

COMPARACIÓN Y ANÁLISIS ESTADÍSTICO ENTRE MODELOS DE PROBABILIDAD DE OCURRENCIA DE INCENDIOS FORESTALES PARA HONDURAS. AÑO 2019

José Cáceres Coello^{1,*}

¹Departamento de Ciencia y Tecnologías de la Información Geográfica
Universidad Nacional Autónoma de Honduras

Recibido: 4/marzo/2023

Aceptado: 6/junio/2023

DOI: <https://doi.org/10.5377/ce.v14i2.16918>

RESUMEN

Honduras por sus condiciones climáticas y fisiográficas en los últimos años ha sufrido impactos negativos a causa que los incendios forestales que ocasionan pérdidas en la economía del país, como el cierre de aeropuertos, en la salud de las poblaciones urbanas y rurales, incidencia de plagas y enfermedades en agricultura y el bosque, mayor degradación y erosión de suelos, alteración de sistemas de recarga hídrica, efectos negativos sobre la biodiversidad así como la pérdida de valores escénicos. Para el 2019 se reportaron 1 177 incendios con una superficie afectada de 72 434.77 hectáreas. Los departamentos con mayor incidencia de incendios forestales fueron; Francisco Morazán, Olancho, y Copán. En esta investigación se analizó estadísticamente el desempeño de tres modelos (ICF, Random Forest y MaxEnt) para la generación de mapas de susceptibilidad a incendios forestales a partir de variables ambientales. Los modelos se calibraron con los datos de los reportes de incendios generados por el ICF. Adicionalmente se seleccionaron puntos de incendios de manera aleatoria obtenidos del Sistema FIRMS. El modelo Random Forest tuvo el mejor rendimiento con un AUC para los datos de prueba de 0.973 y un AUC de datos MODIS-FIRMS de 0.919. Usando las clasificaciones promedio de las medidas de importancia de la variable ambiental, la cercanía a asentamientos humanos fue el mejor predictor de igniciones de incendios forestales, seguido de cerca por la distancia a carreteras no pavimentadas y la elevación, para los modelos basados en algoritmos de aprendizaje autónomo.

Palabras clave: Incendios Forestales; MaxEnt; Random Forest; AUC; FIRMS.

*jcaceres@unah.edu.hn,  <https://orcid.org/0000-0002-6055-7690>

ABSTRACT

Honduras, due to its climatic and physiographic conditions in recent years, has suffered negative impacts due to wildfires that cause losses in the country's economy, such as the closure of airports, in the health of urban and rural populations, incidence of pests and diseases in agriculture and forestry, greater soil degradation and erosion, alteration of water recharge systems, negative effects on biodiversity as well as the loss of scenic values. In 2019, 1,177 fires were reported with an affected area of 72,434.77 hectares. The departments with the highest incidence of wildfires were; Francisco Morazán, Olancho, and Copan. In this research, the performance of three models (ICF, Random Forest and MaxEnt) for the generation of susceptibility maps to wildfires from environmental variables was statistically analyzed. The models were calibrated with the fire report data generated by the ICF. Additionally, fire points were selected randomly obtained from the FIRMS System. The Random Forest model had the best performance with an AUC for test data of 0.973 and an AUC for MODIS-FIRMS data of 0.919. Using the average rankings of the environmental variable importance measures, proximity to human settlements was the best predictor of wildfire ignitions, closely followed by distance to unpaved roads and elevation, for models based on Machine Learning Algorithms.

Keywords: Wildfires; MaxEnt; Random Forest; AUC; FIRMS.

1 Introducción

Honduras por sus condiciones climáticas y fisiográficas en los últimos años ha sufrido impactos negativos a causa que los incendios forestales que ocasionan pérdidas en la economía del país, como el cierre de aeropuertos, en la salud de las poblaciones urbanas y rurales, incidencia de plagas y enfermedades en agricultura y el bosque, mayor degradación y erosión de suelos, alteración de sistemas de recarga hídrica, efectos negativos sobre la biodiversidad así como la pérdida de valores escénicos ([Instituto de Conservación Forestal, 2014](#)).

Históricamente se han realizado estudios de las causas que originan la gran destrucción de los bosques debido a los incendios forestales en todas las regiones del país; logrando identificar de esta manera las coberturas más propensas a los mismos (mapa forestal y cobertura de la tierra de Honduras con imágenes rapidEye), la amenaza (Población), la vulnerabilidad de los bosques a causa de su accesibilidad, (camino y pendientes), etc. ([Instituto de Conservación Forestal, 2014](#); [Moreno Segura et al., 2017](#)).

Con estos elementos la Unidad Nacional de Monitoreo Forestal del Instituto Nacional de Conservación y Desarrollo Forestal, Áreas Protegidas y Vida Silvestre (ICF), ha elaborado mapas de riesgo a incendios a escala nacional en los últimos 3 años utilizando la metodología de ICF que consiste en una matriz que toma como base la amenaza, vulnerabilidad y propensión para construir categorías de riesgo a incendios forestales que permite la identificación y cuantificación de las zonas de riesgo tanto a nivel nacional, regional como municipal.

Esta herramienta ha permitido a la Comité Nacional de Protección Forestal (CONAPROFOR) elaborar planes con información actualizada en los diferentes niveles de planificación con datos reales que hacen más eficiente la aplicación de las diferentes actividades y permiten coordinar las acciones de participación e integración de esfuerzos gubernamentales y de la sociedad civil en una forma planificada y coordinada, que contribuyan al desarrollo de actividades que fortalezcan las capacidades políticas, técnicas y operativas, orientadas a reducir los efectos negativos que originan los incendios a los ecosistemas la salud y la economía nacional.

A pesar de contar con esta metodología se han desarrollado dos metodologías más para la caracterización del riesgo a incendios forestales, por otros proyectos que apoyan al ICF (Máxima Entropía y Random Forest desarrolladas por China-Taiwán). En tal sentido, se busca comparar la eficiencia de cada una de estas 3 herramientas para la definición de riesgos a incendios y así aplicar la herramienta más adecuada en la toma de decisiones de la institución en esta temática.

De las diferentes amenazas a las que se encuentran expuestos los bosques, los incendios forestales es una de las principales que se presentan cada año. Estos siniestros ocurren principalmente en bosques de pino y mixto (pino y roble) que representan el 34.70 % de la cobertura boscosa 2.19 millones de hectáreas (Cáceres Coello, 2020).

Para el 2019 se reportaron 1 177 incendios con una superficie afectada de 72 434.77 hectáreas. Los departamentos con mayor incidencia de incendios forestales fueron; Francisco Morazán con 268, Olancho con 213 y Copán con 132 incendios; estos representan 52.08 % de los incendios ocurridos a nivel nacional. El departamento con la mayor cantidad de área afectada por incendios forestales es Gracias a Dios con 24 372.93 hectáreas, seguido por Olancho con 15 936.62 hectáreas (Instituto de Conservación Forestal, 2020).

En esta investigación se analizó estadísticamente el desempeño de tres modelos para la generación de mapas de susceptibilidad a incendios forestales.

2 Marco Teórico

Los incendios forestales pueden ser originados por causas humanas o naturales. Ambos tipos de igniciones son esencialmente procesos no aleatorios que dependen de impulsores de arriba hacia abajo y de abajo hacia arriba y tienen patrones espaciales explícitos (Krawchuk *et al.*, 2006; Narayanaraj y Wimberly, 2011; Parisien y Moritz, 2009). Los controladores descendentes reflejan la variabilidad espacial a gran escala del tiempo y el clima, que a su vez afectan directamente las igniciones naturales (p. Ej., Rayos) y también cambian la humedad del combustible, lo que cambia la probabilidad de una ignición (Latham y Williams, 2001). Variables de escala local que afectan la combustibilidad del combustible y la proximidad a fuentes de ignición humanas (Krawchuk *et al.*, 2006; Sturtevant y Cleland, 2007; Syphard *et al.*, 2007; Syphard *et al.*, 2008).

