

# **Instituto Tecnológico de Costa Rica**

## **Escuela de Ingeniería en Computación**

### **Programa de Maestría en Computación**

Inclusión de relaciones interespecíficas usando *Maxent*

**Propuesta de tesis sometida a consideración del Departamento de Computación, para optar por el grado de Magíster Scientiae en Computación, con énfasis en Ciencias de la Computación**

**Estudiante: José Morales Ortiz**

**Profesor Asesor: Erick Mata Montero**

### RESUMEN

Los modelos de distribución de especies normalmente se alimentan de ocurrencias, datos georeferenciados donde se ha visto a una especie en particular e información climática tal como la precipitación, la radiación solar, temperatura, entre otros. La distribución geográfica de una especie no solo depende de estas variables, sino también de otros factores como las relaciones entre las especies o las barreras geográficas, por ejemplo cordilleras o lagos.

En la presente investigación se pretende mejorar la calidad de los modelos de distribución de especies al incluir las relaciones interespecíficas. Para esto se definieron tres especies de aves y sus relaciones con las especies de plantas de las cuales se alimentan; la lapa verde (*Ara ambiguus*) y su relación con el almendro amarillo (*Dipteryx panamensis*), también el pájaro campana (*Procnias tricarunculatus*) y el quetzal (*Pharomachrus mocinno*) y su relación con las especies de la familia *Lauraceae*.

Para incluir las relaciones interespecíficas en los modelos de distribución de especies se definió una metodología para construir una capa de "rango de influencia"; con esta nueva capa se plantearon una serie de experimentos, con los cuales se pudo constatar que al incluir las relaciones interespecíficas en estos modelos se mejora su precisión. Para medir el mejoramiento se calcularon las diferentes curvas AUC (Área bajo la curva por sus siglas en inglés) de los modelos con el fin de realizar una comparación numérica.

**Palabras clave:** sistemas de información geográfica, modelos de distribución de especies, *Maxent* y relaciones interespecíficas.

## ABSTRACT

The species' distribution models are normally feed by occurrences, geo referenced data points where a specie has been seen, and climatic information such as precipitation, solar radiation, temperature, among others. The species geographic distribution not only depends on those variables, there are other factors such as the relationship between the species or the geographic barriers, as ridges and lakes.

The present research try to improve the quality of the species' distribution models by including the relationship between the species. In order to achieve that three bird species and the relationship with the plants which serve them as food were defined. The selected species are the lapa verde (*Ara ambiguus*) and its relationship with the almendro amarillo (*Dipteryx panamensis*), the pájaro campana (*Procnias tricarunculatus*) and its relationship with the *Lauracea* family and finally the quetzal (*Pharomachrus mocinno*) and its relationship with the *Lauracea* family as well.

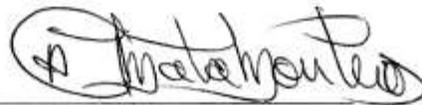
In order to include the relationship between species into the species' distribution models a methodology in order to build a layer of "range of influence" were defined; with this new layer a set of experiments were defined, and those experiments verifies that by including the relationships between the species into the species' distribution models improve their precision. In order to measure the improvements the AUC curve were calculated for each model, in that way a numeric comparison were done.

**Key words:** geographic information systems, species' geographic distribution models, Maxent, interespecific relationships.

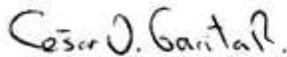
## APROBACIÓN DE LA TESIS

# "Inclusión de las relaciones interespecíficas usando Maxent."

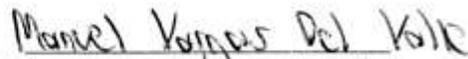
### TRIBUNAL EXAMINADOR



Dr. Erick Mata Montero  
Profesor Asesor



Dr. Cesar Garita Rodríguez  
Profesor Lector



MSc. Manuel Vargas Del Valle  
Profesor Externo



Dr. Roberto Cortés Morales  
Coordinador del Programa  
de Maestría en Computación

Noviembre, 2015

## DEDICATORIA

*Este trabajo se lo dedico a mi querido abuelo quien murió hace unos años, un gran ejemplo a seguir para mí, fue siempre la persona que creyó en mí y de la que siempre tuve un apoyo incondicional.*

## AGRADECIMIENTOS

*Quiero agradecer a mi profesor tutor el doctor Erick Mata Montero, sin el cual este trabajo no hubiera sido posible. También, dar las gracias al doctor Olivier Chassot por una charla provechosa para la investigación y por todos los datos suministrados, los cuales sirvieron de insumo para los experimentos. Finalmente, a todas aquellas personas que me impulsaron a seguir trabajando en este proyecto y aportaron de una u otra forma.*

## CONTENIDO

Resumen .....	2
Abstract .....	3
Dedicatoria .....	5
Agradecimientos.....	5
1. Introducción .....	9
2. Antecedentes De los Modelos de Distribución .....	11
3. Sistemas de información geográfica.....	12
3.1 Definición .....	12
3.2 Componentes.....	13
3.3 Estructuras de datos .....	14
4. Modelos de Distribución de especies.....	15
4.1 Definición .....	15
4.2 Métodos de modelización.....	16
4.2.1 Técnicas discriminantes .....	16
4.2.2 Técnicas descriptivas .....	17
4.2.3 Técnicas Mixtas.....	17
4.3 Usos e Importancia .....	17
4.4 <i>Maxent</i> .....	17
4.4.1 Ventajas .....	19
4.4.2 Desventajas.....	19
4.4.3 El principio de <i>Maxent</i> .....	19
4.4.4 Aplicación de <i>Maxent</i> en el modelado de distribución de especies.....	20
4.5 Evaluación de los modelos .....	21
4.5.1 Aplicación del análisis ROC con datos de solo presencia.....	22
4.6. ¿Por qué la selección de <i>Maxent</i> ? .....	23
5. Relaciones Biológicas entre especies .....	24

6. Propuesta de tesis .....	26
6.1 Hipótesis.....	26
7. Objetivos y alcance .....	26
7.1 Objetivo General .....	26
7.2 Objetivos específicos.....	26
7.3 Alcances y limitaciones .....	26
8. Recolección de Datos .....	28
8.1 Especies Seleccionadas .....	28
8.1.1 <i>Ara ambiguus</i> .....	28
8.1.2 Almendro de montaña ( <i>dipterix panamensis</i> ) .....	29
8.1.3 Quetzal ( <i>Pharomachrus mocinno</i> ) .....	30
8.1.4 Pájaro campana ( <i>Procnias tricarunculatus</i> ) .....	30
8.1.5 Familia <i>lauraceae</i> .....	31
8.2 Información Climática .....	32
9. Metodología para la inclusión de las relaciones interespecíficas en <i>Maxent</i> .....	33
10. Experimentos realizados .....	38
10.1 Efectividad de <i>Maxent</i> usando solo datos climáticos contra Datos climáticos y Capa de rango de Influencia.....	38
10.1.1 Lapa Verde ( <i>Ara Ambiguus</i> ) solo con Datos climáticos.....	38
10.1.2 Lapa Verde ( <i>Ara Ambiguus</i> ) con datos climáticos y la capa de rango de influencia de Almendros Amarillos como alimentación .....	40
10.1.3 Lapa Verde ( <i>Ara Ambiguus</i> ) con datos climáticos y la capa de rango de influencia de Almendros Amarillos como Nidos.....	41
10.1.4 Pájaro Campana ( <i>Procnias tricarunculatus</i> ) utilizando solo datos Climáticos .....	42
10.1.5 Pájaro Campana ( <i>Procnias tricarunculatus</i> ) utilizando datos Climáticos y la capa de rango de influencia ( <i>Lauraceae</i> ) .....	44
10.1.6 Quetzal ( <i>Pharomachrus mocinno</i> ) utilizando solo datos Climáticos.....	45
10.1.7 Quetzal ( <i>Pharomachrus mocinno</i> ) utilizando datos Climáticos y la capa de rango de influencia ( <i>Lauraceae</i> ).....	46

10.2 Capacidad de Extrapolación de <i>Maxent</i> al usar solo datos climáticos contra Usar Datos Climáticos más el rango de influencia .....	48
10.2.1 Lapa verde usando el 50% de ocurrencias y solo datos climáticos.....	48
10.2.2 Lapa verde usando el 50% de ocurrencias y datos climáticos en combinación con la capa de rango de influencia (Alimentación) .....	50
10.2.3 Lapa verde usando el 50% de ocurrencias y datos climáticos en combinación con la capa de rango de influencia (Nidos).....	51
10.2.4 Pájaro campana usando el 50% de ocurrencias y solo datos climáticos.....	53
10.2.5 Pájaro campana usando el 50% de ocurrencias Usando datos climáticos en combinación con la capa de rango de influencia.....	55
10.2.6 Quetzal usando el 50% de ocurrencias y solo datos climáticos.....	56
10.2.7 Quetzal usando el 50% de ocurrencias y datos climáticos en combinación con la capa de rango de influencia .....	58
11. Resumen de resultados y conclusiones .....	60
12. Trabajo Futuro .....	63
13. Bibliografía.....	64
14. Glosario .....	66
15. Apéndice.....	67
15.1 Nidos de Lapa Verde .....	67
15.2 Programa hecho en <i>Java</i> para crear la capa de rango de influencia.....	74

## 1. INTRODUCCIÓN

En la actualidad el medio ambiente se ha visto gravemente afectado, esto por aspectos como la contaminación, la tala, el calentamiento global, la caza ilegal y especies invasoras introducidas involuntariamente por los seres humanos. Debido a esto, muchas organizaciones se han dado a la tarea de revertir la situación para preservar el rico patrimonio natural que tenemos. Por ejemplo, se busca crear fuentes de energía nuevas y más limpias, para evitar seguir liberando las cantidades enormes de gases de efecto invernadero que se vierten a la atmosfera. También, se fomenta la creación de nuevas áreas protegidas, donde no se permite la caza ni la tala de árboles. Además, se buscan nuevos materiales que sean biodegradables para evitar la contaminación, y se realizan diversas investigaciones sobre los efectos de las especies invasoras en los ecosistemas.

Sin embargo, ¿cómo podemos saber qué área debemos proteger?, ¿en dónde se ubican las especies invasoras?, ¿cómo afecta el cambio climático a las especies?, y ¿cómo interactúa una especie con su entorno? Una herramienta que puede ayudar a responder estas y muchas otras preguntas son los SIG (sistemas de información geográfica) o GIS por sus siglas en inglés.

Un uso de estos sistemas de información geográfica es su aplicación a la distribución potencial de una especie. Así por ejemplo para saber qué área debemos proteger para salvar a una especie "x" que está en peligro de extinción, podríamos utilizar un SIG para predecir su distribución potencial y así conoceríamos las áreas en que vive esa especie. Con esa información se tendrían fundamentos para proteger dichas áreas; debido a que estos sistemas se alimentan de gran variedad de variables ambientales, incluidas las climáticas, tales como la temperatura, la precipitación, la humedad, la radiación solar, entre muchas otras. Otra funcionalidad es la realización de correlaciones para saber cómo se comportaría una especie "x" si las condiciones climáticas cambiaran eventualmente en alguna medida. Otro ejemplo de cómo se pueden utilizar los SIG es la predicción del área potencial donde una especie invasora "z" vive, de esta forma se pueden hacer preparativos en las posibles áreas donde la especie puede vivir, y así intentar evitar su propagación.

Debido a la utilidad que tienen los sistemas de información geográfica, en específico su uso para la predicción de la distribución de una especie, es que esta investigación propone mejorar los modelos. Existe una gran variedad de métodos para la generación de estos modelos y las variables que los constituyen [Mateo, Felicísimo, Muñoz, 2011]. Una de las variables que normalmente se excluyen son las relaciones interespecíficas, las cuales pueden influenciar en gran medida la distribución geográfica de una especie. Mediante la inclusión de esta variable en uno de los métodos más utilizados para la generación de modelos de distribución geográfica, el método *Maxent*, se intentará probar como la inclusión de estas relaciones biológicas pueden mejorar la efectividad del método al predecir la distribución geográfica de una especie.

En el segundo capítulo de este documento se da una breve reseña de los antecedentes de los modelos de distribución de especies, además de algunos ejemplos de casos en los cuales se han aplicado. En el capítulo tres se menciona cómo están compuestos los sistemas de información geográfica, de los cuales los modelos de distribución de especies son una variante. En el capítulo cuatro se explica qué son los modelos de distribución de especies, cómo se construyen y sus usos e importancia; además, se habla de *Maxent*, la técnica utilizada en la presente investigación y de cómo se evalúan los modelos. En el capítulo cinco se explica qué son las relaciones interespecíficas, las cuales se quieren agregar a los modelos de distribución de especies. En el capítulo seis se expone la propuesta de tesis; En el siete, los objetivos y los alcances; En el ocho, de dónde se tomaron los datos necesarios para la construcción de los modelos; en el

nueve, se explica la metodología utilizada para incluir las relaciones interespecíficas a los modelos de distribución de especies; En el diez, se describen los experimentos realizados y sus resultados; En el once, se resumen los resultados y se sacan las conclusiones; y en el 12, se describen algunos puntos que pueden servir de inicio para nuevas investigaciones; en el trece, se listan las citas bibliográficas utilizadas; en el catorce, se listan algunos conceptos básicos para el entendimiento de la presente investigación; y por último en el quince (el apéndice), se muestra la información suministrada por el doctor Olivier Chassot del CCT y el código fuente del programa hecho en *Java* que sirve de soporte para la construcción de los modelos realizados en la presente investigación.

## 2. ANTECEDENTES DE LOS MODELOS DE DISTRIBUCIÓN

Los modelos de distribución potencial de especies son de gran utilidad, por ende se han desarrollado una gran cantidad de técnicas para su desarrollo; por ejemplo, en [Mateo, Felicísimo, Muñoz, 2011] se mencionan 3 diferentes técnicas: las discriminantes que necesitan datos de presencia y ausencia, las descriptivas que solo requieren datos de presencia y las mixtas. Solo para las técnicas discriminantes se tienen 8 diferentes tipos, entre los que destacan los modelos aditivos generalizados (GAM por sus siglas en inglés), los modelos lineales generalizados (GLM por sus siglas en Inglés), redes neuronales (ANN Redes neuronales asistidas por sus siglas en inglés) y árboles de clasificación y máxima entropía.

Para cada uno de los tipos de técnicas discriminantes existen diferentes implementaciones; por ejemplo, para los árboles de clasificación existe *CART* [Mateo, Felicísimo, Muñoz, 2011] o para máxima entropía tenemos el *Maxent* [Phillips, Anderson, Schapire, 2006]. A su vez, para las técnicas descriptivas podemos encontrar muchas otras implementaciones como *BIOCLIM*, *HABITAT*, *FLORAMAP*, entre muchas otras [Mateo, Felicísimo, Muñoz, 2011].

Cada uno de esos métodos a su vez se han utilizado en diferentes estudios, por ejemplo en [Owens, Bentley, Townsend, 2012] se utiliza *GARP* y *Maxent* para predecir los ambientes donde potencialmente pueden vivir los *Latimeria chalumnae*, una rara especie de pez. Otro ejemplo interesante es el estudio realizado en [Bond *et al*, 2011] donde se utilizaron varias técnicas de modelado de distribución de especies para predecir los efectos del cambio climático, esto en alrededor de 40 diferentes especies de peces de agua dulce.

Existen muchas otras aplicaciones como las de los métodos mencionados anteriormente. Sin embargo, en la mayoría de estudios, como se menciona en [Mateo, Felicísimo, Muñoz, 2011], no se incluye un componente muy importante, el cual es las relaciones entre las especies. Estas pueden llegar a influenciar la distribución geográfica de una especie, ya que pueden generar presión tanto para la presencia de una especie como para su ausencia (en el caso de la competencia). Por ejemplo, la lapa verde (*Ara ambiguus*) se alimenta principalmente del almendro de montaña (*Dipteryx panamensis*), por lo que si el almendro de montaña está presente en un área específica es posible que también en esa área pueda vivir la lapa verde.

Debido a la gran importancia que tienen estas relaciones es que se quiere incorporar al *Maxent* [Phillips, Anderson, Schapire, 2006] las relaciones interespecíficas, para así mejorar su precisión a la hora de predecir la distribución geográfica de una especie.

### 3. SISTEMAS DE INFORMACIÓN GEOGRÁFICA

#### 3.1 DEFINICIÓN

Un sistema de información geográfica es una herramienta de análisis en la cual se utilizan datos georeferenciados. También, se puede utilizar la definición dada por [Bolstad, 2012]: "Un SIG es un sistema basado en computadora que apunta a la acumulación, mantenimiento, almacenamiento, análisis, salida y distribución de datos espaciales e información".

Los datos georeferenciados normalmente se agrupan por tipo, cada grupo de esos datos se llama capa. En un SIG normalmente se trabaja con varias capas; por ejemplo, en un SIG que muestre una ciudad se puede tener capas para el tipo de suelo, los ríos y acequias, una capa para las carreteras, y otra para las diferentes edificaciones. Esta es una imagen tomada de *Google Maps* que ilustra el ejemplo de un SIG que muestra información de una ciudad.



Figura 1: ejemplo de vista de una ciudad

En la Figura 1, se pueden ver diferentes tipos de información de la ciudad; por ejemplo, se tiene una capa con las calles, otra con la imagen satelital, y otra más con los nombres de diferentes lugares.

En un SIG tener esta separación de las capas es de gran ayuda ya que permite analizar los datos por separado, con esto se podría llegar a un mejor entendimiento de los fenómenos que pasan en la vida real. Por lo tanto, hay una mayor exactitud entre los valores arrojados por el modelo y situaciones reales. Esto le agrega valor al modelo.

Otra definición es la dada por [Foote, Lynch, 1995]: "Un SIG es una base de datos digital de propósito especial en la cual un sistema de coordenadas espaciales es la primer fuente de referencia. Un SIS integral requiere los siguientes parámetros:

- Datos de entrada, desde mapas, fotos aéreas, satélites, encuestas y otras fuentes.
- Almacenamiento de datos, acceso y consulta.

- Transformación de datos, análisis y modelado, incluyendo estadísticas espaciales.
- Reportes de datos, tales como mapas, reportes y planos."

## 3.2 COMPONENTES

Según [Bolstad, 2012], un SIG está compuesto de hardware, software, datos, personas y un conjunto de protocolos organizacionales. Dentro del hardware requerido podemos encontrar computadoras rápidas con gran cantidad de almacenamiento y monitores de buena calidad. En el caso del software este debe proveer herramientas para manipular, analizar y mostrar información espacial. Además, este autor menciona algunas funciones que debe tener el software de un SIG entre las cuales destacan:

- Ingreso de datos:
  - Captura manual de coordenadas.
  - Captura de atributos.
  - Captura digital de coordenadas.
  - Importación de datos.
- Edición:
  - Edición de puntos, líneas y áreas.
  - Edición de atributos.
  - Edición y detección automática de errores.
- Administración de datos:
  - La Copia, partición y unificación de datos.
  - Versionamiento.
  - Resumen y reducción de datos.
  - Compresión.
  - Indexación.
  - Registro de datos y proyección de datos.
- Análisis:
  - Consultas espaciales.
  - Consulta de atributos.
  - Interpolación.
  - Conectividad.
  - Proximidad y adyacencia.
  - Análisis del terreno.
  - Disolución de límites.
  - Algebra para manejo de mapas.
- Salidas:
  - Diseño de mapas y plantillas.
  - Impresión de mapas.
  - Producción gráfica digital.
  - Generación de formatos para exportar.
  - Salida de metadatos.

