1 January 2016.
- [61] R. A. Baquero, A. Márcia Barbosa, D. Ayllón, C. Guerra, E. Sánchez, M. B. Araújo y G. G. Nicola, «Potential distributions of invasive vertebrates in the Iberian Peninsula under projected changes in climate extreme events,» *Diversity and Distributions*, vol. 27, pp. 2262-2276, 22 August 2021.
- [62] T. Hastie, «Generalized Additive Models,» 2020.
- [63] A. Liaw y M. Wiener, «Classification and Regression by randomForest,» *R News*, vol. 2, nº 3, pp. 18-22, December 2002.
- [64] R. Valavi, J. Elith, J. J. Lahoz-Monfort y G. Guillera-Arroita, «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*, vol. 10, pp. 225-232, February 2019.
- [65] M. B. Araújo y R. G. Pearson, «Equilibrium of species’ distributions with climate,»

- Ecography*, vol. 28, nº 5, pp. 693-695, October 2005.
- [66] A. H. Fielding y J. F. Bell, «. A review of methods for the assessment of prediction errors in conservation presence/absence models,» *Environmental Conservation*, vol. 24, nº 1, pp. 38-49, March 1997.
- [67] A. Marcia Barbosa, R. Real, A. Román Muñoz y J. A. Brown, «New measures for assessing model equilibrium and prediction mismatch in species distribution models.New measures for assessing model equilibrium and prediction mismatch in species distribution models.,» *Diversity and Distributions*, vol. 19, nº 10, pp. 1333-1338, 1 October 2013.
- [68] T. A. Lake, R. D. Briscoe Runquist y D. A. Moeller, «Predicting range expansion of invasive species: Pitfalls and best practices for obtaining biologically realistic projections,» *Diversity and Distributions*, vol. 26, nº 12, pp. 1767-1779, 1 December 2020.
- [69] O. Allouche, A. Tsoar y R. Kadmon, «Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS),» *Journal of Applied Ecology*, vol. 43, nº 6, pp. 1223-1232, December 2006.
- [70] M. E. Miller, S. L. Hui y W. M. Tierney, «Validation techniques for logistic regression models,» *Statistics in Medicine*, vol. 10, nº 8, pp. 1213-1226, August 1991.
- [71] J. Pearce y S. Ferrier, «Evaluating the predictive performance of habitat models developed using logistic regression,» *Ecological Modelling*, vol. 133, nº 3, pp. 225-245, 3 September 2000.
- [72] G. Rapacciuolo¹, D. B. Roy, S. Gillings, R. Fox, K. Walker y A. Purvis, «Climatic Associations of British Species Distributions Show Good Transferability in Time but Low Predictive Accuracy for Range Change,» *PLoS ONE*, vol. 7, nº 7, pp. 1-11, 5 July 2012.
- [73] M. B. Araújo, R. P. Anderson, A. Márcia Barbosa, C. M. Beale, C. F. Dormann, R. Early, R. A. Garcia, A. Guisan, . L. Maiorano, B. Naimi, R. B. O’Har, N. E. Zimmerman y C. Rahbek, «Standards for distribution models in biodiversity assessments,» *Science Advances*, vol. 5, nº 1, pp. 1-11, 16 January 2019.
- [74] R. Real, A. Márcia Barbosa y J. M. Vargas, «Obtaining Environmental Favourability Functions from Logistic Regression,» *Environmental and Ecological*

- Statistics*, vol. 13, p. 237–245, June 2006.
- [75] R. Real, A. M. Barbosa y J. W. Bull, «Species Distributions, Quantum Theory, and the Enhancement of Biodiversity Measures,» *Systematic Biology*, vol. 66, nº 3, p. 453–462, May 2017.
- [76] M. B. Araújo, R. G. Pearson, W. Thuiller y M. Erhard, «Validation of species–climate impact models under climate change,» *Global Change Biology*, vol. 11, nº 9, pp. 1504-1513, September 2005.
- [77] J. Soberón, L. Osorio-Olvera y T. Peterson, «Diferencias conceptuales entre modelación de nichos y modelación de áreas de distribución,» *Revista Mexicana de Biodiversidad*, vol. 88, nº 2, Junio 2017.