El enfoque más común para comprender los patrones espaciales de ignición de incendios forestales y sus fuerzas impulsoras es el modelado de distribución basado en ubicaciones históricas de ignición (Bar Massada *et al.*, 2011; Sturtevant y Cleland, 2007; Syphard *et al.*, 2008). Algunos estudios consideraron únicamente igniciones antropogénicas (Syphard *et al.*, 2008; Yang *et al.*, 2007) o por rayos (Díaz-Avalos *et al.*, 2001; Krawchuk *et al.*, 2006), mientras que otros examinaron ambos (Parisien y Moritz, 2009; Sturtevant y Cleland, 2007). El hallazgo general es que las igniciones humanas y naturales de incendios forestales tienen patrones espaciales distintivos que pueden cuantificarse fácilmente utilizando variables explicativas humanas y biofísicas. Sin embargo, es importante distinguir los dos objetivos del modelado de ignición: (1) modelado explicativo para probar hipótesis sobre el papel que juegan diferentes factores en causar igniciones, (2) modelado predictivo para identificar aquellas áreas que son más propensas a igniciones, y donde se pueden apuntar los tratamientos de prevención de incendios o reducción de combustible (Bar Massada *et al.*, 2013).

Conceptual y metodológicamente, el modelado de la distribución de la ignición está estrechamente relacionado con el modelado de la distribución de especies (MDS) (Franklin, 2010). El enfoque básico es

analizar las ubicaciones de ignición de incendios forestales (análogas a las ubicaciones de ocurrencia de especies) en relación con las variables ambientales que se hipotetizan para influir en la distribución espacial de igniciones (o ocurrencias de especies). Los modelos estiman la respuesta de incendios forestales (o especies) a estas variables predictoras ambientales. Al igual que en MDS, hay dos tipos de datos de ignición que pueden analizarse: presencia (la ocurrencia de un evento de ignición en un punto en el espacio) o abundancia (el número de igniciones por unidad de área). El tipo de datos afecta la elección del tipo de modelo porque los datos de presencia generalmente requieren una respuesta binomial, mientras que los datos de abundancia requieren una respuesta continua. Además, aunque los datos de presencia a menudo van acompañados de datos de ausencia, también se han desarrollado métodos especiales de modelado para datos de solo presencia, donde las ubicaciones de presencia se comparan con las condiciones ambientales de fondo ('usado versus disponible') (Elith *et al.*, 2006; Franklin, 2010), porque a menudo es imposible identificar lugares donde no puede ocurrir ignición.

Se ha utilizado un conjunto diverso de modelos para modelar igniciones (o apariciones de especies), incluidos métodos estadísticos como modelos lineales generalizados (Nelder y Wedderburn, 1972), modelos aditivos generales (Hastie y Tibshirani, 1990) y algoritmos de aprendizaje automático como Random Forest (Prasad *et al.*, 2006) y Máxima entropía (Maxent) (Phillips *et al.*, 2006). Dada la amplia gama de tipos de modelos, la pregunta es ¿hasta qué punto la elección del modelo afecta los resultados? ¿Algunos tipos de modelos son consistentemente mejores para explicar o predecir los incendios forestales? En los estudios de vida silvestre, el tipo de modelo afecta fuertemente el rendimiento y los mapas resultantes de los MDS (Elith *et al.*, 2006; Elith y Graham, 2009; Guisan *et al.*, 2007), y por lo tanto, planteamos la hipótesis de que los modelos de ignición también se verían afectados por la elección del modelo..

3 Metodología

3.1 Modelos estadísticos

Para la predicción de la probabilidad de ocurrencia de incendios forestales se utilizaron tres tipos de modelos de distribución de incendios: un modelo paramétrico (ICF) y dos algoritmos de aprendizaje automático (Random Forest y MaxEnt).

En el caso del modelo de ICF (Figura 1), se utilizó una ponderación lineal combinada para construir un mapa de riesgo a Incendios Forestales definido por un modelo de propensión. Este modelo considera las siguientes variables: recurrencia de incendios forestales, uso del suelo (cobertura forestal), sequía meteorológica, infraestructura técnica, productiva y asentamientos humanos, áreas de interés especial (áreas protegidas, zonas productoras de agua, etc.), pendiente y accesibilidad; asignando una valoración para cada categoría (ICF, 2014).

El modelo Random Forest es una técnica de conjunto que amplía el enfoque de modelado de árboles de clasificación (Breiman, 2001) promediando las predicciones de muchos árboles de clasificación individuales, cada uno desarrollado utilizando un subconjunto de los datos de entrenamiento. Este algoritmo supera el problema de la inestabilidad en el uso de árboles de clasificación única, lo que resulta en una mayor precisión de predicción (Franklin, 2010; Sypard y Franklin, 2010). El Random Forest estima el error del modelo y la importancia de la variable al comparar la predicción de cada árbol con un conjunto de datos de prueba que consiste en todas las observaciones que se retrasaron durante su desarrollo (muestras descartadas). La importancia de variable en Random Forest se calcula calculando el error medio cuadrático (RMSE) de cada árbol comparando su predicción basada en las observaciones del subconjunto con

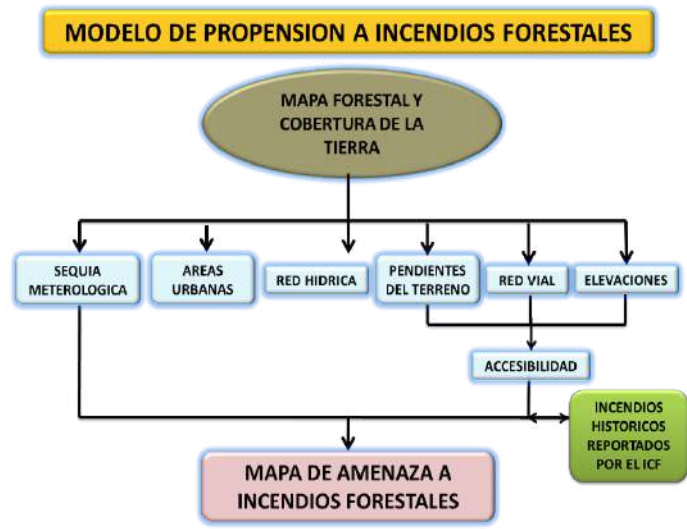


Figura 1: Mapa de Propensión a Incendios Forestales. Metodología ICF.

las basadas en las observaciones *out-of-bag*. Para cada variable, las observaciones de la muestra *out-of-bag* se permutan aleatoriamente, y el RMSE se vuelve a calcular. Para cualquier variable permutada, la diferencia en RMSE para cada árbol (generalmente un aumento), promediada en el bosque y normalizada por la desviación estándar de las diferencias, es una medida de su importancia, con diferencias mayores medidas como incremento porcentual en RMSE que indica más importancia (Cutler *et al.*, 2007). Para la generación del modelo de incendios, utilizando el principio de máxima clasificador Random Forest, se tomaron en consideración trece variables (Figura 2) y se generó el modelo Random Forest utilizando las herramientas desarrolladas por el Marine Geospatial Ecology Tools (Roberts *et al.*, 2010).