Los componentes de cada SIG van a depender en gran medida de su finalidad. Por ejemplo, si el objetivo es un SIG que puede ser accedido por internet se necesitará un servidor web; pero por el contrario,

si el objetivo es un SIG que va a ser utilizado desde una máquina de escritorio, entonces ese servidor web no tendría ninguna funcionalidad o uso.

### 3.3 ESTRUCTURAS DE DATOS

Los datos georeferenciados en los SIG se almacenan principalmente de tres maneras: un procedimiento utiliza datos vectoriales o datos *raster*. También, como menciona [Parmenter, 2007] se utilizan datos TIN (red irregular triangulada, TIN por sus siglas en Inglés) o datos tabulados (una tabla de atributos).

Los datos vectoriales pueden ser de tres tipos:

- Puntos: elementos que son un par de coordenadas x, y.
- Líneas: elementos que están compuestos por un conjunto de coordenadas.
- Polígonos: elementos que están compuestos por conjuntos de pares de coordenadas que muestran los límites del polígono.

Los datos *raster* se representan por una matriz de celdas, donde cada celda tiene un valor específico. En la figura 2 se ejemplifica la comparación entre datos de tipo *raster* y vectoriales, donde se muestra la representación del mundo real en cada tipo de dato:



Figura 2: datos *raster* vs datos vectoriales. Tomada de [www.aulati.net](http://www.aulati.net)

## 4. MODELOS DE DISTRIBUCIÓN DE ESPECIES

### 4.1 DEFINICIÓN

Según se menciona en [Mateo, Felicísimo, Muñoz, 2011] un modelo es una representación parcial de la realidad, en el cual se reflejan algunas de sus propiedades. Los modelos son simplificaciones de la realidad que nos ayudan a entender mejor los fenómenos que se quieren modelar. Una de las características que mejor los define es que nunca están completos, ya sea por la necesidad de simplificarlos, o por desconocimiento de sus propiedades y eventos de las entidades en el mundo real, sumamente complejos de replicar. Los modelos de distribución de especies son representaciones cartográficas de la idoneidad de un espacio para la presencia de una especie en función de las variables empleadas para generar dicha representación.

Los modelos de distribución de especies utilizan distintas variables, se tienen datos de presencia/ausencia o solo presencia de una especie, los cuales son básicamente localizaciones (puntos georeferenciados) donde se ha visto a la especie en estudio. Se toman en cuenta una amplia gama de variables ambientales, como la elevación, la humedad, la temperatura, la cobertura vegetal, uso de suelo, entre otras. Estas variables ambientales pueden ser de tipo numéricas o categóricas. Por ejemplo, una variable tal como la temperatura tiene valores numéricos, mientras que una variable tal como el uso del suelo o el tipo de vegetación son de tipo categórico [Mateo, Felicísimo, Muñoz, 2011].

En un modelo de distribución de especies se genera un valor para cada pixel en el área de estudio, ese valor dice que tan idóneo es la localización para la especie en estudio. Esa idoneidad se obtiene del estudio de la relación de las diferentes variables con respecto a las ocurrencias (lugares donde fue vista la especie en estudio).

En el caso de los datos de ocurrencias, tal y como se menciona en [Mateo, Felicísimo, Muñoz, 2011] o [Phillips, Anderson, Schapire, 2006], se tiene la grave limitante de que en la mayoría de los casos solo se cuenta con datos de presencia. Son raras las ocasiones donde se hace un estudio que incluya ambos tipos de datos, debido a esto no se pueden usar métodos estadísticos estándares donde se utilizan ambos tipos de datos. Se debe entonces crear escenarios tales como: si no se ha visto a la especie, entonces se infiere que no se encuentra en definitiva en esa área, lo cual puede introducir error en múltiples casos. Además, como ilustra [Mateo, Felicísimo, Muñoz, 2011] las ocurrencias tienen otros problemas:

- Generalmente los datos provienen de observadores oportunistas y muestreos no dirigidos.
- Los casos donde se diseñan muestreos expresamente para estimar la distribución de una especie son poco comunes.
- Existen errores de geolocalización e identificación taxonómica.

En la figura 3 se muestra un ejemplo de un modelo de distribución de especies generado con el método *Maxent*:

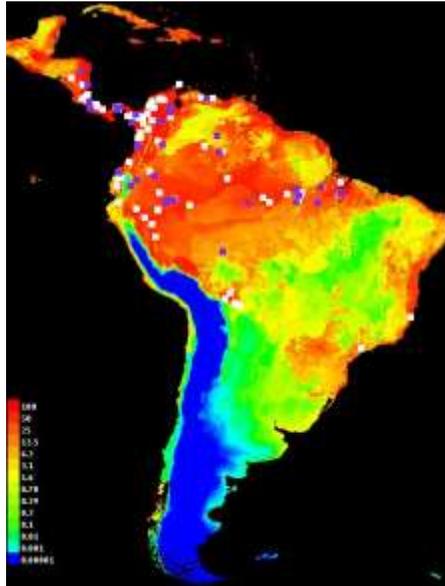


Figura 3: modelo generado con *Maxent*. Tomado de <http://www.oceanecology.ca>

## 4.2 MÉTODOS DE MODELIZACIÓN

Existen muchos métodos para modelar la distribución geográfica de una especie, y según [Bolstad, 2012], se pueden clasificar en dos técnicas distintas:

### 4.2.1 TÉCNICAS DISCRIMINANTES

Son aquellas que necesitan de datos de presencia y ausencia para construir el clasificador, estas se pueden subdividir en:

1. Árboles de clasificación (*CART* árboles de clasificación y regresión) y sus derivados, por ejemplo *random forest*.
2. Técnicas de ordenación, por ejemplo el análisis de correspondencia canónica o CCA por sus siglas en inglés.
3. Redes neuronales ANN (assisted neural network).
4. Modelos lineales generalizados GLM.
5. Modelos aditivos generalizados o GAM.
6. Estadística bayesiana.
7. Máxima entropía, por ejemplo *Maxent*.

---

#### 4.2.2 TÉCNICAS DESCRIPTIVAS

Son aquellas técnicas que solo requieren datos de presencia, entre ellas se encuentran:

- Deductivas, como la opinión del experto.
- Envueltas geográficas o *convex hull*.
- Envueltas ambientales, como *BIOCLIM*.
- Métodos de distancias matemáticas, como *DOMAIN* o *Biomapper*.
- *Expectation, Maximization Algorithm*.

---

#### 4.2.3 TÉCNICAS MIXTAS

En estos métodos se combinan modelos individuales para generar modelos de consenso, un ejemplo de esta técnica es *BIOMOD*.

### 4.3 USOS E IMPORTANCIA

Existen muchos usos para los modelos de distribución de especies como menciona [Mateo, Felicísimo, Muñoz, 2011], entre ellos se pueden enlistar:

- Cuantificación de efectos del cambio climático.
- Identificación de hábitats amenazados.
- Identificación de patrones de diversidad.
- Documentación de la distribución y riqueza de especies.
- Conocimiento de la distribución potencial de una comunidad.
- Identificación de núcleos de especies invasoras.
- Creación de áreas de conservación.
- Reintegración de especies en peligro de extinción.
- Distribución potencial de enfermedades infecciosas.

Los modelos de distribución de especies son herramientas fundamentales para muchos campos científicos debido a la diversidad de información que brindan. Gracias a ellos, se puede entender mejor como se relaciona una especie con su entorno; además, incluye la investigación a futuro de muchas otras relaciones interespecíficas de la especie en su hábitat (una vez determinada la distribución de la especie), identificadas por el uso del modelo, sin dejar de lado las relaciones interespecíficas que son la base de este análisis.

### 4.4 MAXENT

*Maxent* es un método de aprendizaje de computadora de uso general, utilizado para hacer predicciones o inferencias a partir de datos incompletos; según se menciona en [Phillips, Anderson, Schapire, 2006] sus orígenes están en la estadística mecánica, la cual es una rama de la física teórica. La

entropía, como se menciona en [Phillips, Anderson, Schapire, 2006] se define como "Una medida de cuanta 'decisión' está envuelta en la selección de un evento".

La idea de *Maxent* es estimar la distribución más probable de una especie al encontrar su máxima entropía, es decir, su distribución predominante o la más cercana a la uniformidad. Esto lo hace tomando en cuenta las limitantes que representa tener información incompleta. Como se menciona en [Phillips, Anderson, Schapire, 2006] la información disponible sobre la distribución de la especie se presenta como un conjunto de variables con valores reales, las cuales son llamadas "características". Las limitantes son que los valores esperados para cada característica deben igualarse con su promedio empírico (el valor promedio para un conjunto de puntos tomados de la distribución de la especie).

Los pixeles del área de estudio son el espacio donde *Maxent* define la distribución probable de la especie. Los pixeles con ocurrencias conocidas de la especie constituyen puntos de ejemplo, y las características son variables climáticas, de elevación, categoría del suelo, tipo de vegetación, entre muchas otras [Phillips, Anderson, Schapire, 2006].

Los creadores de *Maxent* construyeron un software en *Java*, el cual es de uso libre; debido a que fue construido en *Java* (programa multiplataforma) permite ejecutarlo mediante la instalación de la máquina virtual de *Java* en el equipo a utilizar. En la figura 4 se muestra una imagen de la interfaz gráfica de *Maxent*. A la izquierda se observa el espacio donde se agregan las muestras de las ocurrencias de las especies, mientras a la derecha se muestra el espacio donde se incluye la información climática. Esta tiene que estar en el formato *ASCII raster grid*:

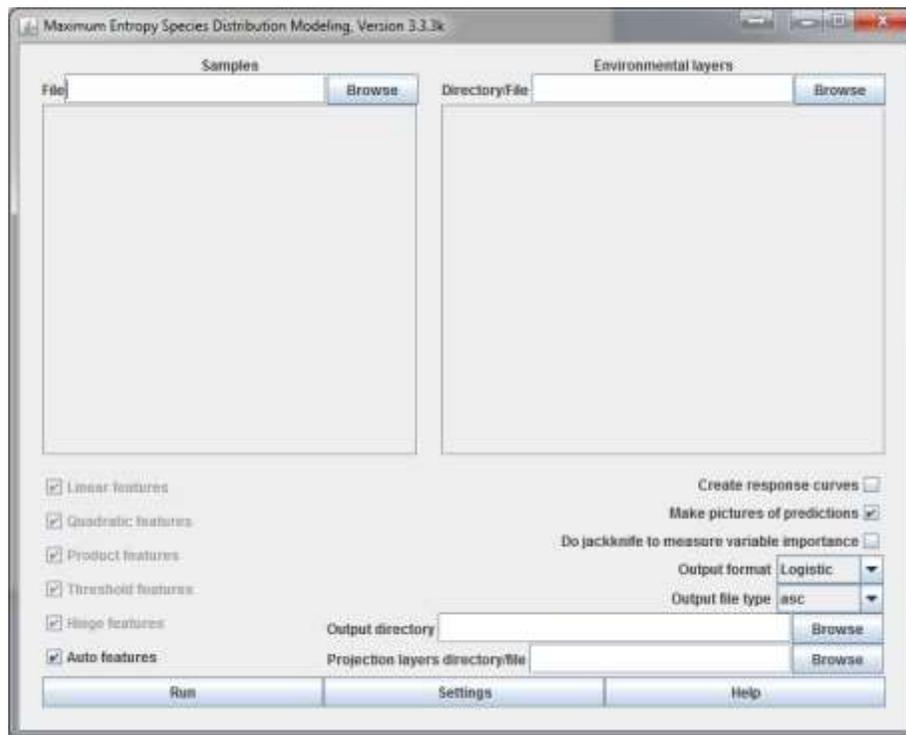


Figura 4: interfaz gráfica de *Maxent*

#### 4.4.1 VENTAJAS

Según sus autores [Phillips, Anderson, Schapire, 2006] *Maxent* tiene las siguientes ventajas:

- Solo requiere datos de presencia e información medio ambiental para el área de estudio.
- Puede utilizar datos ya sea continuos o categóricos y puede incorporar interacciones entre las diferentes variables.
- Se han desarrollado algoritmos determinísticos que garantizan la convergencia a la probable distribución óptima (máxima entropía).
- La distribución probable de *Maxent* tiene una definición matemática concisa, con lo que es posible realizarle diferentes análisis.
- *Maxent* tiene un enfoque generativo en lugar de discriminativo, lo cual puede ser una ventaja cuando se tienen muestras limitadas.

#### 4.4.2 DESVENTAJAS

- No es un método estadístico tan maduro como GLM (métodos lineales generalizados) o GAM (métodos aditivos generalizados) por lo que existe menos información sobre cómo utilizarlo y menos métodos para estimar la cantidad de error en sus predicciones.
- Se necesita más investigación para saber cuánto se debe regular el método.
- Se necesita software especial para usar *Maxent*, ya que los paquetes estadísticos estándar no lo incluyen.

#### 4.4.3 EL PRINCIPIO DE MAXENT

Cuando se está aproximando una posible distribución surge la pregunta ¿Cuál es la mejor distribución? El enfoque es asegurarse que esa posible distribución cumple con todas las limitantes de las cuales tenemos conocimiento, y que sujeta a esas limitantes la distribución tiene máxima entropía. A esto se le llama principio de máxima entropía [Phillips, Anderson, Schapire, 2006].

La distribución desconocida la vamos a denotar como  $\pi$ , la cual se encuentra en un conjunto finito  $X$  (El cual es el conjunto de pixeles en el área de estudio). A cada elemento de  $X$  se le llama punto.

La distribución desconocida  $\pi$  asigna una probabilidad no negativa  $\pi(x)$  a cada punto  $x$  en el conjunto  $X$ , todas estas probabilidades suman 1. La aproximación que se hace a  $\pi$  es también una distribución probable, y a esta se le va a llamar  $\Omega$ ; la entropía de  $\Omega$  se define basado en [Phillips, Anderson, Schapire, 2006] como:

$$H(\Omega) = - \sum_{x \in X} \Omega(x) \ln \Omega(x)$$

La entropía es no negativa y es a sumo el logaritmo natural del número de elementos en  $X$  [Phillips, Anderson, Schapire, 2006]. Como se mencionó anteriormente la entropía se puede definir como la cantidad de "decisiones" que están envueltas en la selección de un evento. Entonces se puede decir que a mayor la entropía más número de decisiones están involucradas.

Como se menciona en [Phillips, Anderson, Schapire, 2006], ya que una probable distribución maximiza la entropía sujeta a las limitantes relacionadas a la información incompleta dada, se puede decir que este es el principio fundamental para poder inferir la distribución, ya que esta se basa en todas las cosas que se conocen, pero evade asumir algo de las cosas que se desconocen.

#### 4.4.4 APLICACIÓN DE *MAXENT* EN EL MODELADO DE DISTRIBUCIÓN DE ESPECIES

Según es mencionado por los autores de *Maxent* en [Phillips, Anderson, Schapire, 2006], para el modelado estadístico de la distribución de especies se necesitan tres componentes: un modelo ecológico, el cual especifica la teoría ecológica a utilizar, un modelo de datos, y un modelo estadístico. *Maxent* es un modelo estadístico por lo que solo falta definir su relación con el modelo ecológico y el modelo de datos.

El modelo de datos es el método con el cual las localidades de presencia fueron recolectadas. Normalmente, como se menciona en varios artículos, [Mateo, Felicísimo, Muñoz, 2011], [Phillips, Anderson, Schapire, 2006] y [Peterson, 2006] concuerdan en que los muestreos tienen varios problemas, entre ellos: fallas en la geolocalización, muestreos mal diseñados, errores de identificación de la especie, entre otros. Debido a estos problemas la distribución desconocida  $\pi$  (la que se busca) y  $\Omega$  (nuestra aproximación) se pueden interpretar como un índice relativo de sustentabilidad ambiental, donde los valores más altos representan una predicción de mejores condiciones para las especies.

Para definir el modelo ecológico se deben establecer un conjunto de características. Las limitantes impuestas por las características representan todos los factores ambientales que limitan la distribución geográfica de una especie. En [Phillips, Anderson, Schapire, 2006] se definen cinco tipos de características:

1. Una variable continua  $f$  es en esencia una "característica lineal". Ella impone la limitante en  $\Omega$  en la que la variable ambiental  $\Omega[f]$  debe estar cerca de su valor observado, lo cual es su significado en las localidades muestreadas.
2. El cuadrado de una variable continua  $f$  es una "característica cuadrática". Cuando se usa con la característica lineal correspondiente, esta característica impone la siguiente limitante en  $\Omega$ : la varianza de la variable ambiental debe estar cerca de su valor observado, ya que la varianza es igual a  $\Omega[f^2] - \Omega[f]^2$ . Esto modela la tolerancia de las especies a la variación de sus condiciones óptimas.
3. El producto de dos variables ambientales continuas  $f$  y  $g$  es una "característica producto"; junto con las características lineales para  $f$  y  $g$ , esta impone una limitante a la covarianza de las dos variables: ambas deben estar cercanas a sus valores observados, ya que la covarianza es igual a  $\Omega[fg] - \pi[f]\pi[g]$ . Las características producto incorporan interacciones entre las variables pronosticadoras.
4. Para una variable ambiental continua  $f$ , una "característica umbral" es igual a 1 cuando  $f$  está encima de un umbral dado, en caso contrario la característica umbral tendrá el valor de 0. Este rasgo impone la siguiente limitante: La proporción de  $\pi$  que tiene valores para  $f$  sobre el umbral deben estar cerca de la proporción observada. Unidas todas las posibles características umbral para  $f$ , le permiten a *Maxent* modelar una curva de respuesta arbitraria para las especies en  $f$ , ya que cualquier función suave puede ser aproximada por una combinación lineal de funciones umbral.
5. Para una variable ambiental categórica que toma valores en  $v_1, \dots, v_k$  se usan  $k$  "características binarias", donde la  $i$ -ésima característica es 1 cuando la variable es igual a  $v_i$  o 0 en caso contrario. Como con las características umbral, estas características binarias limitan la proporción de  $\Omega$  en cada característica para que estén cerca del valor observado.

Para cada tipo de característica se tiene un correspondiente  $\beta_j$  que dice cuál cerca se requiere la expectativa bajo  $\Omega$  del valor observado. Sin regularización, estas tienen que ser iguales.

La aplicabilidad del principio de máxima entropía en la distribución de especies está respaldada por las teorías de la termodinámica de los procesos ecológicos. La segunda ley de la termodinámica dice que en un sistema sin influencias ajenas los procesos se mueven en la dirección de la máxima entropía. Entonces, en la ausencia de influencias fuera de aquellas incluidas por las limitantes en el modelo, la distribución geográfica de una especie va a tender hacia la distribución de máxima entropía [Phillips, Anderson, Schapire, 2006].

#### 4.5 EVALUACIÓN DE LOS MODELOS

La efectividad de un método para predecir la posible distribución de una especie está fundamentada en la evaluación del error, tal y como se menciona en [Mateo, Felicísimo, Muñoz, 2011]. La evaluación es medir la proporción de casos clasificados incorrectamente por el modelo. Hay dos tipos de error:

- El de comisión, que consiste en clasificar una ausencia como presencia.
- El de omisión, que consiste en clasificar una presencia como ausencia.

La evaluación de un modelo y las comparaciones entre ellos se hacen con base en métodos estadísticos que miden su desempeño y consistencia en cuanto a su capacidad para discriminar entre los datos de entrada (datos de presencias y ausencias o pseudoausencias) y datos independientes de contraste. Normalmente para la construcción del modelo se utiliza un porcentaje de la totalidad de ocurrencias de las especies en estudio y el resto se utiliza para realizar pruebas.