CAPÍTULO VII

ANEXOS

Anexo 1. Proceso de Limpieza de Datos en RStudio con el Paquete “scrubr”.

```
80
81
82 # let's do some further data cleaning with functions of the 'scrubr' package (but note this cleaning
83
84 nrow(presences)
85 presences <- coord_incomplete(coord_imprecise(coord_impossible(coord_unlikely(presences))))
86 nrow(presences)
87
88 presences <- coord_uncertain(presences, coorduncertaintyLimit = 7071) # 7071 = distance from the ce
89 nrow(presences)
90
91 # map the cleaned occurrence records on top of the raw ones:
92
93
85:70 GET SPECIES OCCURRENCE DATA DATA : R Script
```

Anexo 2. Proceso de Cálculo y Trazado en RStudio para el Área Bajo la Curva ROC.

```
35
36
37 # EVALUATE MODELS ON THE TRAINING DATA ####
38
39 names(dat)
40
41 # calculate and plot area under the ROC curve:
42 par(mfrow = c(3, 2), mar = c(3, 2, 2, 1)) # set plot margins and 3x2 plots in same window
43 ?AUC
44 with(dat, AUC(obs = presence, pred = GLM_P, main = "GLM"))
45 with(dat, AUC(obs = presence, pred = GAM_P, main = "GAM"))
46 with(dat, AUC(obs = presence, pred = MX_P, main = "MX"))
47 with(dat, AUC(obs = presence, pred = RF_P, main = "RF"))
48 with(dat, AUC(obs = presence, pred = GBM_P, main = "GBM"))
49 with(dat, AUC(obs = presence, pred = BART_P, main = "BART"))
50
51 # calculate and plot threshold-based classification metrics:
52 ?threshMeasures
53
54
46:57 EVALUATE MODELS ON THE TRAINING DATA : R Script
```

Anexo 3. Proceso de Cálculo y Trazado en RStudio para Métricas de Clasificación Basadas en Umbrales.

```
50
51 # calculate and plot threshold-based classification metrics:
52 ?threshMeasures
53
54 classif_metrics <- c("CCR", "Sensitivity", "Specificity", "Precision", "Recall", "TSS", "kappa")
55
56 my_thresh <- "preval" # to use training prevalence as the threshold
57
58 # note you can also optimize the threshold given a specific metric and model prediction:
59 # my_thresh <- with(dat, optithresh(obs = presence, pred = BART_P, interval = 0.001, measures = "TSS", optimize = "each"))
60 # my_thresh
61 # my_thresh <- my_thresh$optimals[, each$threshold]
62
63 par(mfrow = c(3, 2), mar = c(5, 2, 2, 1))
64 with(dat, threshMeasures(obs = presence, pred = GLM_P, thresh = my_thresh, measures = classif_metrics, ylim = c(0, 1), main = "GLM"))
65 with(dat, threshMeasures(obs = presence, pred = GAM_P, thresh = my_thresh, measures = classif_metrics, ylim = c(0, 1), main = "GAM"))
66 with(dat, threshMeasures(obs = presence, pred = MX_P, thresh = my_thresh, measures = classif_metrics, ylim = c(0, 1), main = "MX"))
67 with(dat, threshMeasures(obs = presence, pred = RF_P, thresh = my_thresh, measures = classif_metrics, ylim = c(0, 1), main = "RF"))
68 with(dat, threshMeasures(obs = presence, pred = GBM_P, thresh = my_thresh, measures = classif_metrics, ylim = c(0, 1), main = "GBM"))
69 with(dat, threshMeasures(obs = presence, pred = BART_P, thresh = my_thresh, measures = classif_metrics, ylim = c(0, 1), main = "BART"))
70
71 # calculate and plot Miller calibration line:
72 par(mfrow = c(3, 2), mar = c(3, 2, 2, 1))
73
58:89 EVALUATE MODELS ON THE TRAINING DATA : R Script
```