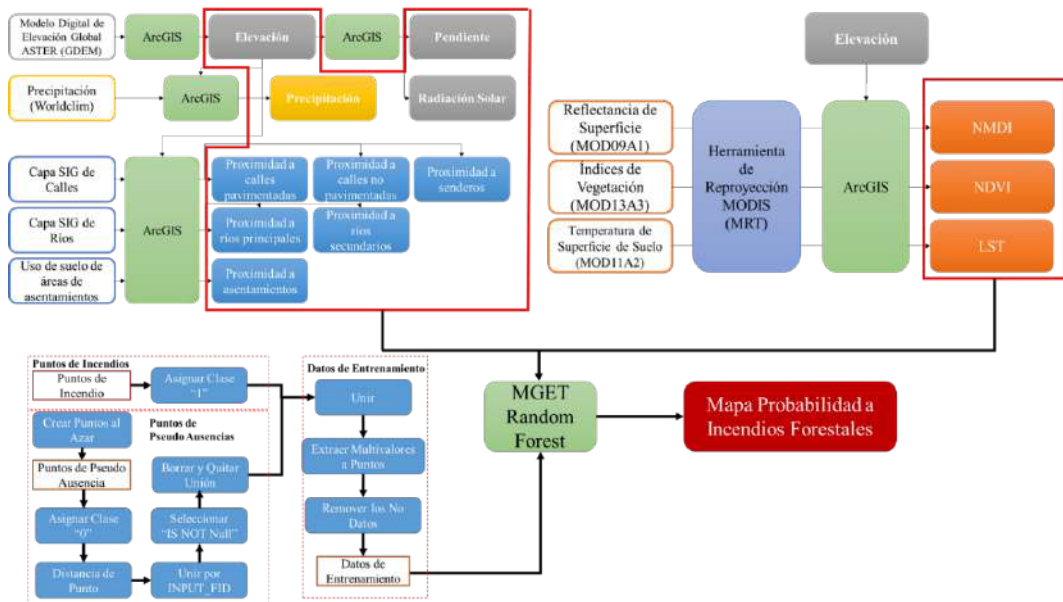


Figura 2: Mapa de Propensión a Incendios Forestales. Metodología Random Forest.

Maximum Entropy es un algoritmo de aprendizaje automático de presencia única que contrasta iterativamente valores de predictores ambientales en ubicaciones de ocurrencia con los de una gran muestra

de antecedentes tomada a lo largo del área de estudio (Elith *et al.*, 2011; Phillips *et al.*, 2006). Maxent se ha utilizado para modelar la ignición de fuego en los EE. UU. (Parisien y Moritz, 2009) y la aparición de incendios en las montañas Ghats de la India (Renard *et al.*, 2012). Maxent representa Π , la distribución de las ubicaciones de ignición para un conjunto x de sitios en el área de estudio (Phillips y Dudík, 2008). Cada sitio x se asigna a un valor no negativo $\Pi(x)$, y estos valores en todos los sitios suman uno. Maxent luego genera un modelo de Π basado en las restricciones ambientales obtenidas a partir de los datos de ocurrencia, donde cada restricción se define como una función de una variable ambiental. El modelo requiere que el valor promedio de cada una de estas funciones esté dentro de un pequeño margen de error del promedio empírico de las variables ambientales correspondientes en todos los sitios de muestra. Entre la gran cantidad de posibles distribuciones de probabilidad, Maxent selecciona la que tiene la entropía máxima (o la más uniforme) como el mejor representante de los datos (Phillips *et al.*, 2006) y asigna una probabilidad de ocurrencia a cada ubicación. Entre los modelos de distribución de especies de solo presencia, Maxent demuestra consistentemente una mayor precisión de predicción, especialmente con tamaños de muestra pequeños (Elith *et al.*, 2006; Pearson *et al.*, 2006). La importancia de variable se cuantifica en función del aumento de la ganancia de entrenamiento regularizada mediante las iteraciones de desarrollo del modelo, donde la ganancia de entrenamiento denota el aumento en la probabilidad de ignición en los lugares de entrenamiento (es decir, variables que contribuyen a aumentar las posibilidades de ignición se consideran más importantes). Para la generación del modelo de incendios, utilizando el principio de máxima entropía, se tomaron en consideración trece variables (Figura 3) y se utilizó el software autónomo Maxent, disponible en: https://biodiversityinformatics.amnh.org/open_source/maxent/.

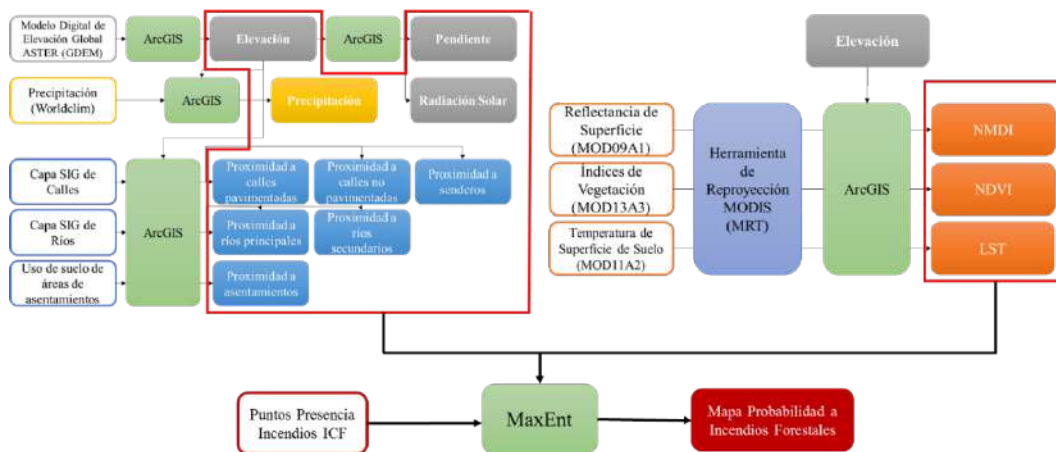


Figura 3: Mapa de Propensión a Incendios Forestales. Metodología MaxEnt

3.2 Comparación y evaluación de los modelos

Para comparar la precisión de predicción entre los tres modelos, se calculó el área bajo la curva (AUC) de la representación de la característica operativa del receptor (ROC) (Hanley y McNeil, 1982). El gráfico ROC (Figura 4) muestra la relación entre la tasa de error falso positivo (1-especificidad, donde la especificidad es la proporción de no igniciones que predijo correctamente) y la tasa verdadero-positiva (sensibilidad o la proporción de igniciones predichas correctamente) para cada umbral valor a la probabilidad de presencia predicha por el modelo, cuanto mayor sea la separación de la curva ROC de la diagonal, mayor será el poder predictivo del modelo. Solo dos (Random Forest y MaxEnt) de los tres modelos evaluados predicen probabilidades continuas de ocurrencia entre cero y uno, por lo tanto, se debe establecer un umbral para determinar si la probabilidad denota presencia o ausencia. El AUC es una métrica independiente del

umbral porque evalúa el rendimiento de un modelo en todos los valores umbral posibles (Franklin, 2010). El AUC varía de 0.5 a 1, donde 0.5 es análogo a una predicción completamente aleatoria y 1 implica una predicción perfecta. Los valores de AUC entre 0.5 y 0.7 indican pobre, entre 0.7 y 0.9 indican moderada, y mayor que 0.9 denota un alto rendimiento del modelo (McCune y Grace, 2002). Aunque el AUC ha sido criticado (Lobo *et al.*, 2008), es el método estándar para evaluar la precisión de predicción debido a su independencia de umbral y la facilidad de interpretar sus resultados.