Según se menciona en [Mateo, Felicísimo, Muñoz, 2011] las dos propiedades más deseables de estos métodos estadísticos para la evaluación de los modelos son la capacidad de describir el desempeño y la consistencia en términos de ajuste al modelo; esto junto a la capacidad de medir si un modelo es generalizable o no.

El método estadístico más empleado para comparar los métodos es el área bajo la curva AUC del análisis ROC, este es el método que se va a utilizar en esta tesis. Algunos autores, como por ejemplo [Peterson, Papes, Soberón, 2008] mencionan que al usar la curva AUC para comparar dos métodos en ocasiones no es justo. Esto debido a que algunos métodos generan predicciones para toda el área de estudio, como por ejemplo el *Maxent*, pero otros como *GARP* no generan predicciones en algunas áreas límites; ya sea por diseño o por la construcción en sí del algoritmo. En nuestro caso esos problemas en el análisis ROC no nos afectan ya que vamos a comparar dos variantes del mismo método, el *Maxent*.

La curva ROC es la representación gráfica de la capacidad discriminativa de un modelo para todos sus posibles puntos de corte, para esta curva se necesitan datos de presencia/ausencia. El eje "x" corresponde al error de omisión, y el cual muestra la sensibilidad a casos positivos bien clasificados, y el eje "y" al error de comisión, el cual evidencia los falsos positivos. En el punto 0,0 están los casos donde no se tuvo error de comisión. Es decir, no hay falsos positivos, pero al no clasificar tampoco ningún caso positivo correctamente sería un método inútil; de igual modo, el punto 1,1 representa a los casos donde todos los casos positivos fueron bien clasificados, pero también se clasificaron como positivos los demás casos, lo que también sería un método inútil. Lo que se busca es estar en el punto 0,1 que representa el ideal, donde todos los casos se clasificaron correctamente.

El valor de la curva AUC está comprendido entre 0 y 1, un valor de 1 indica que todos los casos se clasificaron correctamente y un valor de 0.5 indica que el modelo es igual que clasificar los casos al azar, por eso un valor menor a 0.5 nos dice que el modelo es realmente malo.

Las mayores ventajas del análisis ROC según [Mateo, Felicísimo, Muñoz, 2011] son:

- La posibilidad de usarlo para comparar cualquier método.
- Sus resultados son independientes de la prevalencia.
- Es una medida independiente del punto de corte.

En la figura 5 se muestra un ejemplo de una curva AUC:

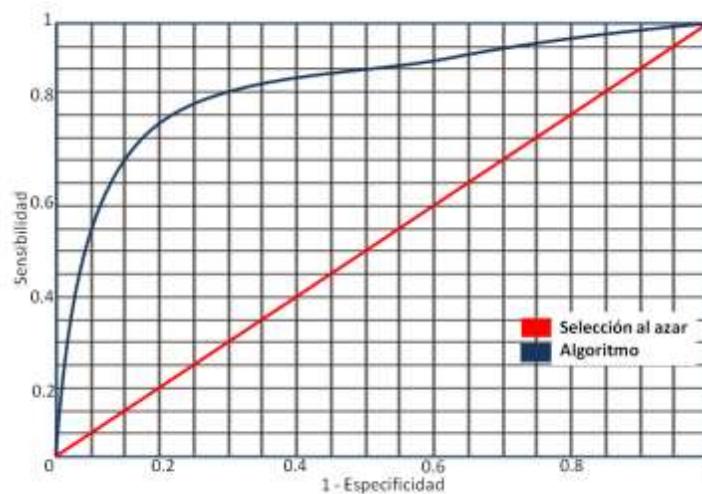


Figura 5: ejemplo de curva AUC

Como se observa en el ejemplo anterior, se tiene en el eje de las "x" la especificidad que representa el error de comisión (falsos positivos), y en el eje "y" la sensibilidad que representa el error de omisión (verdaderos positivos). También, se puede notar la diagonal que muestra el caso donde se usa el azar para realizar la clasificación. Cualquier cifra bajo la diagonal corresponde a un algoritmo realmente malo, entre más se aleja la curva de la diagonal y más se acerca al punto 0,1 mejor es el algoritmo para clasificar.

#### 4.5.1 APLICACIÓN DEL ANÁLISIS ROC CON DATOS DE SOLO PRESENCIA

Como se explicó en la sección anterior, el análisis ROC está diseñado para discernir entre datos de presencia y ausencia, por lo que podría ser inútil para los casos donde solo se tienen datos de presencia. Sin embargo, como se explica en [Phillips, Anderson, Schapire, 2006] se puede adaptar el análisis ROC para que funcione cuando se tiene solo datos de presencia. Para esto, se puede considerar un problema de clasificación diferente, en este se debe distinguir las presencias de los resultados al azar. Formalmente, para cada pixel  $x$  en el área de estudio, se puede definir una instancia negativa  $x_{\text{random}}$ . De manera similar, también se puede definir una instancia positiva  $x_{\text{presencia}}$  de cada pixel  $x$  que esté incluido en la verdadera distribución geográfica. Con esto se pueden hacer predicciones para ambos tipos de instancias, las positivas y las negativas (estas son pixeles de fondo escogidos uniformemente al azar). Con esto ya es posible definir

una curva ROC, la cual puede ser analizada utilizando los métodos estadísticos estándar para el análisis ROC. En otras palabras, se están usando pseudoausencias en lugar de las ausencias en el análisis ROC.

Es importante mencionar, como se explica en [Phillips, Anderson, Schapire, 2006] que la manera de usar el método anterior tiene una gran diferencia con respecto al uso estándar: el AUC máximo que se puede alcanzar es menor a 1. Si la distribución de la especie cubre una fracción  $a$  de los píxeles, entonces el máximo AUC alcanzable es exactamente  $1 - a/2$ . Desafortunadamente, es normal no conocer el valor de  $a$ , por lo tanto no se puede saber cuál cerca se está del óptimo valor del AUC. Sin embargo, de igual manera se puede utilizar el análisis para realizar la comparación de los métodos y la predicción al azar aún corresponde a un AUC de 0.5.

#### 4.6. ¿POR QUÉ LA SELECCIÓN DE *MAXENT*?

En esta propuesta se seleccionó a *Maxent* como el método para probar la hipótesis, debido a que en la literatura revisada es uno de los métodos de mejor rendimiento. Su efectividad se ha puesto a prueba desde diferentes puntos de vista, por ejemplo en [Elith, Graham, 2009] se hace una comparación de diferentes modelos de distribución de especies a partir de tres aspectos: la capacidad de determinar cómo se relacionan las especies con su ambiente, de predecir cuales áreas del medio ambiente son más o menos apropiadas para la especie (al crear un mapa de aptitud relativa) y su capacidad de extrapolación (predecir la distribución de una especie en áreas con pocas o ninguna ocurrencia). En general, en ese estudio *Maxent* tuvo uno de los mejores rendimientos.

Otra comparación de métodos para modelar la distribución de especies se da en [Peterson, Papes, Eaton, 2007], en este caso se compara *Maxent* contra *GARP*, otro método que tuvo un buen rendimiento. En este artículo se comparó los dos métodos y se evaluó su capacidad de extrapolación; basados en el análisis ROC, *Maxent* tuvo un mejor AUC por lo que su rendimiento fue mejor.

Como estos, existen otros artículos que realizan comparaciones, y en general *Maxent* siempre está dentro de los de mejor rendimiento, esta es una de las razones por las que se escogió para realizar las pruebas en esta propuesta. Otra razón es porque como se menciona en [Mateo, Felicísimo, Muñoz, 2011], uno de los problemas más comunes en el modelado de la distribución geográfica de una especie es que normalmente solo se cuenta con datos de presencia, lo que limita el uso de técnicas estadísticas estándar. *Maxent* es uno de los métodos que puede trabajar solo con datos de presencia, lo que le da una ventaja.

## 5. RELACIONES BIOLÓGICAS ENTRE ESPECIES

Dentro de los ecosistemas los organismos mantienen un complejo conjunto de relaciones con los individuos de su propia especie (relaciones intraespecíficas) e individuos de otras especies (relaciones interespecíficas).

Estas interacciones pueden ser favorables o desfavorables para las especies relacionadas, por lo que afectan su supervivencia y ayudan a determinar su área de distribución [García, 2011]. Además, estas interacciones desempeñan un papel esencial en la evolución de las especies; por ejemplo, las presas "aprenden" a ocultarse o a defenderse de sus depredadores, y los depredadores desarrollan nuevas maneras para cazar a sus presas.

Debido a la gran influencia que pueden tener las relaciones interespecíficas en la distribución de las especies es que en esta propuesta se incluye esta información en el método *Maxent* para la predicción de la distribución de las especies.

Las relaciones interespecíficas son muy diversas y pueden ser muy complejas, entre ellas se encuentran [García, 2011]:

- Depredación: donde una especie caza a otra para poder alimentarse. Por ejemplo, los leones cazan gacelas.
- Parasitismo: es un tipo de relación donde una especie, el parásito, se alimenta o reproduce a costa de otro, el hospedador. Por ejemplo, los tipos de plantas (como la *Orobancha rapum-genistae*) que se adhieren a árboles y absorben sus nutrientes.
- Explotación: este tipo de interacción se da cuando una especie obtiene beneficio de otra que no recibe nada a cambio. Por ejemplo, las rémoras de mar, las cuales se adhieren a otros animales y les restringen su movilidad sin dar nada a cambio.
- Mutualismo: este tipo de interacción se da cuando ambas especies se benefician. Por ejemplo, las hormigas (*Pseudomyrmex*), las cuales viven en las espinas ahuecadas de las acacias; así las hormigas reciben un lugar donde vivir y al mismo tiempo la acacia se protege de otros depredadores.
- Simbiosis: es un tipo de mutualismo, pero existe una relación permanente y muy estrecha entre individuos de distintas especies que en algunos casos pueden llegar a fundirse en un organismo único.
- Comensalismo: es un tipo de relación donde una especie se beneficia y a la otra le es indiferente. Por ejemplo, los buitres que esperan a que los depredadores dejen de comer y se alimentan de los restos de comida.
- Competencia: es un tipo de interacción que produce un enfrentamiento de especies similares, esto provoca un reparto de recursos que perjudica a ambas especies.
- Inquilinismo: es un tipo de relación donde una especie vive en la madriguera o refugio de otro. El huésped generalmente no es afectado por el inquilino, sin embargo, el inquilino sí se beneficia de tener el refugio.

En la tabla 1 se muestra un resumen tomado de [García, 2011] donde se muestra la relación de beneficio/perjuicio de las distintas relaciones interespecíficas. Un "+" significa que la especie se beneficia de la relación, un "-" significa que la especie se ve perjudicada y un "0" significa que a la especie le es indiferente:

Tabla 1: tipos de relaciones interespecíficas

Tipo de relación	Especie 1	Especie 2
Depredación	+	-
Parasitismo	+	-
Explotación	+	-
Mutualismo	+	+
Simbiosis	+	+/-
Comensalismo	+	0
Inquilinismo	+	0
Competencia	0	0

Como se nota en la tabla anterior, según el tipo de relación interespecífica se tienen diferentes tipos de influencias, las que intervienen en la posible distribución geográfica de una especie; por ejemplo, cuando se tiene una relación como el inquilinismo es probable que si se encuentra una especie A en su madriguera, también se encuentre la especie B. De igual manera, este tipo de relación es más notoria en relaciones como el mutualismo y la simbiosis.

En esta tesis solo se va a tomar en cuenta la relación que se da entre aves que se alimentan de diferentes especies de árboles, esto por motivos de tiempo y simplicidad. Además, también se ha buscado especies donde exista un grado de dependencia relativamente alto; esto para mostrar el potencial que puede tener la inclusión de las relaciones interespecíficas con el fin de mejorar la eficacia de métodos como el *Maxent*. A pesar de que en esta tesis solo se toma un tipo de relación, todas las relaciones interespecíficas pueden afectar la distribución geográfica de una especie, ya que de una u otra forma influyen en la especie con la cual interacciona. Debido a esto, explorar las demás relaciones interespecíficas puede ser de gran utilidad para futuros estudios de su influencia en los modelos de distribución de especies.

## 6. PROPUESTA DE TESIS

### 6.1 HIPÓTESIS

Al añadir las relaciones biológicas entre las especies (depredación, competencia, simbiosis, entre otras) al método *Maxent* es posible mejorar su capacidad predictiva para obtener la posible distribución geográfica de una especie.

## 7. OBJETIVOS Y ALCANCE

### 7.1 OBJETIVO GENERAL

Incrementar la capacidad para predecir la potencial distribución geográfica para una especie del método *Maxent*, esto mediante la adición de las relaciones de interacción biológica de tipo interespecíficas.

### 7.2 OBJETIVOS ESPECÍFICOS

- Definir las especies de estudio, donde estas especies tengan una relación muy cercana con otras especies, por lo cual exista dependencia entre ellas.
- Recolección de los datos de las especies y las variables medioambientales.
- Medir la efectividad de predicción de los modelos de distribución geográfica generados por el método *Maxent*, esto utilizando las variables ambientales seleccionadas y empleando como parámetro de aceptación el área bajo la curva del análisis ROC.
- Definir una metodología para incluir relaciones biológicas de tipo inter-específicas, que representen una mejora en la capacidad predictiva para la identificación de la distribución geográfica de una especie, empleando el método *Maxent*.
- Construir modelos de distribución geográfica con la incorporación de las mejoras propuestas al método *Maxent*, que demuestren un mejor desempeño cuantitativo del software en relación con los modelos generados por *Maxent* sin relaciones interespecíficas.

### 7.3 ALCANCES Y LIMITACIONES

Con esta investigación se busca probar que la inclusión de las relaciones interespecíficas mejora el rendimiento de un modelo de distribución de especies tal como el *Maxent*. Actualmente, esto no se hace ya que se considera una tarea difícil, principalmente porque no existe información documentada accesible (una base de datos) acerca de las diferentes relaciones entre las especies. Esta información se encuentra principalmente en libros y artículos. En esta investigación no se incluye un procedimiento para obtener estos datos ya que esto es un problema complejo que amerita una investigación independiente.

Como resultado de esta investigación se van a generar dos modelos de distribución de especies para cada especie de estudio: uno generado con el *Maxent* original o línea base y otro con la mejora propuesta en este documento. Los entregables van a ser: el documento de tesis, donde se detalle el resultado de la investigación, sus conclusiones y los modelos de distribución de especies, tanto generados con el método original como con los modelos generados con la mejora.

Debido a limitaciones de tiempo, no se van a producir materiales fuera de los entregables propuestos.

## 8. RECOLECCIÓN DE DATOS

### 8.1 ESPECIES SELECCIONADAS

Los datos de ocurrencias de las especies seleccionadas se han tomado principalmente de GBIF. La Infraestructura Mundial de Información en Biodiversidad (GBIF por sus siglas en inglés) es una infraestructura abierta fundada por gobiernos, donde cualquiera tiene acceso a toda la información de especies en la Tierra, sin límites de naciones, todo compartido a través de internet [GBIF, 2015]. En esta tesis también se realizó un experimento utilizando los nidos de lapa verde. En este caso la información fue suministrada por el doctor Olivier Chassot del Centro Científico Tropical (CCT). En el siguiente enlace se puede acceder a la página principal de GBIF, donde se pueden descargar los datos de ocurrencias para las especies seleccionadas: <http://www.gbif.org/>. Para todos los casos se utilizó el filtro de país con el valor de Costa Rica, esto para delimitar el área de estudio solo a nivel nacional.

Además, cabe destacar que para mejorar la calidad de los datos se eliminaron todas las ocurrencias sin datos de georeferenciación o con latitud y longitud con valor de 0. En el caso de la información de nidos de lapa verde solo se procesaron los nidos en almendro de montaña y la información de georeferenciación fue transformada del formato horas, minutos y segundos a direcciones decimales, las cuales son las utilizadas por *Maxent* y otros programas para el manejo de este tipo de datos. La información de los nidos de lapa verde se encuentra en los apéndices.

#### 8.1.1 *ARA AMBIGUUS*

La lapa verde es un ave de la familia *Psittacidae*, mide unos 80 centímetros de alto y pesa unos 1,442 gramos en promedio. De setiembre a abril, se alimenta principalmente del fruto de almendro de montaña (*Dipteryx panamensis*); cuando los frutos del almendro no están disponibles, se nutre principalmente del fruto del titor. Hay un breve período cuando ninguno de estos dos frutos está disponible, en ese lapso come de otras 37 especies de frutos. La lapa verde con frecuencia utiliza el mismo nido siempre y se mueve unas 550 hectáreas alrededor de su nido [García, 2011].

Como se puede apreciar en la información anterior, existe una relación cercana entre la lapa verde y el almendro de montaña, por lo cual se han incluido estas dos especies en la investigación. Es importante aclarar que los modelos generados solo serán representativos de los meses entre setiembre y abril, ya que luego de este período las aves migran. Además, en esta tesis se van a generar modelos para dos tipos de relaciones entre la lapa verde y el almendro de montaña: una es como alimentación y otra como refugio para anidar. En el caso de los nidos, se dispone de información tomada del Centro Científico Tropical, en el cual se tiene información de unos 90 nidos en almendros de montaña. Esto en combinación del rango hogareño será utilizado para la construcción de los modelos.

En la figura 6 se muestra una imagen de la lapa verde:



Figura 6: Ara ambiguus. Tomada de <http://atta2.inbio.ac.cr/>

#### 8.1.2 ALMENDRO DE MONTAÑA (*DIPTERYX PANAMENSIS*)

El almendro de montaña es un árbol con alturas que van de 15 a 50 metros y con diámetros de 1 a 1.5 metros. Es una especie endémica en Costa Rica y Panamá, además se encuentra en zonas bajas y planicies de la costa atlántica. Se encuentran en altitudes desde los 20 a los 1300 metros sobre el nivel del mar [Jiménes *et all*, 2011].

En la figura 7 se muestra una imagen de la especie:



Figura 7: Dipteryx panamensis. Tomada de <http://atta2.inbio.ac.cr/>

---

### 8.1.3 QUETZAL (*PHAROMACHRUS MOCINNO*)

El quetzal es un ave de la familia *Trogonidae*, mide unos 36 cm y tiene hasta 64 cm correspondientes a la cola del macho. Los quetzales prefieren los bosques montanos muy húmedos, así como los bosques lluviosos del Caribe. Se ubican normalmente en las montañas cubiertas total o parcialmente de bosque de la Cordillera de Tilarán hasta Panamá [Luis, 1999a].

Los quetzales se alimentan de frutos, insectos pequeños, ranas, lagartijas pequeñas y caracoles. Sus frutos favoritos incluyen varias especies de la familia *Lauraceae* y también *Symplocos* [Luis, 1999a].

Según [Ávila, Hernández, Velarde, 1996], en el bosque lluvioso de montaña de la reserva El Triunfo en México, la dieta de los quetzales depende en un 63% de los frutos de la familia *Lauraceae*. Estos pueden volar hasta 3 km de distancia para alimentarse. Este es un dato importante ya en esta tesis se van a utilizar las ocurrencias de las diferentes especies de la familia *Lauraceae*, y el rango de alimentación para construir los diferentes modelos.

En la figura 8 se muestra una imagen del quetzal:



Figura 8: quetzal por Luis Vargas. Tomada de <http://atta2.inbio.ac.cr/>

---

### 8.1.4 PÁJARO CAMPANA (*PROCNIAS TRICARUNCULATUS*)

El pájaro campana es un miembro de la familia *Cotingidae*. El macho mide 30 cm y pesa 220 gramos, la hembra mide 25 cm y pesa 145 gramos. El macho adulto presenta toda la cabeza, cuello y pecho de color blanco, el resto del plumaje es café, más oscuro por debajo. La hembra es verde oliva por encima y amarilla por debajo [Luis, 1999b].