Anexo 4. Proceso de Cálculo y Trazado en RStudio para la Línea de Calibración de Miller

```
72  
73  
74  
75  
76 # calculate and plot Miller calibration line:  
77 par(mfrow = c(3, 2), mar = c(3, 2, 2, 1))  
78 with(dat, MillerCalib(obs = presence, pred = GLM_P, main = "GLM"))  
79 with(dat, MillerCalib(obs = presence, pred = GAM_P, main = "GAM"))  
80 with(dat, MillerCalib(obs = presence, pred = MX_P, main = "MX"))  
81 with(dat, MillerCalib(obs = presence, pred = RF_P, main = "RF"))  
82 with(dat, MillerCalib(obs = presence, pred = GBM_P, main = "GBM"))  
83 with(dat, MillerCalib(obs = presence, pred = BART_P, main = "BART"))  
84  
85  
86 # BUT THIS JUST EVALUATES HOW THE MODELS FIT THE SAME DATA ON WHICH THEY WERE TRAINED  
87 # YOU CAN SET ASIDE SOME DATA TO LEAVE OUT OF MODEL TRAINING AND USE FOR TESTING OUT  
88 # BLOCK CROSS VALIDATION (below) IS CURRENTLY THE MOST APPROPRIATE METHOD  
89  
75:1 # EVALUATE MODELS ON THE TRAINING DATA
```

Anexo 5. Código en RStudio: Conversión de probabilidad a favorabilidad.

```
208 boxplot(crossval, col = rep(1:length(measures), each = length(names(models))), las = 2)  
209 abline(h = 1, col = "darkgrey", lty = 2) # remember Miller calibration slope (MCS) should ide  
210  
211  
212 # CONVERT PROBABILITY TO FAVOURABILITY ###  
213  
214 # after validating the probability predictions, you can convert them to prevalence-independent  
215 # presence probability (given by presence/absence models) is necessarily bound by the general  
216 # you can remove this effect of prevalence (and make predictions directly comparable) with the  
217  
218 ?Fav  
219 dat$GLM_Fav <- Fav(pred = dat$GLM_P, sample.preval = prevalence(model = models[["GLM"]]))  
220 dat$GAM_Fav <- Fav(pred = dat$GAM_P, sample.preval = prevalence(model = models[["GAM"]]))  
221 # (Maxent left out because it predicts suitability, not presence probability)  
222 dat$RF_Fav <- Fav(pred = dat$RF_P, sample.preval = prevalence(model = models[["RF"]]))  
223 dat$GBM_Fav <- Fav(pred = dat$GBM_P, sample.preval = prevalence(model = models[["GBM"]]))  
224 dat$BART_Fav <- Fav(pred = dat$BART_P, sample.preval = prevalence(model = models[["BART"]]))  
225 head(dat)  
226  
227  
228  
211:1 # EVALUATE EACH MODEL ON ITS VALIDATION FOLD
```

Anexo 6. Código en RStudio: mapear la favorabilidad.

```

227 # map favourability (and Maxent suitability) from the table:
228
229 names(dat)
230 dat_spatial <- vect(dat, geom = c("x", "y"))
231
232 plot(dat_spatial, c("GLM_Fav", "GAM_Fav", "MX_P", "RF_Fav", "GBM_Fav", "BART_Fav"), type = "continuous", col = hcl.colors(100))
233
234 # or you can plot each column in turn:
235 plot(dat_spatial, "GLM_Fav", type = "continuous", col = hcl.colors(100), cex = 0.8)
236 plot(dat_spatial, "GAM_Fav", type = "continuous", col = hcl.colors(100), cex = 0.8)
237 plot(dat_spatial, "MX_P", type = "continuous", col = hcl.colors(100), cex = 0.8)
238 plot(dat_spatial, "RF_Fav", type = "continuous", col = hcl.colors(100), cex = 0.8)
239 plot(dat_spatial, "GBM_Fav", type = "continuous", col = hcl.colors(100), cex = 0.8)
240 plot(dat_spatial, "BART_Fav", type = "continuous", col = hcl.colors(100), cex = 0.8)
241
242
243 # you can also calculate favourability directly on the raster maps of probability predictions:
244
245 GLM_Fav <- fav(ford, mod=raster[["GLM"]], geom=c("x", "y"), scale=1, model=f[["GLM"]])
246
247
239:83 CONVERT PROBABILITY TO FAVOURABILITY :

```

Anexo 7. Generación de mapas binarios presencia-ausencia mediante la herramienta "reclassify".