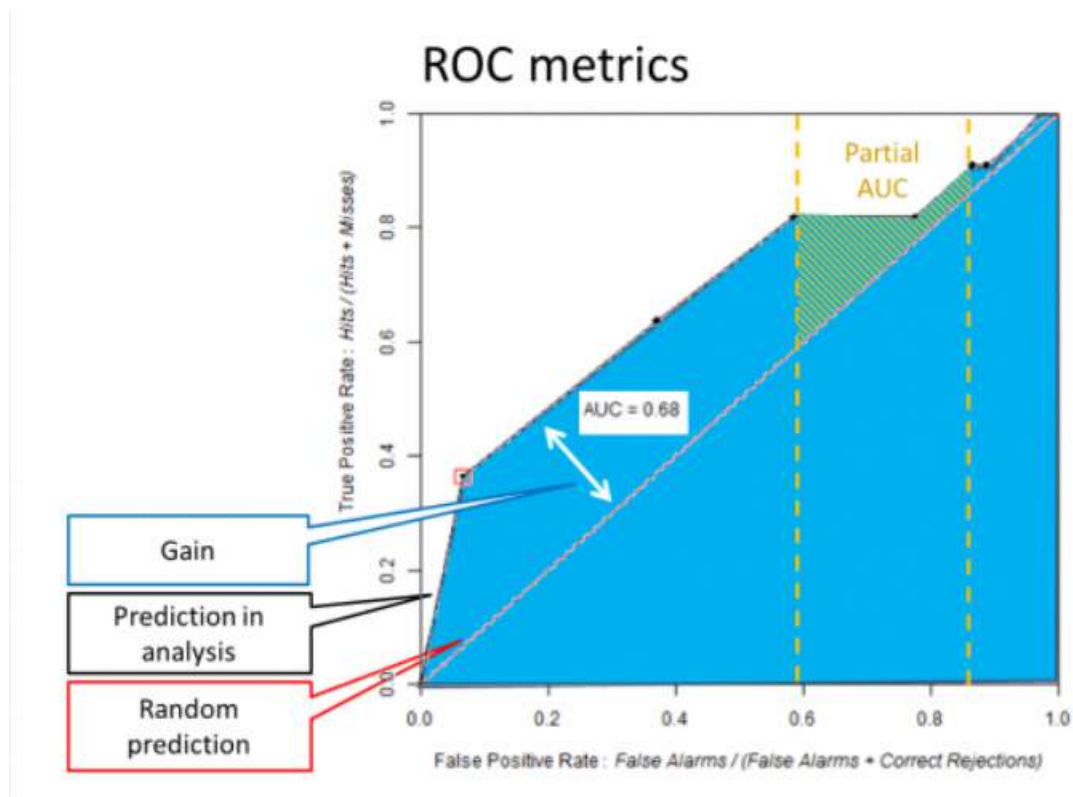


Figura 4: Curva ROC y métricas respectivas. Fuente: Tomado de Mas *et al.* (2013)

3.3 Evaluación de la importancia de la variable

Dadas las diferentes características de los modelos, se utilizó un estimador jackknife de importancia de variable basado en el cambio en AUC usando los datos de prueba. Esto produce resultados directamente comparables en dos de los tres tipos de modelos (MaxEnt y RF). El enfoque consiste en eliminar las variables de predicción del modelo completo de una a la vez, entrenar al modelo y calcular el AUC utilizando los datos de prueba. La diferencia entre el modelo completo y parcial (sin la variable) AUC indica la contribución de cada variable al modelo. Por lo tanto, representa la información proporcionada por una variable determinada que no está presente en otras variables. Además, se cuantificó el AUC del modelo usando una variable a la vez, se compararon los valores AUC de los modelos de variable única y se clasificó las variables en consecuencia.

3.4 Comparación de los mapas de predicción de la probabilidad de ignición

Para dos de los tres modelos, se generaron mapas de predicción de probabilidad de ignición basados en mapas de trama de todas las variables predictores. Estos mapas son equivalentes a los mapas de idoneidad

del hábitat en modelos de distribución de especies y representan la idoneidad de un píxel dado para la ignición, en relación con todos los demás píxeles. Los valores de píxel oscilan entre 0 (no apto) y 1 (perfectamente adecuado para el encendido). Debido a que los valores de píxel contenidos en el modelo derivado de la metodología de ICF representan una variable categórica, y no una probabilidad, se realizó una reclasificación de los valores entre los tres modelos (Tabla 1) para llevar los valores de los tres modelos a la misma escala. Luego se seleccionó una muestra aleatoria de 800 puntos de incendios recopilados por ICF para el año 2015, se extrajeron los valores de cada modelo para cada uno de esos puntos, y se realizó un análisis multivariado correlacional para los tres modelos.

Tabla 1: Reclasificación de valores entre los tres modelos

Modelo	Valor Original	Valor Nuevo
ICF	Nulo	1
	Bajo	2
	Medio	3
	Alto	4
MaxEnt	0-0.25	1
	0.25-0.5	2
	0.5-0.75	3
	0.75-1	4
Random Forest	0-0.25	1
	0.25-0.5	2
	0.5-0.75	3
	0.75-1	4

Por último, para evaluar las diferencias entre los modelos, se comparó visualmente los tres mapas y calculamos el coeficiente de correlación de Spearman entre cada par de mapas.

4 Resultados

4.1 Rendimiento de los enfoques de los modelos

El rendimiento predictivo basado en datos de prueba independientes fue adecuado para todos los dos modelos y similar entre los tipos de modelo.

Los modelos se calibraron con el 80 % de los datos de los reportes de incendios para el año 2015 generados por el ICF, por lo cual se utilizó el 20 % restante como datos de prueba para analizar el rendimiento predictivo de los modelos. Adicionalmente se seleccionaron puntos de incendios de manera aleatoria obtenidos del Sistema FIRMS (Kaufman *et al.*, 1998).

Los valores de AUC obtenidos para ambos modelos y para ambos datos de prueba se presentan en la Tabla 2.

Tabla 2: Valores de AUC obtenidos para ambos modelos

Modelo	AUC Test Data	AUC Modis Data
MaxEnt	0.883108601	0.867837585
RF	0.973117276	0.918905323

El modelo Random Forest tuvo el mejor rendimiento con un AUC para los datos de prueba de 0.973 y un AUC de datos MODIS-FIRMS de 0.919. La curva ROC de ambos modelos, y su comparación contra un modelo aleatorio, se presenta en la Figura 5.

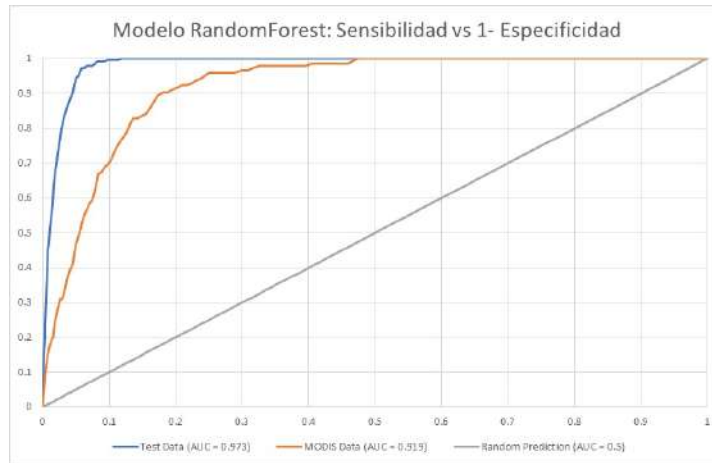


Figura 5: Curvas ROC del modelo Random Forest

El modelo MaxEnt Forest tuvo rendimiento moderado (McCune y Grace, 2002) con un AUC para los datos de prueba de 0.883 y un AUC de datos MODIS-FIRMS de 0.868. La curva ROC de ambos modelos, y su comparación contra un modelo aleatorio, se presenta en la Figura 6.