El pájaro campana vive en los niveles medios y altos de los bosques montanos en la época reproductiva; también, en algunas ocasiones visitan bosques secundarios y árboles altos en bosques semiabiertos. Se alimentan principalmente de frutos de especies de *Lauraceae*, dos terceras partes de los 307 tipos de frutos con los cuales se alimentan en el área de la Reserva Biológica Bosque Nuboso Monteverde, corresponden a especies de la familia *Lauraceae* [Luis, 1999b].

En la figura 9 se muestra una imagen del pájaro campana:



Figura 9: pájaro campana por Fernando Zeledón. Tomada de <http://atta2.inbio.ac.cr/>

#### 8.1.5 FAMILIA LAURACEAE

La familia de plantas *Lauraceae* tiene alrededor de 2000 a 2500 especies distribuidas en 54 géneros; pueden ser tanto árboles como arbustos y son frecuentemente aromáticos. Además, están distribuidas principalmente en el Trópico [Watson, Dallwithz, 1992].

Tanto el quetzal como el pájaro campana se alimentan de diferentes especies de esta familia de plantas. Por ejemplo, uno de los frutos favoritos del quetzal es el aguacatillo (*Persea caerulea*), por esta razón esta familia de plantas es de gran relevancia para la investigación.

En la figura 10 se muestra una imagen de la planta de aguacatillo:



Figura 10: aguacatillo. Tomada de <http://www.elhogarnatural.com/>

## 8.2 INFORMACIÓN CLIMÁTICA

Para la construcción de los modelos de distribución geográfica de las especies necesitamos la información climática en un formato de matriz que *Maxent* pueda utilizar. Por esta razón, se utilizó el conjunto de datos , para el tiempo presente (el cual incluye información climática desde 1950 hasta el 2000). Este conjunto de datos se puede obtener gratuitamente en el siguiente enlace: <http://www.worldclim.org/current>. Existen dos versiones: la matriz genérica y la matriz ESRI. En esta tesis se utilizó la matriz de ESRI como herramienta de apoyo ARCMAP, pues es compatible con esta. Es importante también aclarar que se utilizó el conjunto de datos de "30-arc-seconds", con los cuales se tienen celdas de un tamaño cercano a un kilómetro cuadrado. Esta es la resolución más alta ofrecida en el *WorldClim*. A continuación, se lista las diferentes variables climáticas utilizadas en la presente tesis:

- Temperatura anual media.
- Rango de temperaturas diurnas medias.
- Isotérmico.
- Estacionalidad de la temperatura.
- Temperatura máxima del mes más cálido.
- Temperatura mínima del mes más frío.
- Rango de temperaturas anuales.
- Temperatura media del cuatrimestre más lluvioso.
- Temperatura media del cuatrimestre más seco.
- Temperatura media del cuatrimestre más cálido.
- Temperatura media del cuatrimestre más frío.
- Precipitación anual.
- Precipitación del mes más seco.
- Precipitación del mes más húmedo.
- Estacionalidad de la precipitación.
- Precipitación del cuatrimestre más húmedo.
- Precipitación del cuatrimestre más seco.
- Precipitación del cuatrimestre más cálido.
- Precipitación del cuatrimestre más frío.

Cabe destacar que los datos del *WorldClim* son para toda la Tierra, en esta tesis el área de estudio se delimitó a Costa Rica. Por este motivo, se utilizó *ArcMap* para tomar la información solo para Costa Rica. Por lo tanto, se crea un subconjunto de datos y se guarda en un formato de *ASCII grid*, el cual puede ser utilizado por *Maxent*. En la figura 11 se muestra una imagen de ejemplo sobre la matriz de precipitación anual para Costa Rica:



Figura 11: precipitación anual para Costa Rica

## 9. METODOLOGÍA PARA LA INCLUSIÓN DE LAS RELACIONES INTERESPECÍFICAS EN MAXENT

En la presente investigación, se utilizó un rango de influencia para la inclusión de las relaciones interespecíficas. El rango de influencia es el área donde la presencia de una especie incide sobre otra. Por ejemplo, un almendro amarillo puede ejercer influencia sobre una lapa verde solo si este es accesible para el ave. Por lo tanto, depende de la relación que existe entre las especies y las capacidades de cada una de ellas. Además, se tiene un ejemplo de un comportamiento especial de la lapa verde con respecto a su anidación; según el doctor Olivier Chassot del CCT las lapas verdes no se mueven más de 500 hectáreas de su nido, el cual es normalmente un agujero en árboles de almendro amarillo; estos a su vez sirven de alimentación.

*Maxent* utiliza *ASCII grids* como entrada para la generación de modelos; cada capa de datos climáticos está también en este formato, por lo que la introducción de la nueva capa de "rango de influencia" en forma de *ASCII grid* es natural. Para la construcción de esta nueva capa se cuenta con las ocurrencias de las especies relacionadas y con el rango que se ha definido para la relación. Además, se sabe que cada celda de la matriz que representa el *ASCII grid* mide aproximadamente un kilómetro cuadrado para los experimentos de esta tesis. Otra característica importante de los *ASCII grids* es que las matrices se encuentran georeferenciadas. Con esta información es posible entonces construir la nueva capa de rango de influencia. Cada ocurrencia tiene una latitud y una longitud, y a su vez las celdas de la matriz también tienen una latitud y una longitud inicial. Además, sabemos que todas las celdas tienen un tamaño fijo, debido a esto es posible saber en qué celda de la matriz se encuentra cada ocurrencia.

A continuación un ejemplo de un encabezado de un *ASCII grid*:

```
ncols    982
nrows    655
xllcorner -88.125439160921
yllcorner 6.3629258146449
cellsize  0.00833333337679505
NODATA_value -9999
```

Las propiedades *xllcorner* y *yllcorner* representan la esquina inferior izquierda de la matriz, la propiedad *cellsize* representa el tamaño de cada celda y las propiedades *ncols* y *nrows* representan el tamaño de la matriz. Al utilizar esta información es posible saber exactamente las coordenadas geográficas de cada celda. En la figura 12 se muestra una imagen que representa gráficamente la matriz:

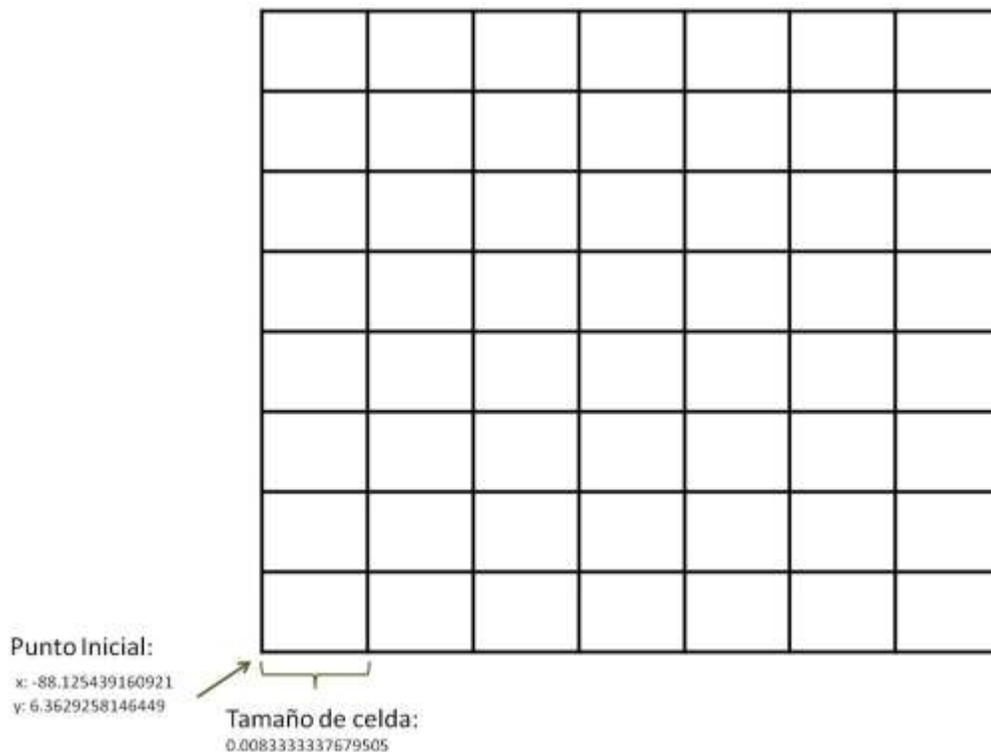


Figura 12: representación de *ASCII grid*

Una vez que se ha definido la celda donde una ocurrencia se encuentra, se puede empezar a construir la nueva capa de rango de influencia. El *ASCII grid* en un inicio tiene dos tipos de valores en sus celdas, el "NODATA" y 0. La diferencia entre el "sin datos" y el "0" es que el valor "sin datos" representa a las celdas fuera del área de estudio y cualquier otro valor significa que una celda si es parte del área de estudio.

Durante la construcción de la capa de rango de influencia, si una celda contiene una ocurrencia a su valor se le suma 1; además, según el rango de influencia a las celdas vecinas también se les suma 1, esto de una manera recursiva. Por ejemplo, si el rango de influencia definido para el experimento es 2, entonces la celda que contiene la ocurrencia y todas las celdas alrededor de ella se les suma 1 a su valor actual. El rango de influencia depende de la relación de cada especie, anteriormente se mencionó que la lapa verde no se aleja más de 500 hectáreas de su nido; 500 hectáreas son alrededor de 5 kilómetros cuadrados, por lo que para este caso el rango de influencia sería 5.

En la figura 13 se muestra un ejemplo de un rango de influencia de 3:

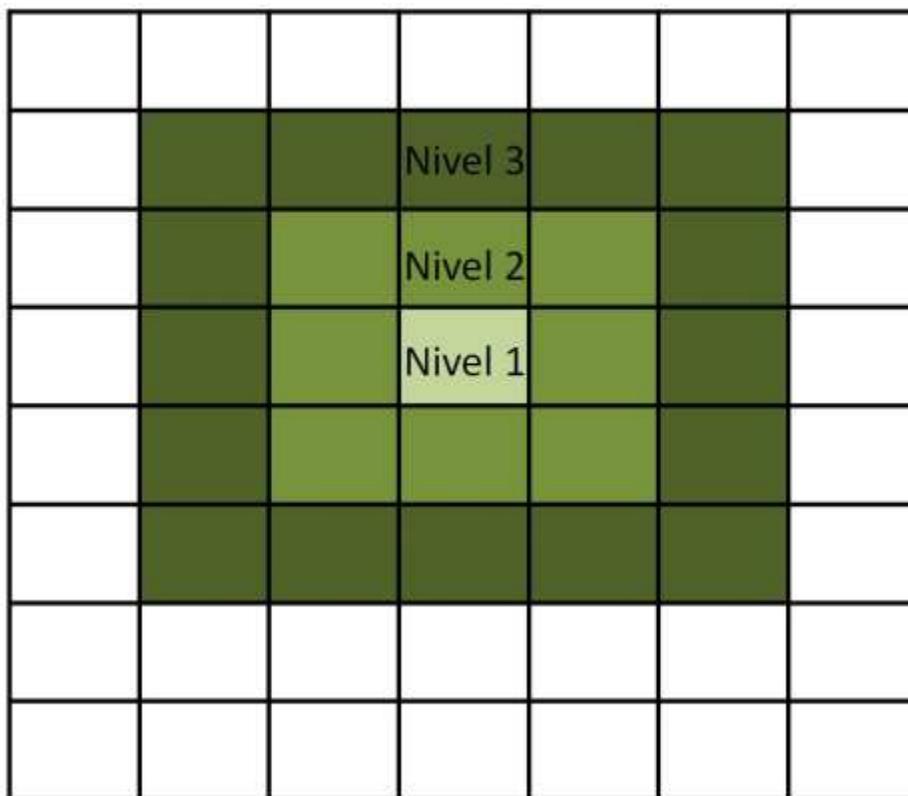


Figura 13: ejemplo niveles de influencia

Ya que en el mundo real una especie puede ser vista varias veces en una zona, puede haber zonas más ricas que otras. Para simular esto en la capa de rango de influencia el algoritmo trabaja recursivamente con las celdas alrededor de la celda donde se encuentra la ocurrencia, esto al sumarle 1 al valor que actualmente posee una celda. Con esto es posible simular esas zonas donde las especies son vistas más veces, ya que una celda puede ser afectada por varias ocurrencias, al aumentar su valor conforme las ocurrencias que se le avecinan.

Como se mencionó anteriormente, al inicio el *ASCII grid* tiene dos tipos de valores: el sin datos y las celdas con valores de 0. Estos valores definen el área de estudio, es por esto que para construir la capa de influencia se necesita esta definición. En esta tesis el área de estudio está definida para Costa Rica y lo que se hizo para darle el valor inicial a la capa de influencia fue usar como base una capa climática y ponerle el valor de 0 a todas las celdas que tenían un valor diferente al "sin datos". Esta capa inicial se hizo con un software en *Java* construido especialmente para esta tesis. El código fuente de ese programa se encuentra

en los apéndices de este texto. En la figura 14 se muestra una imagen de como se ve la capa de rango de influencia al inicio del proceso:



Figura 14: capa de influencia base

Una vez con la base para la capa de rango de influencia se construyó un software especializado que sigue el proceso descrito anteriormente. Primero, lee la capa base, luego lee las ocurrencias de las especies asociadas y calcula en que celdas de la matriz se encuentran esas ocurrencias. A continuación, a partir de la utilización del valor para el rango de influencia definido por parámetro se procede a inspeccionar cada una de las celdas con ocurrencias, y recursivamente le da valor a las celdas que rodean cada una de esas celdas con ocurrencias; esto toma en cuenta el parámetro definido para el rango de influencia. En la figura 15 se muestra un ejemplo de una capa de rango de influencia construida con la información de nidos suministrados por Olivier Chassot del CCT, al utilizar un rango de influencia de 5 km:



Figura 15: capa de rango de influencia de nidos de lapa verde

En la imagen anterior se puede observar el área de estudio, Costa Rica con un color negro. Ese es el valor base para todas las celdas del área de estudio; además, si se observa la zona norte del país se puede ver que hay unos recuadros en blanco, esas son celdas donde una ocurrencia se encuentra. También, si se observa más detalladamente se puede ver que hay zonas grises y otras más blancas, las zonas blancas son zonas donde se encuentra una cantidad mayor de nidos. En la figura 16 se muestra una imagen donde se ve la capa de rango de influencia y las ocurrencias de los nidos de lapa verde:



Figura 16: capa de rango de influencia de nidos de lapa verde con ocurrencias

En la imagen anterior se pueden observar varios puntos rojos, los cuales representan las ocurrencias de nidos de lapa verde; como se puede observar las ocurrencias concuerdan con las zonas donde las celdas tienen valores más altos que 0.

## 10. EXPERIMENTOS REALIZADOS

Para validar la hipótesis planteada en esta tesis se realizaron dos conjuntos de experimentos, esto con las 3 especies de aves incluidas en este estudio y su especie(s) relacionada(s). En el primer conjunto de experimentos se busca comparar la efectividad de *Maxent* para calcular la distribución geográfica de cada ave cuando solo se agregan datos climáticos contra la efectividad de *Maxent* cuando además de esos datos climáticos, se agrega la capa de rango de influencia descrita anteriormente en esta tesis. En el segundo conjunto de experimentos, además de la efectividad de *Maxent* con ambas variaciones se busca medir su capacidad de extrapolación. Para esto se divide el área de estudio en varias secciones y se deja áreas de estudio sin ninguna ocurrencia de las aves. Luego, durante la construcción del modelo se mide que tan efectivo fue *Maxent* para calcular la distribución de la especie para esas áreas donde no se incluyó ninguna ocurrencia de entrenamiento. También, es importante mencionar que se dividió el área de estudio de tal forma que la cantidad de ocurrencias excluidas sea aproximadamente el 50% del total de las ocurrencias para el ave en estudio, siempre garantizando que existan esas áreas grandes sin ocurrencias para el entrenamiento del *Maxent*. En el caso de la evaluación si se utiliza el 100% de las ocurrencias de las aves, así es posible medir la capacidad de extrapolación del *Maxent* al usar solo los datos climáticos contra los modelos que además incluyen la capa de rango de influencia.

Para todos los experimentos se utilizaron los datos climáticos y la información de ocurrencias de las aves de estudio que se mencionan en la sección 8 del presente documento. En las siguientes subsecciones se muestran los resultados obtenidos de cada experimento; en específico los modelos generados por *Maxent*, el gráfico para la curva AUC y las capas más importantes para la generación del modelo. Además, se describe qué componentes se utilizaron para cada uno.

En los modelos generados por *Maxent* los cuadrados morados son ocurrencias utilizadas como prueba y los cuadrados blancos son ocurrencias utilizadas para entrenamiento. La escala indica que los colores más fríos, como el azul, son zonas poco aptas para que la especie en estudio sobreviva. Por el contrario, los colores más cálidos, como el rojo, son zonas donde las condiciones son más favorables para que la especie en estudio se desarrolle.

### 10.1 EFECTIVIDAD DE *MAXENT* USANDO SOLO DATOS CLIMÁTICOS CONTRA DATOS CLIMÁTICOS Y CAPA DE RANGO DE INFLUENCIA

#### 10.1.1 LAPA VERDE (*ARA AMBIGUUS*) SOLO CON DATOS CLIMÁTICOS

Para este experimento las ocurrencias de lapa verde se obtuvieron de GBIF. Se utilizaron el 70% del total de ocurrencias para entrenamiento y 30% para pruebas (los porcentajes se configuran en la aplicación creada por los autores de *Maxent*). Los datos climáticos fueron obtenidos del *WorldClim*, como se describe en la sección 8.