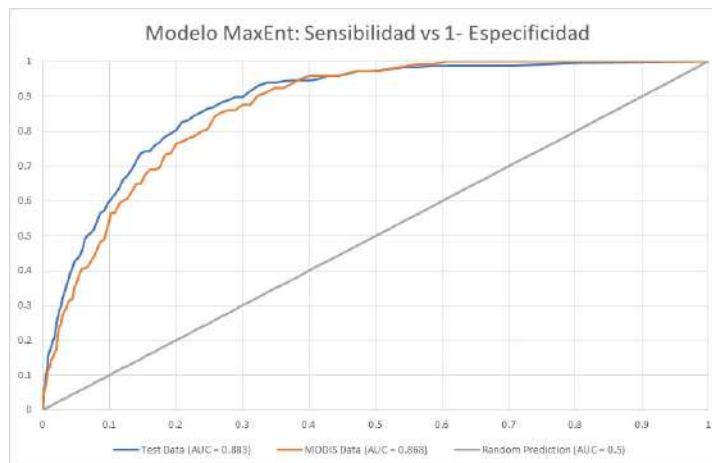


Figura 6: Curvas ROC del modelo MaxEnt

4.2 Importancia de las variables

Se realizó un análisis multivariado de las 13 variables ambientales de los modelos probabilísticos (RF y MaxEnt), y se generó una tabla que muestra el resumen estadístico para cada una de las variables seleccionadas (Tabla 3). Incluye medidas de tendencia central, de variabilidad, y de forma.

Luego se realizó una gráfica (Figura 7) que muestra todos los diagramas de dispersión entre cada par de las variables seleccionadas en el análisis. Cada par de variables se grafica dos veces, una con la primer variable en el eje-X, y otra con esa en el eje-Y. Por ejemplo, todos los diagramas en la primera fila tienen a DEM en el eje-Y. Todos los diagramas en la primera columna tienen a DEM en el eje-X. Este es el equivalente gráfico de una matriz de correlaciones y se utiliza para ayudar a determinar cuáles variables están más fuertemente relacionadas con otras. Las variables con mayor correlación (positiva o negativa) son el DEM con la pendiente y la temperatura de superficie.

En dos modelos (RF y MaxEnt), y en base a medidas de importancia nativa y de jackknife de importancia, la variable ambiental con mayor ganancia cuando se usa aisladamente es la cercanía a caminos no pavimentados, que por lo tanto parece tener la información más útil por sí misma. La variable ambiental que disminuye la ganancia más cuando se omite es la cercanía a centros poblados, que por lo tanto parece tener la mayor cantidad de información que no está presente en las otras variables. En la figura 8 las barras denotan el área bajo la curva de la curva característica del operador receptor (AUC de ROC). La barra roja representa el AUC de modelo completo, mientras que las barras azules indican el AUC de los modelos univariados para las variables correspondientes y las barras aqua indican el AUC de los modelos con todas las variables excepto la variable correspondiente. Se observa que la proximidad a caminos no pavimentados, proximidad a asentamientos humanos, radiación solar y NDVI son las variables más importantes para el modelo.

Tabla 3: Resumen estadístico para cada una de las variables seleccionadas.

	DEM	LST	NDVI	NMDI	Precipitacion	Prox Main_River	Prox Non_Paved	Prox Paved	Prox Sec_River	Prox Settlement	Prox Trail	Slope	Solar Radiation
	2901	2901	2901	2901	2901	2901	2901	2901	2901	2901	2901	2901	2901
Recuento	785.759	29.9189	0.610153	0.420565	40.4636	8327.23	5474.48	24385.5	563.815	8356.15	1829.44	5.94213	327135
Promedio	794	29.605	0.60385	0.409793	35	7158.91	1500	16807.7	500	5099.02	500	5.13297	327425
Mediana	470.592	4.16546	0.140824	0.0648751	16.8223	6110.69	9240.41	23615.4	796.015	10036.9	4194.25	4.42917	17747.9
Desviación Estándar	59.89 %	13.92 %	23.08 %	15.43 %	41.57 %	73.38 %	168.79 %	96.84 %	141.18 %	120.11 %	229.26 %	74.54 %	5.43 %
Coefficiente de Variación	0	0	-0.12605	0	0	0	0	0	0	0	0	0	0
Mínimo	2200	46.9667	0.9221	0.80719	130.5	61077.8	62211.3	118242	11423.7	68229.4	36962.8	30.247	379364
Máximo	2200	46.9667	1.04815	0.80719	130.5	61077.8	62211.3	118242	11423.7	68229.4	36962.8	30.247	379364
Rango													

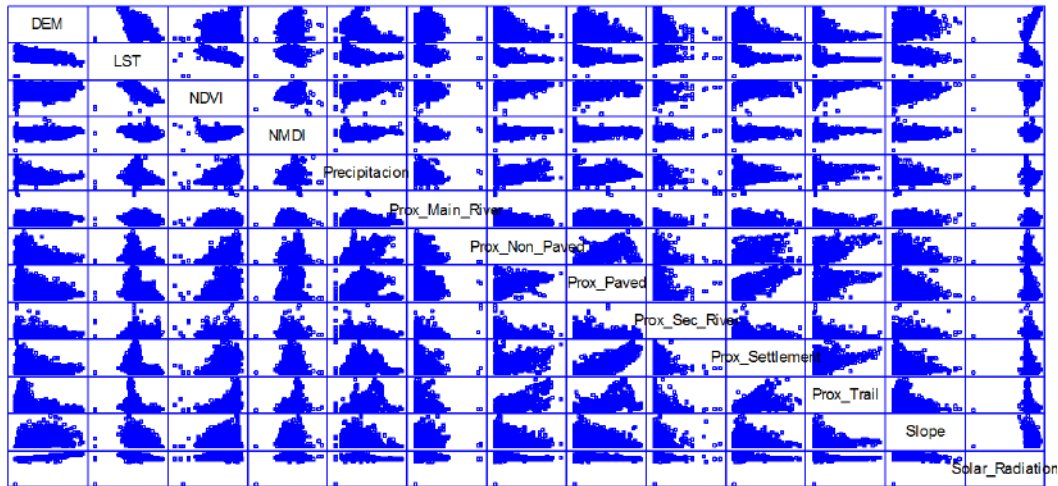


Figura 7: Diagramas de dispersión entre cada par de las variables seleccionadas.

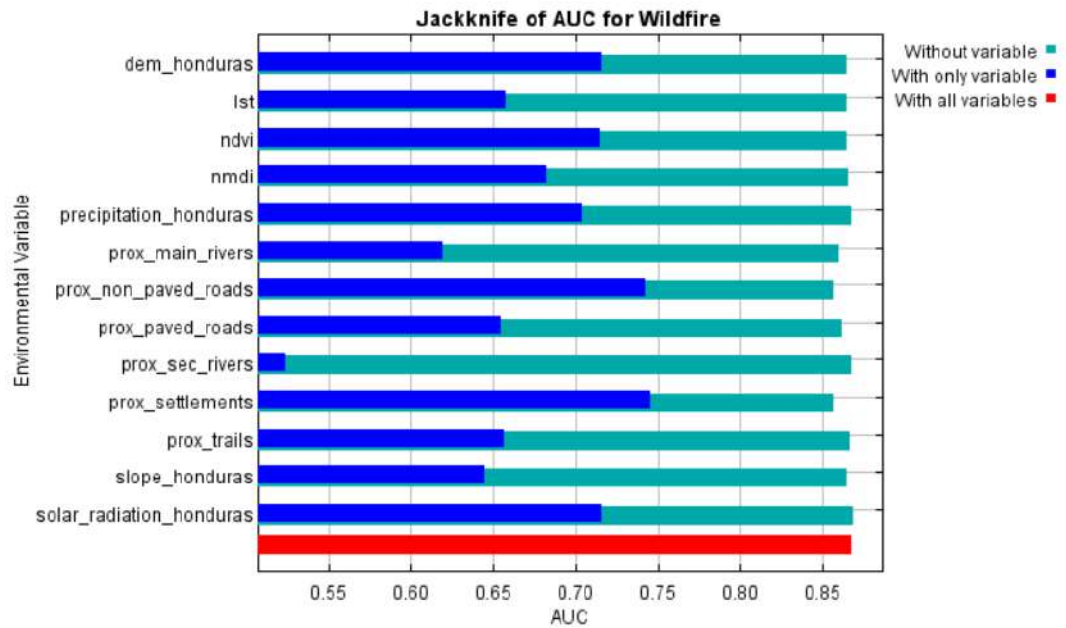


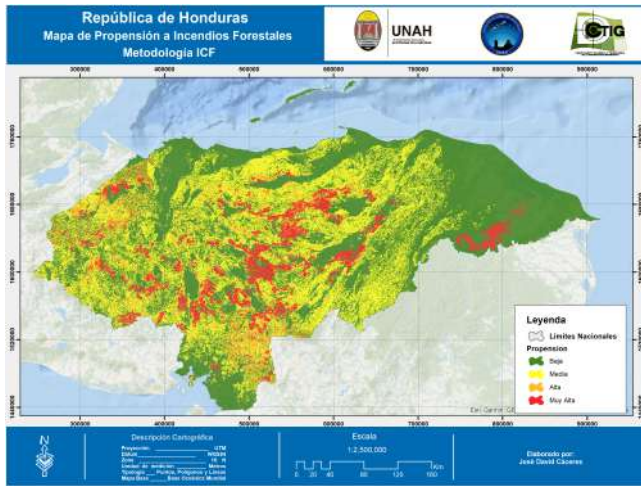
Figura 8: Estimaciones Jackknife de importancia de la variable para dos tipos de modelos de distribución de incendios.

Usando las clasificaciones promedio de las medidas de importancia de la variable ambiental, la cercanía a asentamientos humanos fue el mejor predictor de igniciones de incendios forestales, seguido de cerca por la distancia a carreteras no pavimentadas y la elevación, para los modelos basados en algoritmos de aprendizaje autónomo.

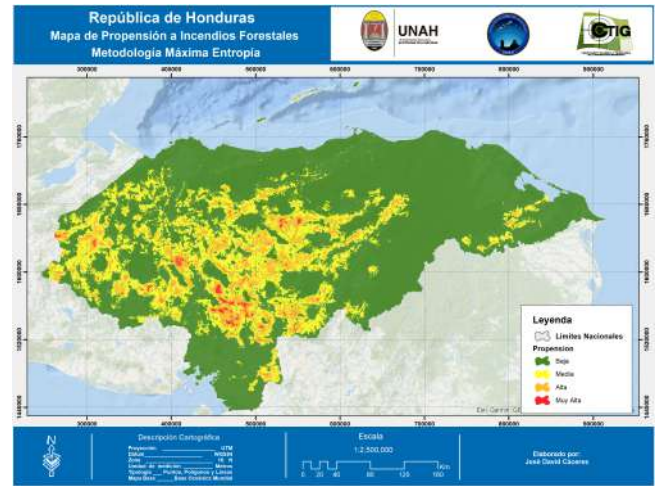
4.3 Patrones espaciales de probabilidades de ocurrencia de incendios

Los resultados de los tres modelos revelaron patrones espaciales distintivos de probabilidades de ocurrencia de incendios forestales (Figura 9). Los valores indican la idoneidad de cada píxel para tener una ocurrencia de incendio.

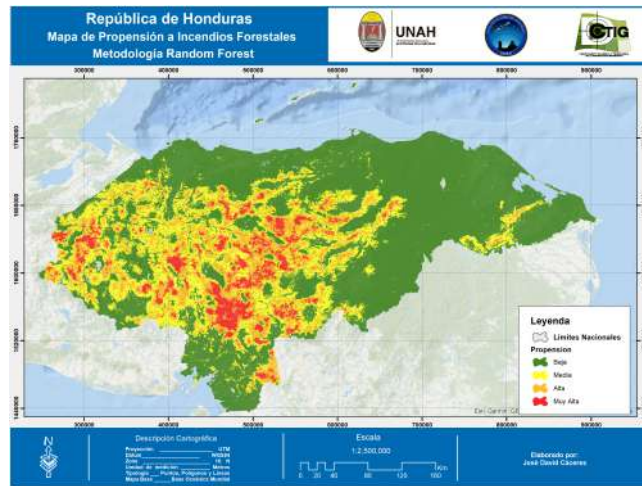
Para poder realizar la comparación estadística entre los tres modelos, debido a que los valores difieren entre sí, se realizó una reclasificación de los valores para llevarlos a una sola escala ordinal de medición, y así poder comparar el desempeño de los mismos. Se seleccionaron valores enteros en el rango de 1 a 4, como se ilustró en la Tabla 1. Valores de 1 representan la menor probabilidad de ocurrencia (colores verdes) y valores de 4 (colores rojos) representan la mayor probabilidad de ocurrencia de incendios.



(a)



(b)



(c)

Figura 9: Mapas predictivos de probabilidad de ocurrencia de incendios: (a) ICF, (b) MaxEnt, (c) Random Forest.

Una vez que los valores de los tres modelos se encontraban en la misma escala, se procedió a extraer los valores de los respectivos modelos utilizando los puntos de incendios reportados por ICF. Con estos valores se procedió a realizar un análisis multivariado de los mismos y se generó una tabla que muestra el resumen estadístico para cada uno de los modelos (Tabla 4). Incluye medidas de tendencia central, de variabilidad, y de forma. Siendo que los puntos seleccionados son puntos de incendios reales, se espera que los valores extraídos de los tres modelos posean en su mayoría la mayor categoría posible (4). Como se puede observar en la Tabla 4, el modelo con el promedio mayor, y por ende la mayor concordancia con la ocurrencia de incendios, es el modelo Random Forest, en concordancia con el modelo que obtuvo la mayor AUC. El modelo Random Forest es también el modelo con la menor desviación estándar (0.218081). Se puede observar también que en el modelo RF el valor que más se repite, la moda, pertenece a la categoría de mayor probabilidad de ocurrencia de incendios.

Tabla 4: Resumen estadístico de los tres modelos aplicados.

	ICF	MaxEnt	RF
Recuento	800	800	800
Promedio	2.28875	2.68875	3.95
Mediana	2.0	3.0	4.0
Moda	1.0	3.0	4.0
Desviación Estándar	1.21541	0.845735	0.218081
Coefficiente de Variación	53.1036 %	31.4546 %	5.52105 %
Mínimo	1.0	1.0	3.0
Máximo	4.0	4.0	4.0
Rango	3.0	3.0	1.0

En la Figura 10 se muestran los diagramas de dispersión entre cada par de los modelos de ocurrencia de incendios desarrollados, donde se muestra que MaxEnt y RF tienen una menor dispersión entre sí, en comparación con el modelo ICF.