En la figura 17 se muestra el modelo de distribución geográfica generado por *Maxent*:

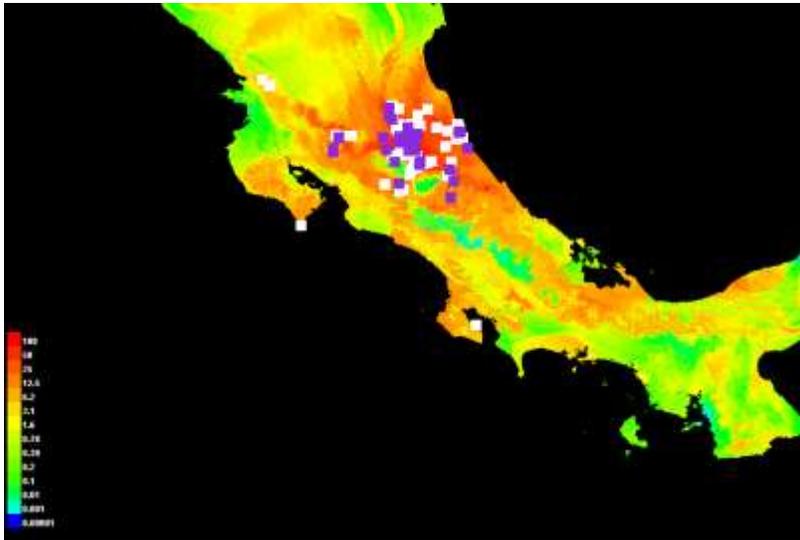


Figura 17: distribución de lapa verde usando solo datos climáticos

En la figura 18 se muestra el gráfico AUC obtenido para el modelo anterior:

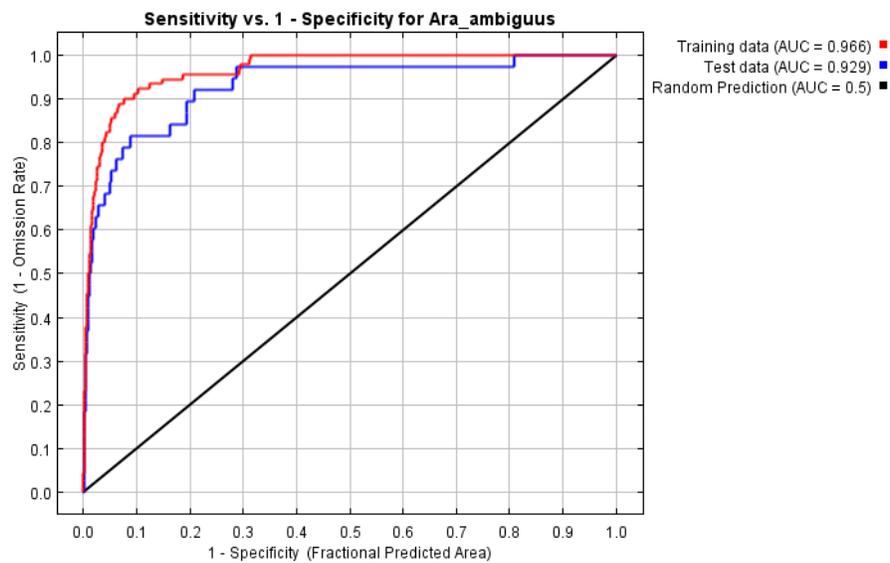


Figura 18: gráfico AUC para la distribución de lapa verde usando solo datos climáticos

Como se puede observar en el gráfico anterior, el modelo obtuvo un AUC de 0.966 para los datos de entrenamiento y un 0.929 para los datos de pruebas.

Según los resultados dados por Maxent para este modelo las dos principales capas son:

1. La precipitación anual con un 27.9%.
2. La precipitación del mes más seco con un 18%.
3. La precipitación del cuatrimestre más seco con un 12.4%.

### 10.1.2 LAPA VERDE (ARA AMBIGUUS) CON DATOS CLIMÁTICOS Y LA CAPA DE RANGO DE INFLUENCIA DE ALMENDROS AMARILLOS COMO ALIMENTACIÓN

Para este experimento se construyó una capa de rango de influencia con la información de almendros amarillos obtenidos de GBIF; de allí se eliminaron las ocurrencias cuya latitud y longitud fuera 0 o no tuvieran esta información. También, para la capa de rango de influencia se utilizó 5 como parámetro para su construcción. Esto significa que cada ocurrencia de almendro amarillo podría ser detectada por una lapa verde si esta se encuentra a una distancia no mayor a 5 kilómetros. El parámetro de influencia con el valor de 5 kilómetros fue utilizado ya que mediante experimentación se encontró que al utilizar ese valor se obtienen los mejores resultados para este modelo. En el caso de las ocurrencias de lapa verde, del total se utilizaron el 70% para entrenamiento y 30% para pruebas (los porcentajes se configuran en la aplicación creada por los autores de *Maxent*). Los datos climáticos fueron obtenidos del *WorldClim* como se describe en la sección 8.

En la figura 19 se muestra el modelo generado por *Maxent* para este caso:

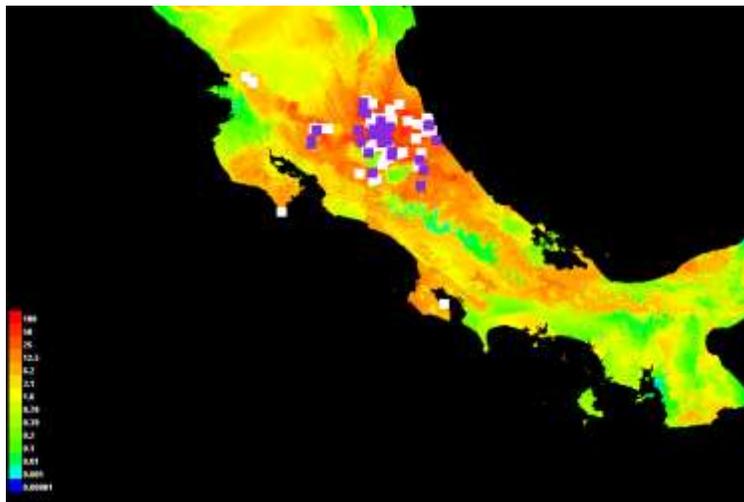


Figura 19: distribución de lapa verde usando datos climáticos y capa de rango de influencia

En la figura 20 se muestra el gráfico AUC obtenido para el modelo anterior:

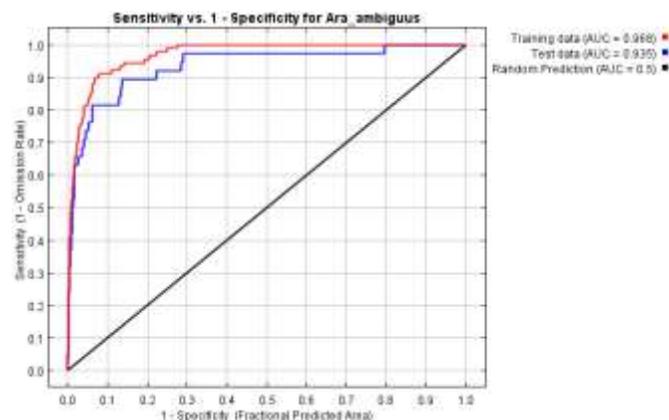


Figura 20: gráfico AUC para la distribución de lapa verde usando datos climáticos y capa de rango de influencia

Como se puede observar en la gráfica anterior para los datos de entrenamiento el modelo obtuvo un AUC de 0.968 y para los datos de prueba el modelo obtuvo un AUC de 0.935.

Las capas más importantes para la construcción del modelo fueron las siguientes:

1. La precipitación anual con un 23.4%.
2. La precipitación del mes más seco con un 22.5%.
3. La capa de rango de influencia con un 12.9%.

### 10.1.3 LAPA VERDE (ARA AMBIGUUS) CON DATOS CLIMÁTICOS Y LA CAPA DE RANGO DE INFLUENCIA DE ALMENDROS AMARILLOS COMO NIDOS

Para este experimento las ocurrencias de Lapa Verde se obtuvieron de GBIF. Se utilizaron del total de ocurrencias el 70% para entrenamiento y 30% para pruebas (los porcentajes se configuran en la aplicación creada por los autores de *Maxent*). Los datos climáticos fueron obtenidos del *WorldClim* como se describe en la sección 8. Para el caso de la capa de rango de influencia como nido, la información de nidos de lapa verde fue suministrada por el doctor Olivier Chassot del CCT. El parámetro de rango de influencia utilizado fue 5, esto ya que según indica el doctor Chassot existe evidencia de que las lapas verdes no se alejan más de 500 hectáreas de sus nidos, 500 hectáreas son alrededor de 5 kilómetros cuadrados; de ahí la utilización de 5 como parámetro.

En la figura 21 se muestra el modelo generado por *Maxent* para este experimento:

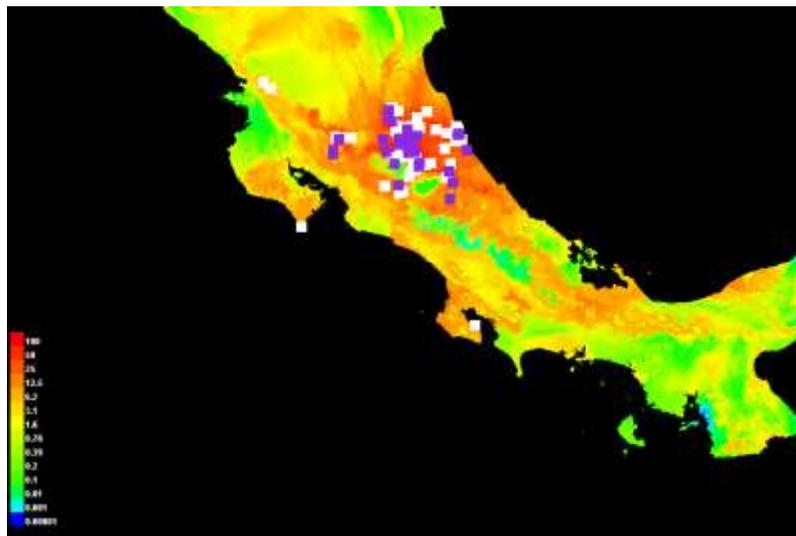


Figura 21: modelo de distribución de la lapa verde usando los datos climáticos y la capa de rango de influencia (nidos)

En la figura 22 se muestra el gráfico AUC obtenido para el modelo anterior:

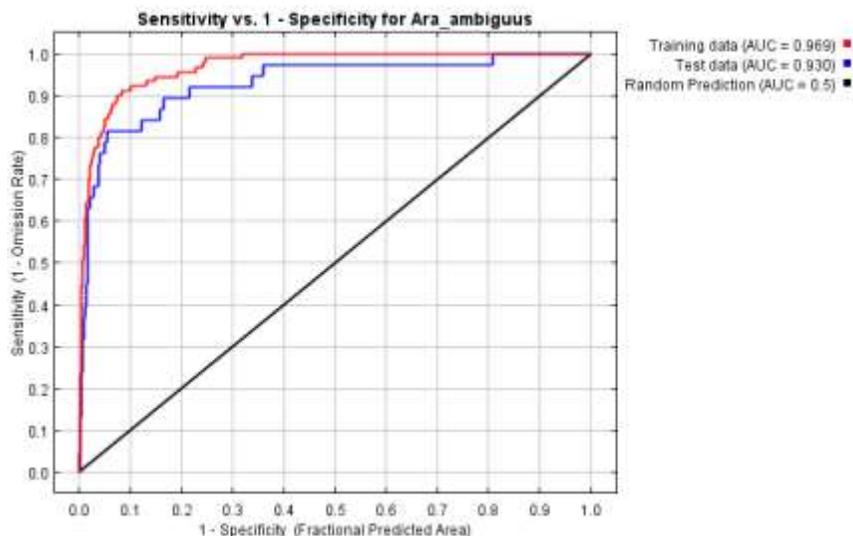


Figura 22: gráfico AUC para el modelo de distribución de la lapa vede usando los datos climáticos y la capa de rango de influencia (nidos)

Como se puede observar en el gráfico anterior el modelo obtuvo un AUC de 0.969 para los datos de entrenamiento y un AUC de 0.930 para los datos de prueba.

Las capas más importantes para la construcción del modelo fueron las siguientes:

1. La precipitación anual con un 28.4%.
2. La precipitación del mes más seco con un 20.2%.
3. La precipitación del cuatrimestre más seco con un 9.6%.

#### 10.1.4 PÁJARO CAMPANA (PROCNIAS TRICARUNCULATUS) UTILIZANDO SOLO DATOS CLIMÁTICOS

Para este experimento las ocurrencias del pájaro campana se obtuvieron de GBIF, del total de ocurrencias se utilizaron el 70% para entrenamiento y 30% para pruebas (los porcentajes se configuran en la aplicación creada por los autores de *Maxent*). Los datos climáticos fueron obtenidos del *WorldClim* como se describe en la sección 8.

En la figura 23 se muestra el modelo de distribución geográfica generado por *Maxent*:

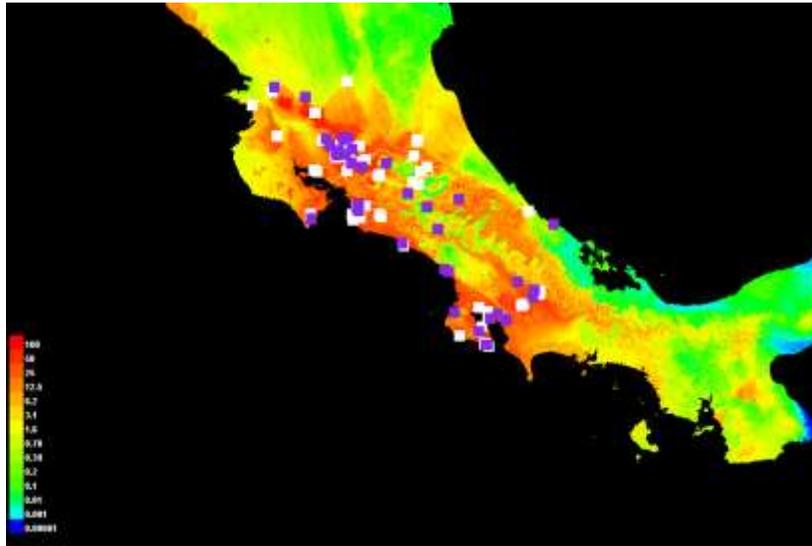


Figura 23: distribución del pájaro campana usando solo datos climáticos

En la figura 24 se muestra el gráfico AUC obtenido para el modelo anterior:

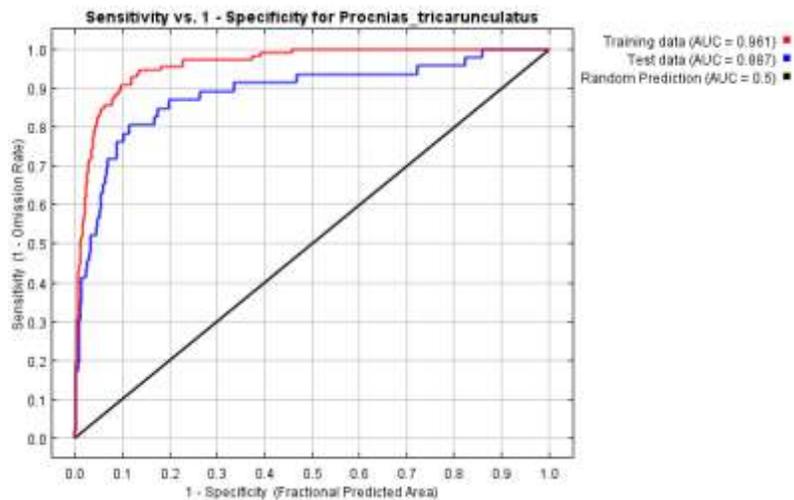


Figura 24: gráfico AUC para la distribución del pájaro campana usando solo datos climáticos

Como se puede observar en el gráfico anterior el modelo obtuvo un AUC de 0.961 para los datos de entrenamiento y un AUC de 0.887 para los datos de prueba.

Las capas más importantes para la construcción del modelo fueron las siguientes:

1. El rango anual de temperaturas con un 18.6%.
2. La precipitación del cuatrimestre más cálido con un 17.5%.
3. La temperatura media del cuatrimestre más frío con un 12%.



En la figura 26 se muestra el gráfico AUC obtenido para el modelo anterior:

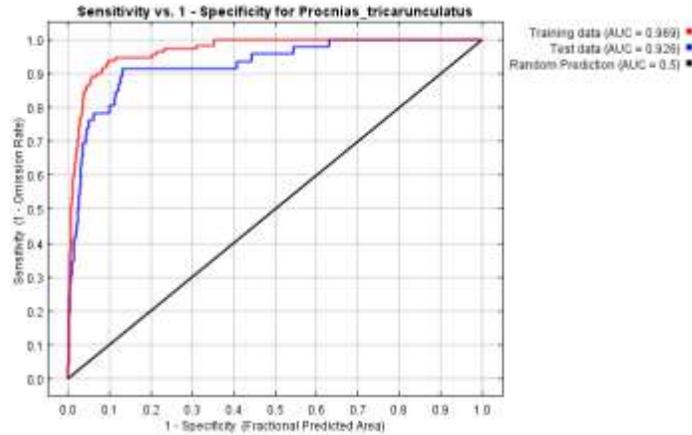


Figura 26: gráfico AUC para el modelo de distribución del pájaro campana usando los datos climáticos y la capa de rango de influencia (*Lauraceae*)

Como se puede observar en el gráfico anterior el modelo obtuvo un AUC de 0.969 para los datos de entrenamiento y un AUC de 0.926 para los datos de prueba.

Las capas más importantes para la construcción del modelo fueron las siguientes:

1. La capa de rango de influencia con un 64.1%.
2. La temperatura media del cuatrimestre más cálido con un 15.1%.
3. El rango anual de temperaturas con un 7.2%.

#### 10.1.6 QUETZAL (*PHAROMACHRUS MOCINNO*) UTILIZANDO SOLO DATOS CLIMÁTICOS

Para este experimento las ocurrencias del quetzal se obtuvieron de GBIF; del total de ocurrencias se utilizaron el 70% para entrenamiento y 30% para pruebas (Los porcentajes se configuran en la aplicación creada por los autores de *Maxent*). Los datos climáticos fueron obtenidos del *WorldClim* como se describe en la sección 8.

En la figura 27 se muestra el modelo de distribución geográfica generado por *Maxent*:

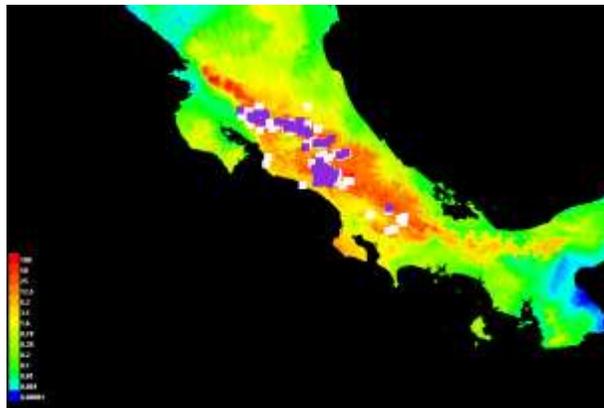


Figura 27: distribución del quetzal usando solo datos climáticos

En la figura 28 se muestra el gráfico AUC obtenido para el modelo anterior:

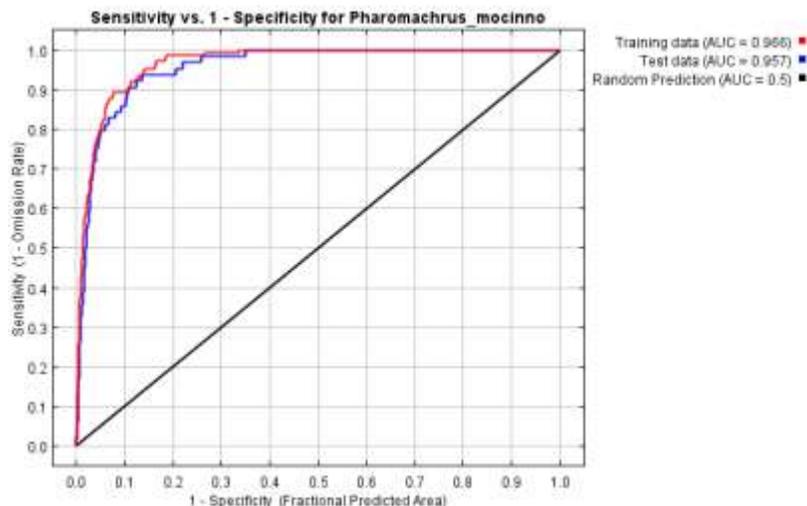


Figura 28: gráfico AUC para la distribución del quetzal usando solo datos climáticos

Como se puede observar en el gráfico anterior el modelo obtuvo un AUC de 0.966 para los datos de entrenamiento y un AUC de 0.957 para los datos de prueba.

Las capas más importantes para la construcción del modelo fueron las siguientes:

1. La temperatura media del cuatrimestre más frío con un 38.1%.
2. La temperatura mínima del mes más frío con un 16.3%.
3. La temperatura media del mes más seco con un 8.1%.

#### 10.1.7 QUETZAL (*PHAROMACHRUS MOCINNO*) UTILIZANDO DATOS CLIMÁTICOS Y LA CAPA DE RANGO DE INFLUENCIA (*LAURACEAE*)

Para este experimento se construyó una capa de rango de influencia con la información de especies pertenecientes a la familia *Lauraceae* obtenidos de GBIF; de las ocurrencias de la familia *Lauraceae* se eliminaron las ocurrencias cuya latitud y longitud fuera 0 o no tuvieran esta información. Además, para la capa de rango de influencia se utilizó 5 como parámetro para su construcción. El parámetro de influencia con el valor de 5 kilómetros fue utilizado ya que mediante experimentación se encontró que al utilizar ese valor se obtienen los mejores resultados para este modelo. En el caso de las ocurrencias del quetzal, del total de ocurrencias se utilizaron el 70% para entrenamiento y 30% para pruebas (los porcentajes se configuran en la aplicación creada por los autores de *Maxent*). Los datos climáticos fueron obtenidos del *WorldClim* como se describe en la sección 8.

En la figura 29 se muestra el modelo generado por *Maxent* para este caso:

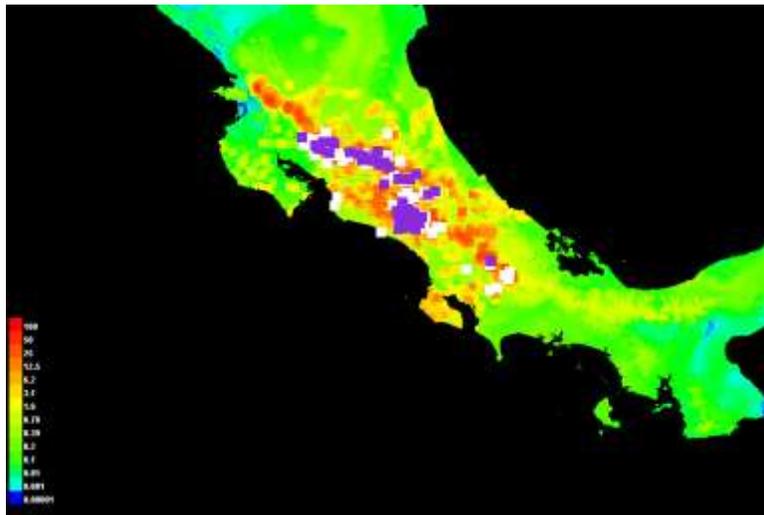


Figura 29: modelo de distribución del quetzal usando los datos climáticos y la capa de rango de influencia (*Lauraceae*)

En la figura 30 se muestra el gráfico AUC obtenido para el modelo anterior:

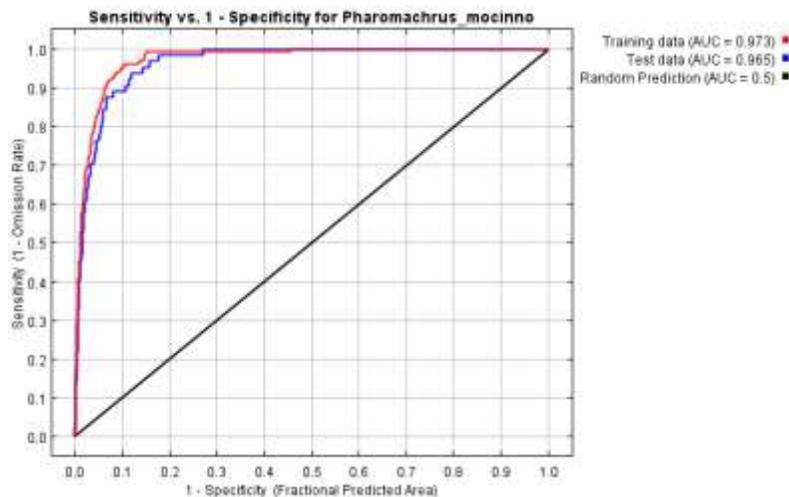


Figura 30: gráfico AUC para el modelo de distribución del quetzal usando los datos climáticos y la capa de rango de influencia (*Lauraceae*)

Como se puede observar en el gráfico anterior el modelo obtuvo un AUC de 0.973 para los datos de entrenamiento y un AUC de 0.965 para los datos de prueba.

Las capas más importantes para la construcción del modelo fueron las siguientes:

1. La capa de rango de influencia con un 47.3%.
2. La temperatura media del cuatrimestre más frío con un 10.7%.
3. La temperatura mínima del mes más frío con un 8.7%.

## 10.2 CAPACIDAD DE EXTRAPOLACIÓN DE MAXENT AL USAR SOLO DATOS CLIMÁTICOS CONTRA USAR DATOS CLIMÁTICOS MÁS EL RANGO DE INFLUENCIA

### 10.2.1 LAPA VERDE USANDO EL 50% DE OCURRENCIAS Y SOLO DATOS CLIMÁTICOS

Para este experimento los datos de ocurrencias de la Lapa verde se tomaron de GBIF; se dividió el área de estudio para simular las áreas sin muestreos de la siguiente manera:



Figura 31: división del área de estudio para la lapa verde

En la figura 31 se nota que el área de estudio se dividió en 4 cuadrantes, los puntos rojos son las ocurrencias de la lapa verde tomados de GBIF. Solo las ocurrencias que se encuentran en los cuadrantes superior izquierdo e inferior derecho fueron tomados en cuenta. De esta manera, se logró tomar aproximadamente el 50% de las ocurrencias para entrenamiento y se dejaron grandes zonas del área de estudio sin muestras. En la figura 32 se evidencia una imagen del área de estudio solo con las ocurrencias utilizadas para entrenamiento:



Figura 32: ocurrencias de entrenamiento de la lapa verde utilizadas en el experimento de capacidad de extrapolación

En la figura 32 los puntos verdes son las ocurrencias utilizadas en el entrenamiento. En el caso de la evaluación del modelo se utilizó el 100% de las ocurrencias de lapa verde obtenidas de GBIF.

En el caso de la capa de rango de influencia, las ocurrencias de almendros amarillos fueron obtenidas de GBIF. De estas se eliminaron las que tenían latitud y longitud 0 o no tuvieran esta información. Además, para la capa de rango de influencia se utilizó 5 como parámetro para su construcción; esto quiere decir que cada ocurrencia de almendro amarillo podría ser detectada por una lapa verde si esta se encuentra a una distancia no mayor a 5 kilómetros. El parámetro de influencia con el valor de 5 kilómetros fue utilizado ya que se encontró mediante experimentación que al utilizar ese valor se obtienen los mejores resultados para este modelo. Los datos climáticos fueron obtenidos del *WorldClim* como se describe en la sección 8.

En la figura 33 se muestra el modelo generado por *Maxent* para este caso:

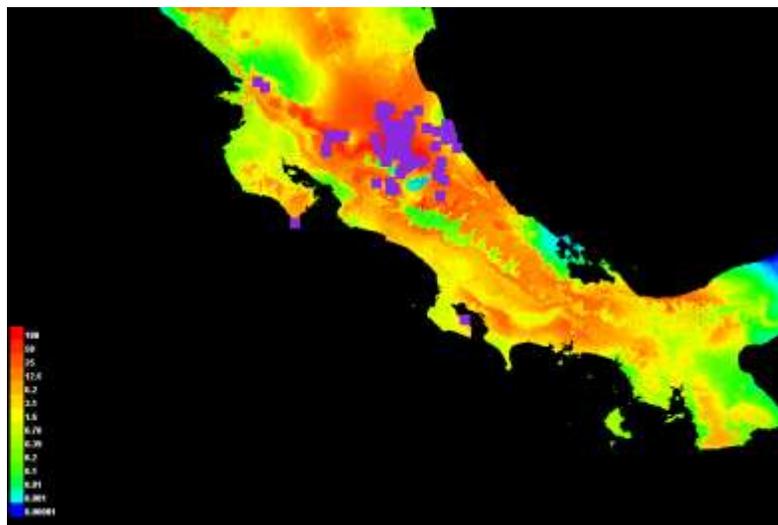


Figura 33: modelo de distribución para la lapa verde usando el 50% de las ocurrencias

En la figura 34 se muestra el gráfico AUC obtenido para el modelo anterior:

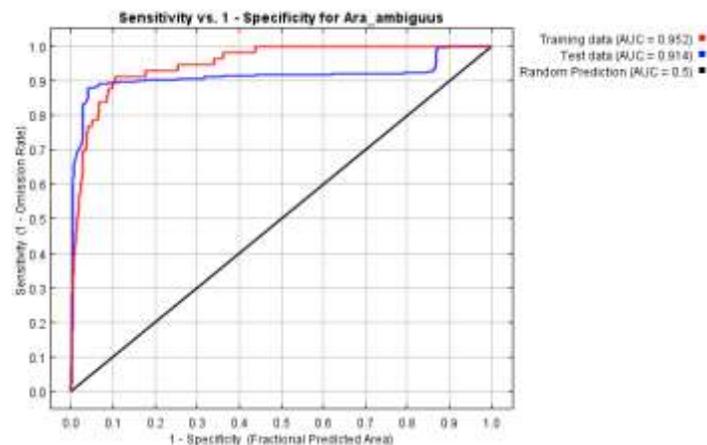


Figura 34: gráfico AUC para el modelo de distribución para la lapa verde usando el 50% de las ocurrencias

Como se puede observar en el gráfico anterior el modelo obtuvo un AUC de 0.952 para los datos de entrenamiento y un AUC de 0.914 para los datos de prueba.

Las capas más importantes para la construcción del modelo fueron las siguientes:

1. Precipitación de cuatrimestre más húmedo con un 46%.
2. Rango anual de temperaturas con un 19.5%.
3. Precipitación de cuatrimestre más húmedo con un 8%.

### 10.2.2 LAPA VERDE USANDO EL 50% DE OCURRENCIAS Y DATOS CLIMÁTICOS EN COMBINACIÓN CON LA CAPA DE RANGO DE INFLUENCIA (ALIMENTACIÓN)

Para este experimento se construyó una capa de rango de influencia con la información de almendros amarillos obtenidos de GBIF, de las ocurrencias aportadas de GBIF se eliminaron aquellas cuya latitud y longitud fuera 0 o no tuvieran esta información. Además, para la capa de rango de influencia se utilizó 5 como parámetro para su construcción. El parámetro de influencia con el valor de 5 kilómetros fue utilizado ya que mediante experimentación se encontró que al utilizar ese valor se obtienen los mejores resultados para este modelo. En el caso de las ocurrencias de lapa verde se usaron las mismas ocurrencias descritas en la sección 10.2.1. Los datos climáticos fueron obtenidos del *WorldClim* como se describe en la sección 8.

En la figura 35 se muestra el modelo generado por *Maxent* para este caso:

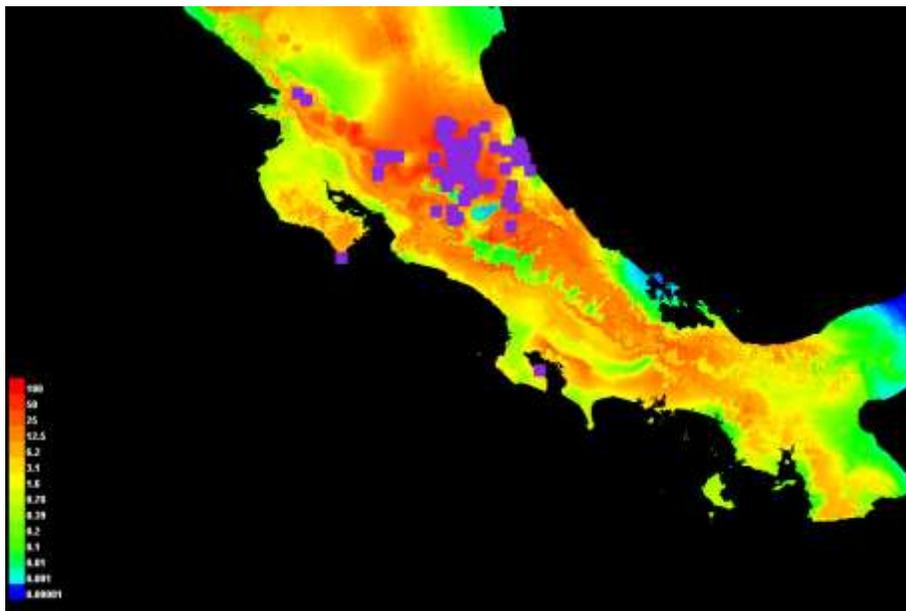


Figura 35: modelo de distribución para la lapa verde usando el 50% de las ocurrencias y la capa de rango de Influencia (alimentación)

En la figura 36 se muestra el gráfico AUC obtenido para el modelo anterior:

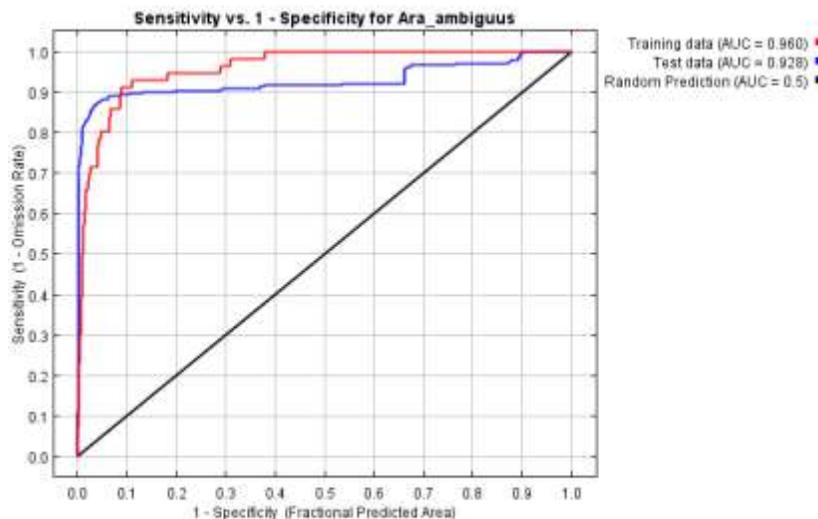


Figura 36: gráfico AUC para el modelo de distribución para la lapa verde usando el 50% de las ocurrencias y la capa de rango de influencia (alimentación)

Como se puede observar en el gráfico anterior el modelo obtuvo un AUC de 0.952 para los datos de entrenamiento y un AUC de 0.914 para los datos de prueba.

Las capas más importantes para la construcción del modelo fueron las siguientes:

1. La capa de rango de influencia con un 38.8%.
2. Precipitación del cuatrimestre más seco 27.8%.
3. Rango de temperaturas anual 18.1%.

### 10.2.3 LAPA VERDE USANDO EL 50% DE OCURRENCIAS Y DATOS CLIMÁTICOS EN COMBINACIÓN CON LA CAPA DE RANGO DE INFLUENCIA (NIDOS)

Para la construcción de la capa de rango de influencia como nido, la información de nidos de lapa verde fue suministrada por el doctor Olivier Chassot del CCT. El parámetro de rango de influencia utilizado fue 5, ya que el doctor Olivier Chassot menciona que existe evidencia de que las lapas verdes no se alejan más de 500 hectáreas de sus nidos (alrededor de 5 kilómetros cuadrados), de ahí la utilización de 5 como parámetro. En el caso de las ocurrencias de lapa verde se usaron las mismas ocurrencias descritas en la sección 10.2.1. Los datos climáticos fueron obtenidos del *WorldClim* como se describe en la sección 8.

En la figura 37 se muestra el modelo generado por *Maxent* para este experimento:

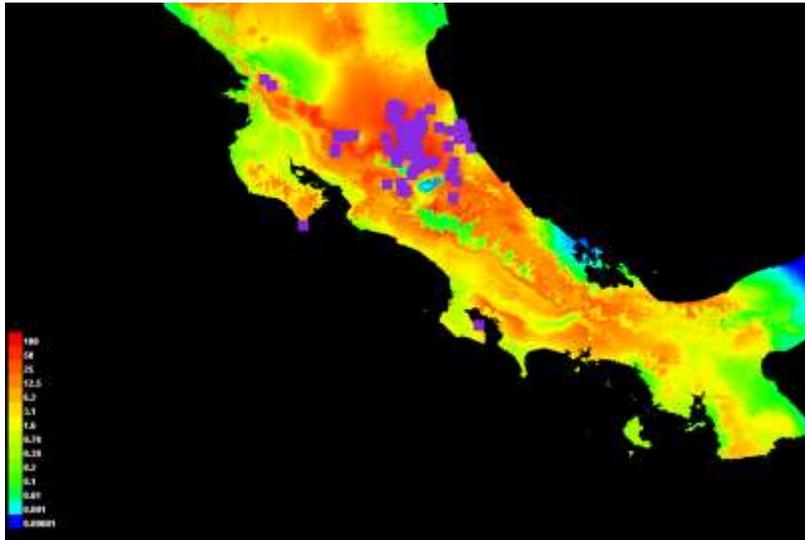


Figura 37: modelo de distribución para la lapa verde usando el 50% de las ocurrencias y la capa de rango de influencia (nidos)

En la figura 38 se muestra el gráfico AUC obtenido para el modelo anterior:

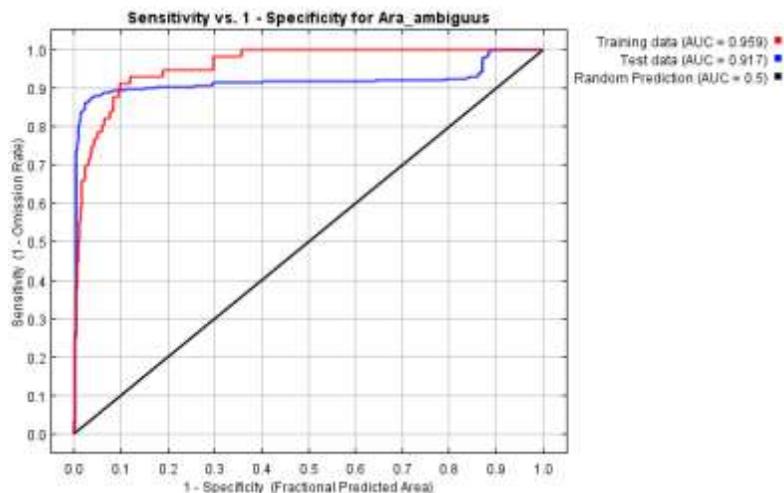


Figura 38: gráfico AUC para el modelo de distribución para la lapa verde usando el 50% de las ocurrencias y la capa de rango de influencia (nidos)

Como se puede observar en el gráfico anterior el modelo obtuvo un AUC de 0.959 para los datos de entrenamiento y un AUC de 0.917 para los datos de prueba.

Las capas más importantes para la construcción del modelo fueron las siguientes:

1. Precipitación del cuatrimestre más seco con 47.7%.
2. Rango anual de temperatura con 24.8%.
3. Precipitación del mes más seco con 7.6%.