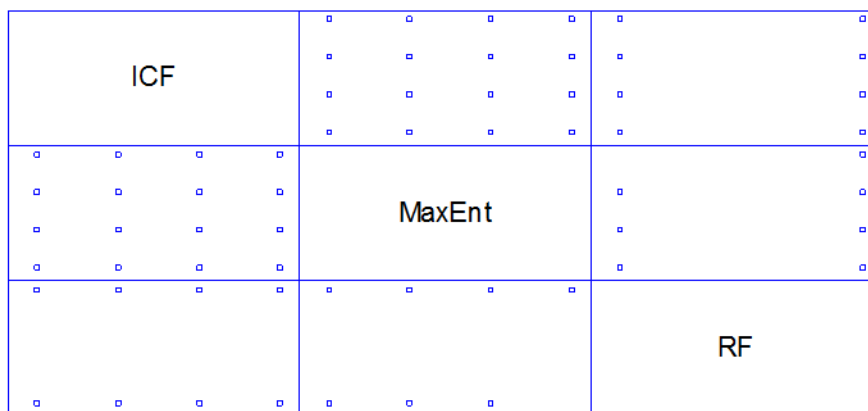


Figura 10: Diagramas de dispersión entre los modelos aplicados.

En la Tabla 5 se muestran las correlaciones momento producto de Pearson, entre cada par de variables. El rango de estos coeficientes de correlación va de -1 a +1, y miden la fuerza de la relación lineal entre las variables. También se muestra, entre paréntesis, el número de pares de datos utilizados para calcular cada

coeficiente. El tercer número en cada bloque de la tabla es un valor-P que prueba la significancia estadística de las correlaciones estimadas. Valores-P abajo de 0.05 indican correlaciones significativamente diferentes de cero, con un nivel de confianza del 95.0%. El modelo RF y MaxEnt son los que presentan la mayor correlación de Pearson.

Tabla 5: Correlaciones de Pearson entre cada par de modelos.

	ICF	MaxEnt	RF
ICF		-0.0695 (800)	-0.0257 (800)
		0.0490	0.4671
MaxEnt	-0.0695 (800)		0.3634 (800)
	0.0490		0.0000
RF	-0.0257 (800)	0.3634 (800)	
	0.4671	0.0000	

Por último, se tomaron los tres mapas derivados de los tres modelos y se procedió a calcular la correlación de Spearman (Tabla 6). El coeficiente de Correlación de Spearman (ρ) es una medida de la correlación (la asociación o interdependencia) entre dos variables aleatorias continuas. Para calcular ρ , los datos son ordenados y reemplazados por su respectivo orden. La interpretación del coeficiente de Spearman es igual que la del coeficiente de correlación de Pearson, Oscila entre -1 y +1, indicando asociaciones negativas o positivas respectivamente, o cero, significa no correlación pero no independencia.

Tabla 6: Correlación de Spearman entre cada par de modelo.

Modelo	RF	MaxEnt	ICF
RF	1.00000	0.80963	0.27354
MaxEnt	0.80963	1.00000	0.21082
ICF	0.27354	0.21082	1.00000

Es importante mencionar que, a pesar de las correlaciones moderadas entre los mapas predichos, hubo diferencias notables en los patrones y la magnitud. Los modelos RF y MaxEnt produjeron mapas similares de predicción ($\rho = 0.81$) mientras que el modelo ICF tuvo una baja correlación con el modelo MaxEnt ($\rho = 0.21$) y con el modelo RF ($\rho = 0.27$).

5 Conclusiones

Se analizaron las diferencias en el rendimiento, la importancia variable y los mapas de predicción de la probabilidad de ocurrencia de incendios forestales entre tres tipos comunes de modelos estadísticos y encontramos que los algoritmos de aprendizaje automático (Random Forest y Maxent) tuvieron un desempeño ligeramente mejor que el enfoque paramétrico (ICF) en términos de precisión de predicción. A pesar del poder predictivo similar de los diferentes tipos de modelos, los mapas de predicción resultantes de la probabilidad de ocurrencia de incendios fueron muy diferentes; siendo el modelo Random Forest

el más diferente ya que predijo mayores probabilidades de ignición en proporciones más grandes del país. Debido a que no existe una herramienta de modelado única y perfecta, los estudios de probabilidad de ocurrencia de incendios forestales pueden beneficiarse del uso de enfoques múltiples, lo que arroja un rango de predicciones en lugar de un solo mapa. Para el caso particular de los datos analizados, el modelo Random Forest fue el que arrojó mejores resultados.

Se recomienda observar en los modelos, el efecto de eliminar la/las variables que menos aportan a la adecuación del modelo (en este caso la proximidad a ríos secundarios) para analizar los cambios que podrían ocurrir en el ajuste del modelo. Así mismo, resulta importante incluir dentro de las variables del modelo la cobertura del suelo, ya que en el caso particular de Honduras, la cobertura de bosque de pino presenta una mayor ocurrencia de incendios que cualquier otra cobertura del suelo.

Si bien algunas de las variables ambientales del modelo pueden ser actualizadas cada 8 días, se recomienda que siempre que se tengan nuevos reportes de incendios, se vuelva a calibrar el modelo para poder identificar posibles patrones espaciales que esté siguiendo algún incendio, aún y cuando las variables ambientales permanezcan inalteradas.

6 Referencias

Bar Massada, A., Syphard, A., Hawbak, T., Stewart, S., y Radeloff, V. (2011). Effects of ignition location models on the burn patterns of simulated wildfires. *Environmental Modelling Software*, 26(5):583–592. Disponible en: <https://doi.org/10.1016/J.ENVSOFT.2010.11.016>.

Bar Massada, A., Syphard, A., Stewart, S., y Radeloff, V. (2013). Wildfire ignition-distribution modeling: a comparative study in the huron–manistee national forest, michigan, usa. *International Journal of Wildland Fire*, 22(2):174–183. Disponible en: <https://doi.org/10.1071/WF11178>.

Breiman, L. (2001). Random forests. *Machine Learning*, 45:5–32. Disponible en: <https://doi.org/10.1023/A:1010933404324>.

Cutler, D., Edwards, T., Beard, K., Cutler, A., Hess, K., Gibson, J., y Lawler, J. (2007). Random forests for classification in ecology. *The Ecological Society of America*, 88(11):2783–2792. Disponible en: <https://doi.org/10.1890/07-0539.1>.

Cáceres Coello, J. (2020). Análisis multitemporal de pérdida de cobertura en 65 microcuencas del corredor seco hondureño. año 2010 - 2017. *Ciencias Espaciales*, 13(1):42–59. Disponible en: <https://doi.org/10.5377/ce.v13i1.11255>.

Díaz-Avalos, C., Peterson, D., Alvarado, E., Ferguson, S., y Besag, J. (2001). Space–time modelling of lightning-caused ignitions in the blue mountains, oregon. *Canadian Journal of Forest Research*, 31(9):1579–1593. Disponible en: <https://doi.org/10.1139/x01-089>.

Elith, J. y Graham, C. (2009). Do they? how do they? why do they differ? on finding reasons for differing performances of species distribution models. *Ecography*, 32:66–77. Disponible en: <https://doi.org/10.1111/J.1600-0587.2008.05505.X>.