#### 10.2.4 PÁJARO CAMPANA USANDO EL 50% DE OCURRENCIAS Y SOLO DATOS CLIMÁTICOS

Para este experimento los datos de ocurrencias del pájaro campana se tomaron de GBIF. Se dividió el área de estudio para simular las áreas sin muestreos de la siguiente manera:

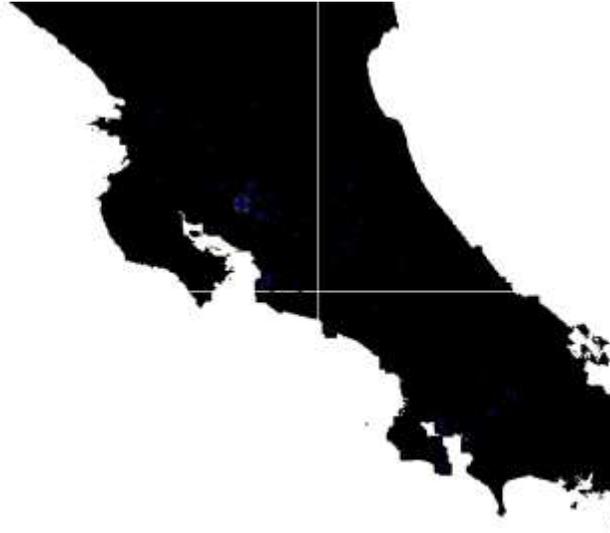


Figura 39: división del área de estudio para el pájaro campana

En la figura 39 se puede notar que el área de estudio se dividió en 4 cuadrantes, los puntos azules son las ocurrencias del pájaro campana tomadas de GBIF. Solo las ocurrencias que se encuentran en los cuadrantes superior derecho e inferior derecho fueron tomadas en cuenta. De esta manera, se logró tomar aproximadamente el 50% de las ocurrencias para entrenamiento y se dejaron grandes zonas del área de estudio sin muestras. A continuación se muestra una imagen del área de estudio solo con las ocurrencias utilizadas para entrenamiento:

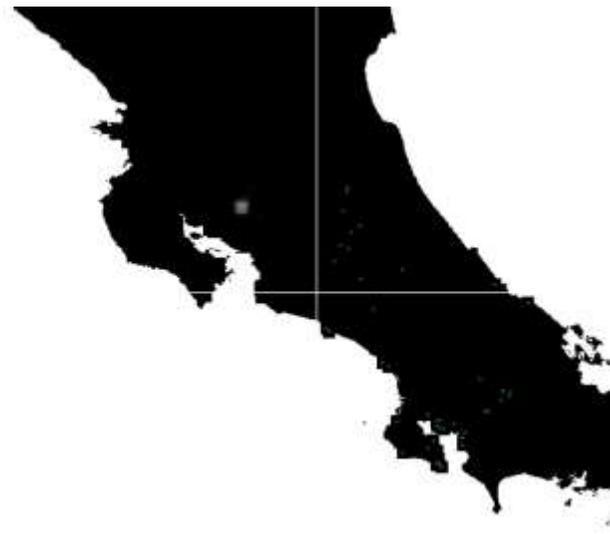


Figura 40: ocurrencias de entrenamiento del pájaro campana utilizadas en el experimento de capacidad de extrapolación

En la figura 40 los puntos verdes son las ocurrencias utilizadas en el entrenamiento; en el caso de la evaluación del modelo se utilizó el 100% de las ocurrencias de pájaro campana obtenidas de GBIF.

En la figura 41 se muestra el modelo generado por *Maxent* para este caso:

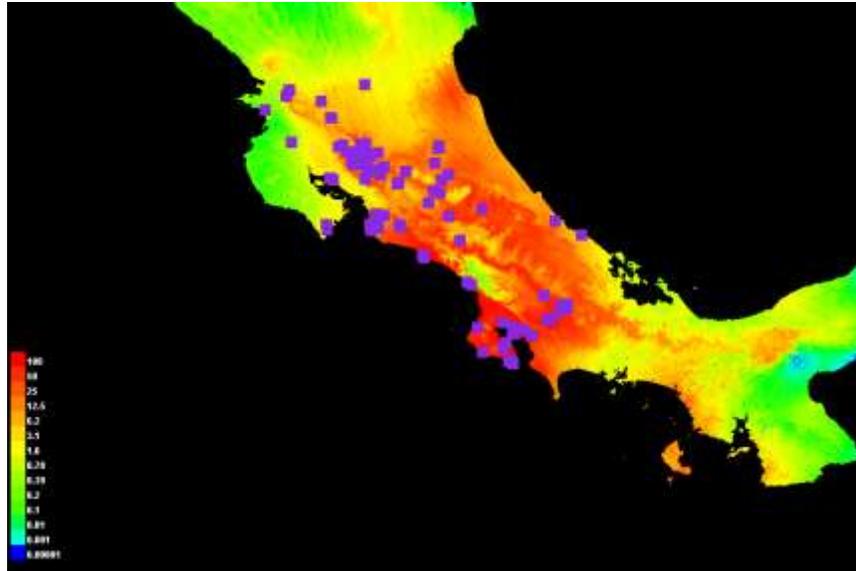


Figura 41: modelo de distribución para el pájaro campana usando el 50% de las ocurrencias

En la figura 42 se muestra el gráfico AUC obtenido para el modelo anterior:

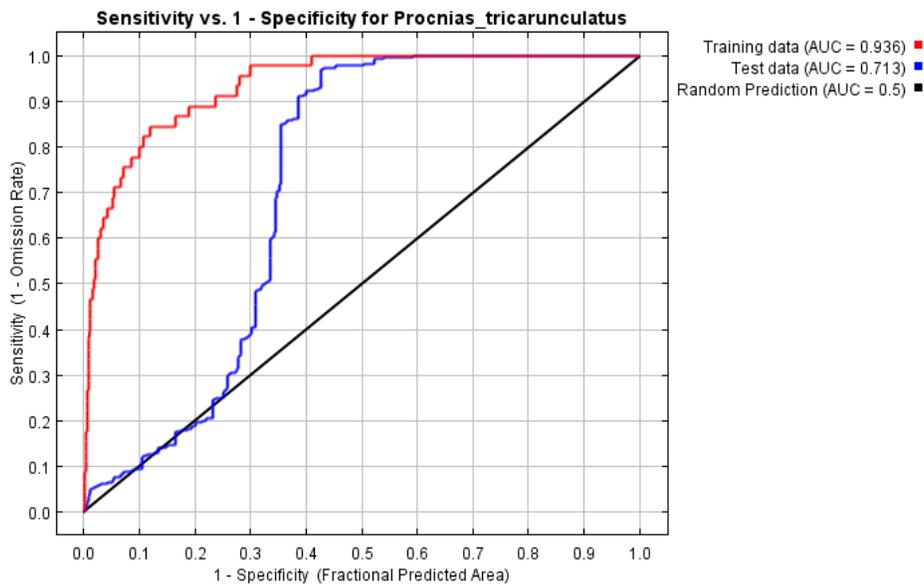


Figura 42: gráfico AUC para el modelo de distribución para el pájaro campana usando el 50% de las ocurrencias

Como se puede observar en el gráfico anterior el modelo obtuvo un AUC de 0.936 para los datos de entrenamiento y un AUC de 0.713 para los datos de prueba.

Las capas más importantes para la construcción del modelo fueron las siguientes:

1. Rango de temperatura medio con 32.9%.
2. Precipitación del mes más seco con 22.4%.
3. Precipitación del mes más húmedo con 16.9%.

#### 10.2.5 PÁJARO CAMPANA USANDO EL 50% DE OCURENCIAS USANDO DATOS CLIMÁTICOS EN COMBINACIÓN CON LA CAPA DE RANGO DE INFLUENCIA

Para este experimento se construyó una capa de rango de influencia con la información de especies pertenecientes a la familia *Lauraceae* obtenidos de GBIF; de las ocurrencias de la familia *Lauraceae* obtenidas de GBIF se eliminaron aquellas cuya latitud y longitud fuera 0 o no tuvieran esta información. Además, para la capa de rango de influencia se utilizó 3 como parámetro para su construcción. El parámetro de influencia con el valor de 3 kilómetros fue utilizado ya que mediante experimentación se encontró que usar ese valor se obtienen los mejores resultados para este modelo. En el caso de las ocurrencias de pájaro campana se utilizaron las mismas descritas en la sección 10.2.4. Los datos climáticos fueron obtenidos del *WorldClim* como se describe en la sección 8.

En la figura 43 se muestra el modelo generado por *Maxent* para este caso:

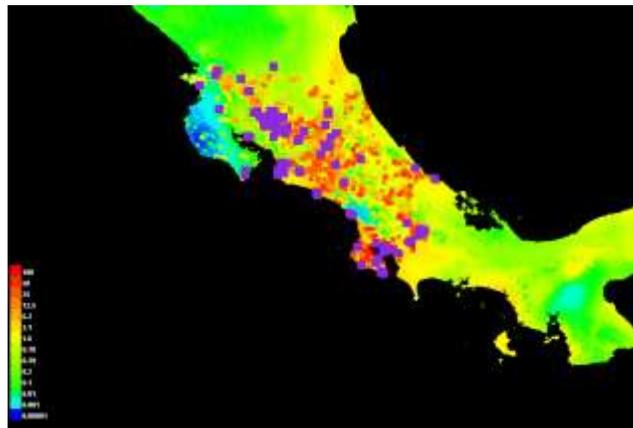


Figura 43: modelo de distribución para el pájaro campana usando el 50% de las ocurrencias y la capa de rango de influencia

En la figura 44 se muestra el gráfico AUC obtenido para el modelo anterior:

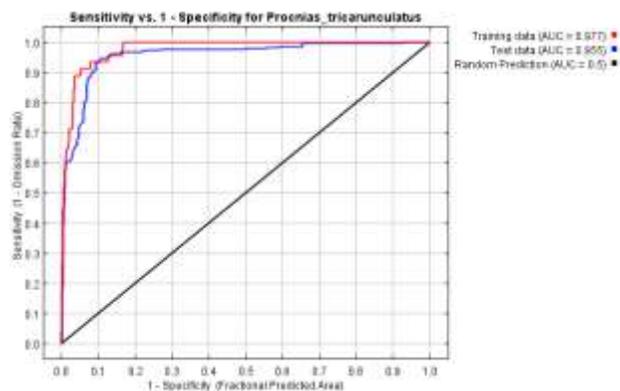


Figura 44: gráfico AUC para el modelo de distribución para el pájaro campana usando el 50% de las ocurrencias y la capa de rango de influencia

Como se puede observar en el gráfico anterior el modelo obtuvo un AUC de 0.977 para los datos de entrenamiento y un AUC de 0.955 para los datos de prueba.

Las capas más importantes para la construcción del modelo fueron las siguientes:

1. Capa de rango de influencia con un 71.9%.
2. Precipitación del mes más seco con un 15.5%.
3. Rango anual de temperaturas con un 2.8%.

#### 10.2.6 QUETZAL USANDO EL 50% DE OCURRENCIAS Y SOLO DATOS CLIMÁTICOS

Para este experimento los datos de ocurrencias del quetzal se tomaron de GBIF; para simular las áreas sin muestreos se dividió el área de estudio de la siguiente manera:

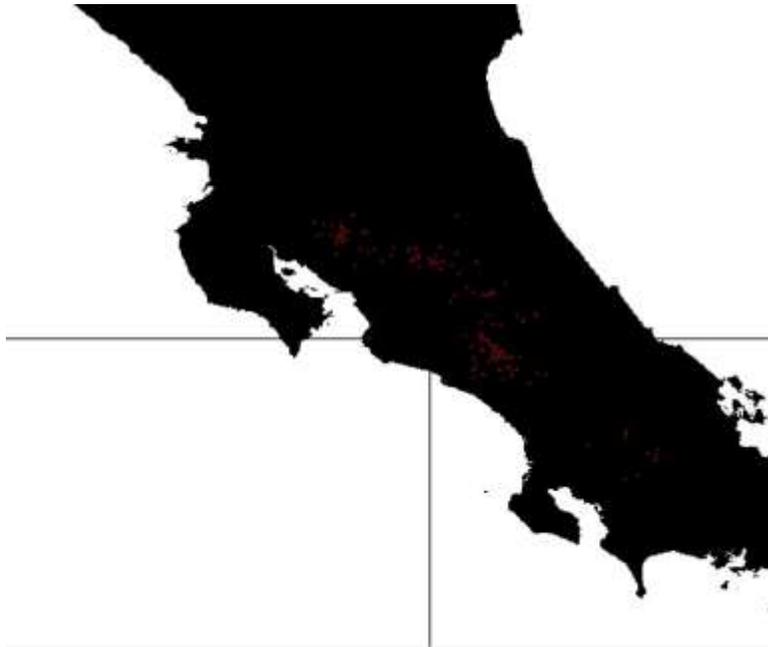


Figura 45: división del área de estudio para el quetzal

En la figura 45 se observa que el área de estudio se dividió en 4 cuadrantes, los puntos rojos son las ocurrencias del quetzal tomados de GBIF. Solo las ocurrencias que se encuentran en los cuadrantes superior izquierdo e inferior izquierdo fueron tomados en cuenta. De esta manera, se logró tomar aproximadamente el 50% de las ocurrencias para entrenamiento y se dejaron grandes zonas del área de estudio sin muestras.

En la figura 46 se muestra una imagen del área de estudio solo con las ocurrencias utilizadas para entrenamiento:



Figura 46: ocurrencias de entrenamiento del quetzal utilizadas en el experimento de capacidad de extrapolación

En la imagen anterior los puntos verdes son las ocurrencias utilizadas en el entrenamiento. En el caso de la evaluación del modelo se utilizó el 100% de las ocurrencias de pájaro campana obtenidas de GBIF. En la figura 47 se muestra el modelo generado por *Maxent* para este caso:

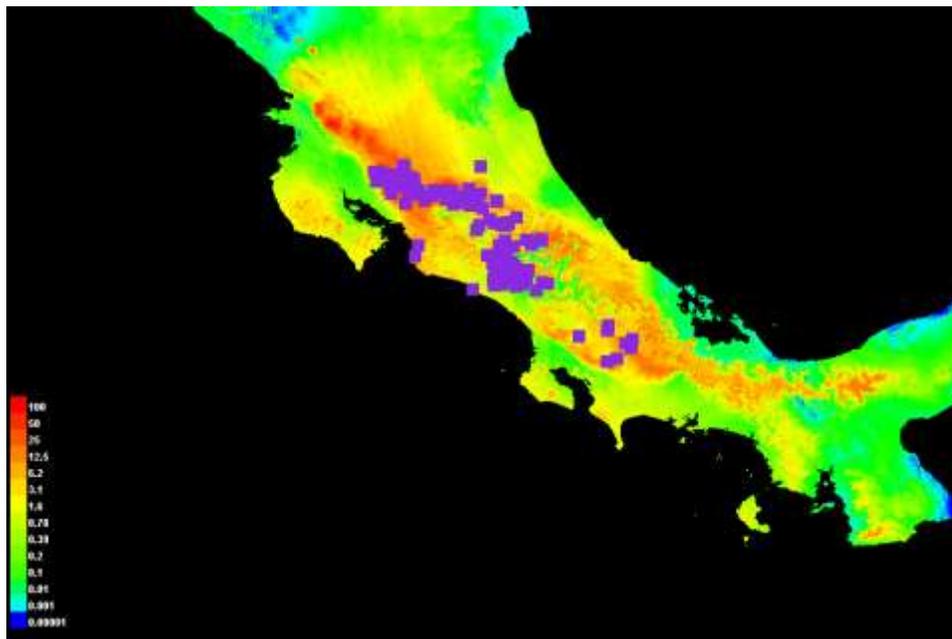


Figura 47: modelo de distribución para el quetzal usando el 50% de las ocurrencias

En la figura 48 se muestra el gráfico AUC obtenido para el modelo anterior:

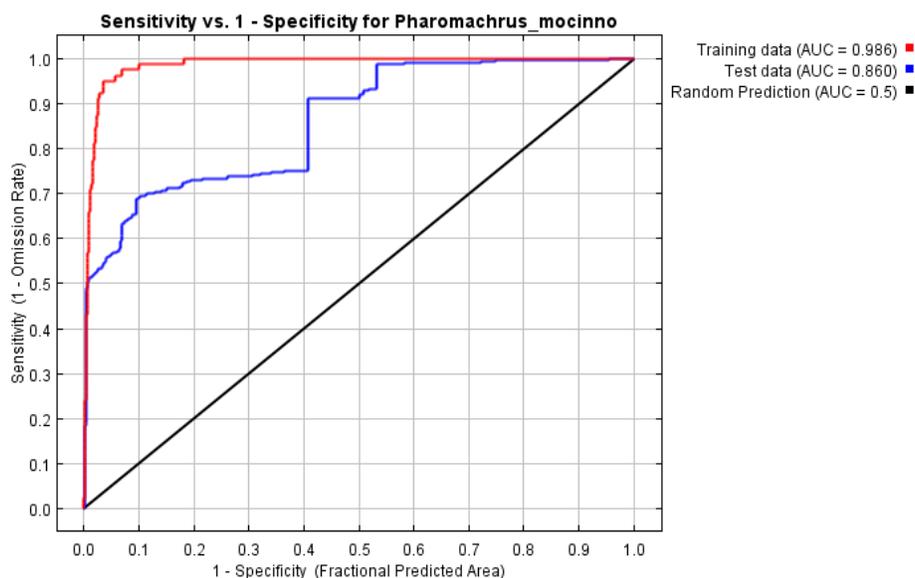


Figura 48: gráfico AUC para el modelo de distribución para el quetzal usando el 50% de las ocurrencias

Como se puede observar en el gráfico anterior el modelo obtuvo un AUC de 0.986 para los datos de entrenamiento y un AUC de 0.860 para los datos de prueba.

Las capas más importantes para la construcción del modelo fueron las siguientes:

1. La temperatura media del cuatrimestre más frío con 55.1%.
2. Isotérmico con 14.2%.
3. La precipitación del cuatrimestre más cálido con 9.8%.

### 10.2.7 QUETZAL USANDO EL 50% DE OCURRENCIAS Y DATOS CLIMÁTICOS EN COMBINACIÓN CON LA CAPA DE RANGO DE INFLUENCIA

Para este experimento se construyó una capa de rango de influencia con la información de especies pertenecientes a la familia *Lauraceae* obtenido de GBIF; de las ocurrencias de la familia *Lauraceae* se eliminaron aquellas cuya latitud y longitud fuera 0 o no tuvieran esta información. Además, para la capa de rango de influencia se utilizó 5 como parámetro para su construcción. El parámetro de influencia con el valor de 5 kilómetros fue utilizado ya que mediante experimentación se encontró que con ese valor se obtienen los mejores resultados para este modelo.

En el caso de las ocurrencias del quetzal se utilizaron las mismas descritas en la sección 10.2.6. Los datos climáticos fueron obtenidos del *WorldClim* como se describe en la sección 8.