Elith, J., Graham, C., Anderson, R., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R., Huettman, F., Leathwick, J., Lehmann, A., Li, J., Lohmann, L., Loiselle, B., Manion, G., Mortiz, C., Nakamura, M., Nakazawa, Y., Overton, J., Townsend Peterson, A., Phillips, S., Richardson, K., Scachetti-Pereira, R.,

- Schapire, R., Soberón, J., Williams, S., Wisz, M., y Zimmermann, N. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29(2):129–151. Disponible en: <https://doi.org/10.1111/j.2006.0906-7590.04596.x>.
- Elith, J., Phillips, S., Hastie, T., Dudík, M., Chee, Y., y Yates, C. (2011). Na statistical explanation of maxent for ecologists. *Diversity and Distributions*, 17(1):43–57. Disponible en: <https://doi.org/10.1111/j.1472-4642.2010.00725.x>.
- Franklin, J. (2010). *Mapping species distributions: spatial inference and prediction*. Ecology, Biodiversity and Conservation. Disponible en: https://books.google.hn/books?hl=es&lr=&id=CkshAwAAQBAJ&oi=fnd&pg=PR15&dq=Mapping+species+distributions:+spatial+inference+and+prediction.+Ecology,+Biodiversity+and+Conservation.+Cambridge+University+Press&ots=6reRauKG3l&sig=UG3NNfzJDchp43Kb_LCdoK02W0k#v=onepage&q&f=false.
- Guisan, A., Zimmermann, N., Elith, J., Graham, C., Phillips, S., y Peterson, A. (2007). What matters for predicting the occurrences of trees: Techniques, data, or species' characteristics? *Ecological Monographs*, 77(4):615–630. Disponible en: <https://doi.org/10.1890/06-1060.1>.
- Hanley, J. y McNeil, B. (1982). The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143(1):29–36. Disponible en: <https://doi.org/10.1148/radiology.143.1.7063747>.
- Hastie, T. y Tibshirani, R. (1990). *Generalized Additive Models*. Monographs on statistics and applied probability: Vol. 43. Chapman and Hall., 1 st edición.
- Instituto de Conservación Forestal (2014). Análisis técnico para la definición de riesgo a incendios forestales. Technical report.
- Instituto de Conservación Forestal (2020). *Anuario Estadístico Forestal de Honduras 2019*. 34 ed., Tegucigalpa: Centro de Información y Patrimonio Forestal, Unidad de Estadísticas Forestales. Disponible en: <https://www.icf.gob.hn/?portfolio=cipf-2>.
- Kaufman, Y., Justice, C., Flynn, L., Kendall, J., Prins, E., Giglio, L., Ward, D., Menzel, W., y Setzer, A. (1998). Potential global fire monitoring from eos-modis. *Journal of Geophysical Research: Atmospheres*, 103:32215–32238. Disponible en: <https://doi.org/10.1029/98JD01644>.
- Krawchuk, M., Cumming, S., M.D.Flannigan, y R.W.Wein (2006). Biotic and abiotic regulation of lightning fire initiation in the mixedwood boreal forest. *Ecology*, 87:458–468. Disponible en: <https://doi.org/10.1890/05-1021>.
- Latham, D. y Williams, E. (2001). *Forest Fires*, capítulo Chapter 11 - Lightning and Forest Fires, pp. 375–418. Academic Press. <https://doi.org/10.1016/B978-012386660-8/50013-1>.
- Lobo, J., Jiménez-Valverde, A., y Real, R. (2008). Auc: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 17:145–151. Disponible en: <https://doi.org/10.1111/j.1466-8238.2007.00358.x>.
- Mas, J.-F., Filho, B., Pontius, R., Gutiérrez, M., y Rodrigues, H. (2013). A suite of tools for roc analysis of spatial models. *Revista internacional de información geográfica de ISPRS*, 2(3):869–887. Disponible en: <https://doi.org/10.3390/ijgi2030869>.

McCune, B. y Grace, J. (2002). *Analysis of Ecological Communities*. MjM Software Design, 1 edición.

Moreno Segura, E., Mondragón Rivera, C., Cáceres Coello, J., y Carias Arias, A. (2017). Identificación de riesgos por desastres naturales en las cuencas hidrográficas matasnillo (panamá) y goascorán (honduras), y su relación con el cambio climático utilizando tecnologías de información geográfica. *Ciencias Espaciales*, 9(1):53–80. Disponible en: <https://doi.org/10.5377/ce.v9i1.3126>.

Narayanaraj, G. y Wimberly, M. (2011). Influences of forest roads on the spatial pattern of wildfire boundaries. *Journal of the International Association of Wildland Fire*, 20(6):792–803. Disponible en: <https://doi.org/10.1071/WF10032>.

Nelder, J. y Wedderburn, R. (1972). Generalized linear models. *Journal of the Royal Statistical Society.*, 135(3):370–384. Disponible en: <https://doi.org/10.2307/2344614>.

Parisien, M. y Moritz, M. (2009). Environmental controls on the distribution of wildfire at multiple spatial scales. *Ecological Monographs*, 79(1):127–154. Disponible en: <https://doi.org/10.1890/07-1289.1>.

Pearson, R., Raxworthy, C., Nakamura, M., y Townsend Peterson, A. (2006). Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in madagascar. *Journal of Biogeography*, 34(1):102–117. Disponible en: <https://doi.org/10.1111/j.1365-2699.2006.01594.x>.

Phillips, S., Anderson, R., y Schapire, R. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190:231–259. Disponible en: <https://doi.org/10.1016/j.ecolmodel.2005.03.026>.

Phillips, S. y Dudík, M. (2008). Modeling of species distributions with maxent: new extensions and a comprehensive evaluation. *Ecography*, 31:161–175.

Prasad, A., Iverson, L., y Liaw, A. (2006). Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems*, 9:181–199. Disponible en: <https://doi.org/10.1007/s10021-005-0054-1>.

Renard, Q., Péliissier, R., Ramesh, B., y Kodandapani, N. (2012). Environmental susceptibility model for predicting forest fire occurrence in the western ghats of india. *International Journal of Wildland Fire*, 21(4):368–379. Disponible en: <https://doi.org/10.1071/WF10109>.

Roberts, J., Best, B., Dunn, D., Treml, E., y Halpin, P. (2010). Environmental modelling & software. *Environmental Modelling & Software*, 25:1197–1207. Disponible en: <https://doi.org/10.1016/j.envsoft.2010.03.029>.

Sturtevant, B. y Cleland, D. (2007). Human and biophysical factors influencing modern fire disturbance in northern wisconsin. *International Journal of Wildland Fire*, 16(4):398–413. Disponible en: <https://doi.org/10.1071/WF06023>.

Syphard, A. y Franklin, J. (2010). Species traits affect the performance of species distribution models for plants in southern california. *Journal of Vegetation Science.*, 21:177–189. Disponible en: <https://doi.org/10.1111/j.1654-1103.2009.01133.x>.

Syphard, A., Radeloff, V., Keeley, J., Hawbaker, T., Clayton, M., Stewart, S., y Hammer, R. (2007). Human influence on california fire regimes. *Ecological Applications.*, 17:1388–1402. Disponible en: <https://doi.org/10.1890/06-1128.1>.

Syphard, A., Radeloff, V., Keuler, N., Taylor, R., Hawbaker, T., Stewart, S., y Clayton, M. (2008). Predicting spatial patterns of fire on a southern california landscape. *International Journal of Wildland Fire.*, 17:602–613. Disponible en: <https://doi.org/10.1071/WF07087>.

Yang, J., He, H., Shifley, S., y Gustafson, E. (2007). Spatial patterns of modern period human-caused fire occurrence in the missouri ozark highlands. *Forest Science.*, 53(1):1–15.