En la figura 49 se muestra el modelo generado por *Maxent* para este caso:

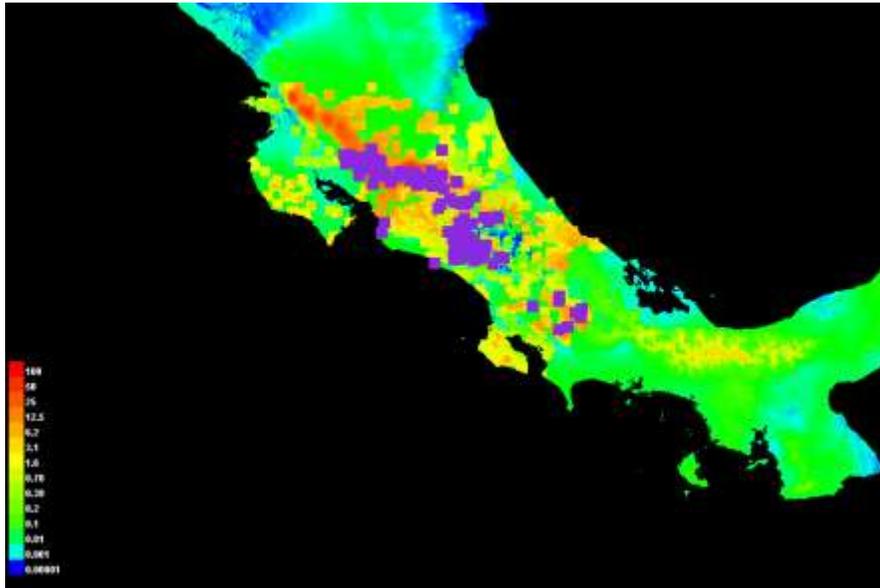


Figura 49: modelo de distribución para el quetzal usando el 50% de las ocurrencias y la capa de rango de influencia

En la figura 50 se muestra el gráfico AUC obtenido para el modelo anterior:

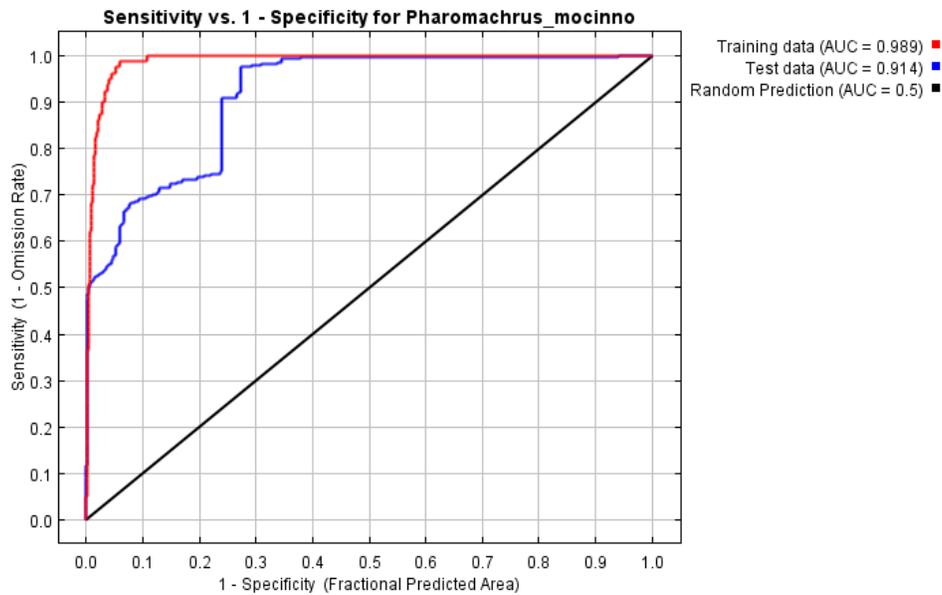


Figura 50: gráfico AUC para el modelo de distribución para el quetzal usando el 50% de las ocurrencias y la capa de rango de influencia

Como se puede observar en el gráfico anterior el modelo obtuvo un AUC de 0.989 para los datos de entrenamiento y un AUC de 0.914 para los datos de prueba.

Las capas más importantes para la construcción del modelo fueron las siguientes:

1. La capa de rango de influencia con un 76.7%.
2. La temperatura media del cuatrimestre más frío con un 6.2%.
3. Isotérmico con un 4.7%.

## 11. RESUMEN DE RESULTADOS Y CONCLUSIONES

En esta sección se muestra el resumen de los resultados obtenidos de los experimentos y algunas de las conclusiones obtenidas de ellos. En la tabla 2 se evidencia el resumen del primer conjunto de experimentos, en el cual se hace una comparación de los modelos generados al usar la capa de rango de influencia y los modelos sin esta capa; en las últimas dos columnas se muestra la diferencia entre el AUC del modelo sin la capa de rango de influencia contra el modelo que sí usa esta capa. Esta diferencia se muestra en porcentajes:

Tabla 2: resumen de resultados para el primer conjunto de experimentos

Especie	Especie relacionada	AUC en datos de entrenamiento	AUC en datos de prueba	capa más importante para el modelo	Diferencia en entrenamiento	Diferencia en pruebas
<b><i>Ara ambiguus</i> (lapa verde)</b>	No aplica	0.966	0.929	Precipitación anual	No aplica	No aplica
<b><i>Ara ambiguus</i> (lapa verde)</b>	<i>Dipteryx panamensis</i> (almendro amarillo como nido)	0.969	0.930	Precipitación anual	0.3%	0.1%
<b><i>Ara ambiguus</i> (lapa verde)</b>	<i>Dipteryx panamensis</i> (almendro amarillo como alimentación)	0.969	0.935	Precipitación anual	0.3%	0.5%
<b><i>Procnias tricarunculatus</i> (pájaro campana)</b>	No aplica	0.961	0.887	Rango anual de temperaturas	No aplica	No aplica
<b><i>Procnias tricarunculatus</i> (pájaro campana)</b>	Especies de la familia <i>Lauraceae</i>	0.969	0.926	Rango de influencia	0.8%	3.9%
<b><i>Pharomachrus mocinno</i> (quetzal)</b>	No aplica	0.966	0.957	Temperatura media del cuatrimestre más frío	No aplica	No aplica
<b><i>Pharomachrus mocinno</i> (quetzal)</b>	Especies de la familia <i>Lauraceae</i>	0.973	0.965	Rango de influencia	0.7%	0.8%

Como se observa en la tabla 2, en todos los experimentos se dio una mejoría tanto en el AUC para los datos de entrenamiento como en el AUC en los datos de pruebas. Además, cabe destacar que tanto para

el pájaro campana como para el quetzal al generar el modelo la capa de rango de influencia es la que más valor le agrega; esto da a entender que en el modelo generado por *Maxent* la influencia que genera la especie relacionada es realmente importante para determinar la distribución geográfica de la especie en estudio. Otra variable a destacar es que las mejoras son más significativas en los datos de pruebas; esto puede deberse a que *Maxent* al entrenarse con esos datos tiene un mayor porcentaje de acierto, por lo que los AUC tienden a ser más altos.

En el caso de la lapa verde y sus relaciones con los almendros amarillos, durante una charla con el doctor Olivier Chassot se discutió que en realidad existía un sobreabastecimiento de almendros. Por lo tanto, se pensaba que lo que más afectaba la distribución de la lapa verde eran los almendros con características específicas para poder tener nidos en ellos; debido a esto, se esperaba que los resultados de los AUC de los experimentos con nidos fueran mejores a los AUC de los experimentos con alimentación. Sin embargo, sorprende que fuera al contrario. Esto puede deberse a que existen muy pocas ocurrencias de almendros para realizar el estudio, y para el caso de los experimentos de nidos se tienen aún menos datos de ocurrencias, ya que son un subconjunto de los datos utilizados en los experimentos de alimentación. En el caso de los experimentos de alimentación, como se mencionó en la sección 10.1.2, se incluyeron los datos suministrados por el CCT más los obtenidos en GBif; en el caso de los experimentos de nidos solo se utilizaron los datos del CCT.

En la tabla 3 se muestra un resumen de los resultados del conjunto de experimentos dos, donde se intenta saber si la capacidad de extrapolación del modelo mejora al utilizar la capa de rango de influencia. En las últimas dos columnas se muestra la diferencia entre el AUC del modelo sin la capa de rango de influencia contra el modelo que sí usa esta capa, esta diferencia se muestra en porcentajes:

Tabla 3: resumen de resultados para el segundo conjunto de experimentos

Especie	Especie relacionada	AUC en ratos de entrenamiento	AUC en datos de prueba	Capa más importante para el modelo	Diferencia en entrenamiento	Diferencia en pruebas
<b><i>Ara ambiguus</i> (lapa verde)</b>	No Aplica	0.952	0.914	Precipitación del cuatrimestre más húmedo	No aplica	No aplica
<b><i>Ara ambiguus</i> (lapa verde)</b>	<i>Dipteryx panamensis</i> (almendro amarillo como nido)	0.959	0.917	Rango de influencia	0.7%	0.3%
<b><i>Ara ambiguus</i> (lapa verde)</b>	<i>Dipteryx panamensis</i> (almendro amarillo como alimentación)	0.960	0.928	Rango de influencia	0.8%	1.1%
<b><i>Procnias tricarunculatus</i> (pájaro campana)</b>	No aplica	0.936	0.713	Rango medio de temperaturas	No aplica	No aplica
<b><i>Procnias</i></b>	Especies de la	0.977	0.955	Rango de	4.1%	24.2%

<b><i>tricarunculatus</i></b> (pájaro campana)	familia <i>Lauraceae</i>			influencia		
<b><i>Pharomachrus mocinno</i></b> (quetzal)	No aplica	0.986	0.860	Temperatura media del cuatrimestre más frío	No aplica	No aplica
<b><i>Pharomachrus mocinno</i></b> (quetzal)	Especies de la familia <i>Lauraceae</i>	0.989	0.914	Rango de influencia	0.3%	5.4%

Como en el experimento anterior para todos los casos se dio una mejora en los modelos que utilizaron la capa de rango de influencia, los AUC fueron mejores tanto para los datos de entrenamiento como para los datos de prueba. Además, cabe resaltar que en el caso de los experimentos de extrapolación en todos los casos donde se utilizó la capa de rango de influencia esta fue la que más aportó al modelo; lo cual quiere decir de nuevo que para *Maxent* la influencia de las relaciones interespecíficas es realmente importante para definir la distribución geográfica de las especies en estudio.

Cabe destacar también que con la capa de rango de influencia para algunos de los experimentos las mejoras son muy significativas. En el caso del pájaro campana se pasa de un AUC para datos de entrenamiento de 0.936 a 0.977 y de un AUC para datos de prueba de 0.713 a un 0.955, lo que corresponde a un 4.1% y un 24.2%. Otro experimento donde se dio una gran mejora es con el quetzal donde se pasó de un AUC para datos de prueba de un 0.860 a un 0.914, lo que corresponde a un 5%.

De acuerdo con los resultados obtenidos en los distintos experimentos se puede concluir que en efecto la inclusión de las relaciones interespecíficas ayuda a mejorar la capacidad del *Maxent* para predecir la distribución geográfica de una especie.

## 12. TRABAJO FUTURO

Con los experimentos planteados en esta investigación se logró concluir que la inclusión de las relaciones interespecíficas mejora la capacidad del *Maxent* para predecir la distribución geográfica de una especie; sin embargo, aún se puede ahondar en muchas otras aristas de esta investigación. Por ejemplo, solo se tomaron en cuenta dos tipos de relaciones interespecíficas: una como alimentación y la otra como nido. No obstante, como se mencionó en la sección 5 de la presente investigación existen muchas otras relaciones que no se tomaron en cuenta por razones de tiempo.

Además, por la forma en que se desarrolló la metodología para la inclusión de las relaciones interespecíficas es posible incluir esta misma capa de información en otras herramientas tales como *GARP*, con lo cual tal y como se hizo con *Maxent* se podrían hacer experimentos para saber si la inclusión de esta capa mejora la capacidad de esa otra herramienta para predecir la distribución geográfica de la especie.

Otro punto en el cual se pueden desarrollar trabajos posteriores es averiguar cómo afecta el parámetro de rango de influencia en el modelo. La mejora que se da en la capacidad de predicción de la distribución de una especie depende mucho de ese parámetro; durante el desarrollo de esta investigación se utilizaron varios valores para los cuales se daban mejores y peores resultados, por esto sería de gran utilidad realizar otra investigación de cómo este parámetro afecta al modelo.

## 13. BIBLIOGRAFÍA

[Araújo, Guisan, 2009]

Araújo, M.; Guisan, A. (2006). *Five (or so) challenges for species distribution modelling*.

[Ávila, Hernández, Velarde, 1996]

Ávila, M.; Hernández, V.; Velarde, E. (1996). The Diet of Resplendent Quetzal (*Pharomachrus mocinno mocinno*: Trogonidae) in a Mexican Cloud Forest.

[Bolstad, 2012]

Bolstad, P. (2012). *GIS Fundamentals*, 4th Edition.

[Bond *et al*, 2011]

Bond, N; Thomson, J; Reich, P; Stein, J. (2011). Using species distribution models to infer potential climate change-induced range shifts of freshwater fish in south-eastern Australia.

[Elith, Graham, 2009]

Elith, J.; Graham C. (2009). Do they? Why do they differ? -- on finding reasons for differing performances of species distribution models.

[Foote, Lynch, 1995]

Foote, K; Lynch M. (1995). *Geographic Information Systems as an Integrating Technology: Context, Concepts, and Definition*. Disponible online en la dirección <http://www.colorado.edu/geography/gcraft/notes/intro/intro.html>

[Garcia, 2011]

Garcia, J. (2011). *Las interacciones biológicas*. Disponible online en la dirección <http://ocw.unican.es/ciencias-sociales-y-juridicas/biogeografia/materiales/tema-2/2.2.1-las-interacciones-biologicas>.

[GBIF, 2015]

What is GBIF? Página web institucional de GBIF. Disponible online en la dirección <http://www.gbif.org/>

[Jiménes *et al*, 2011]

Jiménez, Q.; Rojas, F; Rodríguez, L; Rojas, V. (2011). *Árboles maderables de Costa Rica, ecología y silvicultura*.

[Luis, 1999a]

Elizondo, L. (1999). *Pharomachrus Mocinno*. Disponible online en la dirección <http://atta2.inbio.ac.cr/neoportal-web/species/Pharomachrus%20mocinno>

[Luis, 1999b]

Elizondo, L. (1999). Procnias Tricarunculatus. Disponible online en la dirección <http://atta2.inbio.ac.cr/neoportal-web/species/Procnias%20tricarunculatus>

[Mateo, Felicísimo, Muñoz, 2011]

Mateo, R.; Felicísimo, A.; Muñoz, J. (2011). *Modelos de distribución de especies: Una revisión sintética*.

[Monge, 2011]

Monge, G. (2011) Ara Anbiguus. Disponible online en la dirección <http://atta2.inbio.ac.cr/neoportal-web/species/Ara%20ambiguus>

[Owens, Bentley, Townsend, 2012]

Owens, H; Bentley, A; Townsend, A. (2012). Predicting suitable environments and potential occurrences for coelacanths (*Latimeria* spp.).

[Parmenter, 2007]

Parmenter, B. (2007). GIS Data Structures How do we represent the world in a GIS database?

[Peterson, Papes, Eaton, 2007]

Peterson, T.; Papes, M.; Eaton, M. (2007). *Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent*.

[Peterson, 2006]

Peterson, T. (2006). *Uses and requirements of ecological niche models and related distributional models*.

[Peterson, Papes, Soberón, 2008]

Peterson, T.; Papes, M.; Soberón, J. (2008) Rethinking receiver operating characteristic analysis applications in ecological niche modeling.

[Phillips, Anderson, Schapire, 2006]

Phillips, S.; Anderson, R; Schapire, R; (2006). *Maximum entropy modeling of species geographic distributions*.

[Watson, Dallwithz, 1992]

Watson, L.; Dallwitz, M. (1992). The families of flowering plants: descriptions, illustrations, identification, and information retrieval.

## 14. GLOSARIO

- **Especie invasora:** especie que no es nativa de una zona en particular.
- **Maxent:** *Maxent* es un método de aprendizaje de computadora de uso general, utilizado para hacer predicciones o inferencias a partir de datos incompletos [Phillips, Anderson, Schapire, 2006].
- **Modelo de distribución de una especie:** los modelos de distribución de especies son representaciones cartográficas de la idoneidad de un espacio para la presencia de una especie en función de las variables empleadas para generar dicha representación [Mateo, Felicísimo, Muñoz, 2011].
- **Modelo:** un modelo es una representación parcial de la realidad [Mateo, Felicísimo, Muñoz, 2011].
- **Ocurrencia:** punto en el espacio georeferenciados donde se ha visto a una especie en particular.
- **Relación interespecífica:** relaciones entre especímenes de diferentes especies.
- **Sistema de información geográfica:** un sistema de información geográfica es una herramienta de análisis en la cual se utilizan datos georeferenciados [Bolstad, 2012].

## 15. APÉNDICE

### 15.1 NIDOS DE LAPA VERDE

A continuación se muestran los nidos que el doctor Olivier Chassot del Centro Científico Tropical (CCT) suministró y fueron utilizados en esta investigación:

NIDOS LAPA VERDE (1994-2013)											
ID	LOCATION	OWNER	LAT GRA D	LAT MI N	LON G GRA D	LON G MIN	SPECIES	PLA TE	DBH	HABIT AT	HEIG HT
N1	Cureña (Yucatan)	Carlos Araya	10	38. 95	84	8.65	<i>Vochysia ferruginea</i>		100. 00	4	
N1 0	Boca Tapada	Luz Mirio Alvarez (hueco bajo)	10	38. 26	84	14.5 5	<i>D. panamensis</i>	976	97.0 0	4	10.00
N1 1	Boca Tapada	Luz Mirio Alvarez (por la casa)	10	38. 68	84	14.7 0	<i>D. panamensis</i>	719	150. 00	4	
N1 2	Boca Tapada	Luz Mirio Alvarez (potrero de adentro)	10				<i>D. panamensis</i>	986	90.0 0	4	
N1 3	Cureñita	Alex Alfaro B	10	42. 82	84	5.66	<i>D. panamensis</i>	740	109. 00	4	19.00
N1 4	Camino a Ochoa		10	46. 17	84	8.95	<i>D. panamensis</i>				
N1 5	Camino a Ochoa	Hnos. Miranda					<i>D. panamensis</i>				
N1 6	Camino a Ochoa	Hnos. Miranda					<i>D. panamensis</i>				
N1 7	Camino a la Unión del Toro	Marcos Vargas	10	38. 23	84	3.15	<i>D. panamensis</i>	716	91.0 0	6	
N1 8	Casa de Lata	Andrés Pérez	10	43. 21	84	9.56	<i>D. panamensis</i>	714	104. 00	5	
N1 9	Casa de Lata	Alvaro	10	43. 46	84	10.2 2	<i>D. panamensis</i>	694	145. 00	4	30.30
N2	Alemán						<i>D. panamensis</i>				

							<i>is</i>				
<b>N20</b>	Castelmare	Orlando Barrantes	10	37.37	84	17.60	<i>D. panamensis</i>	105.00		4	
<b>N21</b>	Castelmare	Orlando Barrantes (adentro)	10				<i>D. panamensis</i>				
<b>N22</b>	Caño Colpachi	Pánfilo Sequeira	10	39.74	84	1.89	<i>D. panamensis</i>	523 R	145.00	2	18.00
<b>N23</b>	Camino a Ochoa		10	45.61	84	9.25	<i>Vatairea lundellii</i>				
<b>N24</b>	Colpachi	Jaime Viquez	10	42.85	83	59.20	<i>D. panamensis</i>	524 R	166.00	4	22.90
<b>N25</b>	Const. Belén	Alvaro	10	38.96	84	10.05	<i>D. panamensis</i>	44		4	
<b>N26</b>	Cruce Palo Seco		10	45.11	84	9.26	<i>D. panamensis</i>				
<b>N27</b>	Cureña	Alex Alfaro A	10	42.74	84	5.94	<i>D. panamensis</i>	6	140.00	4	25.00
<b>N28</b>	Cureñita	Churillas	10	42.07	84	6.52	<i>Vochysia ferruginea</i>	539 R	114.00	4	18.00
<b>N29</b>	Cureñita	Gerardo Miranda					<i>D. panamensis</i>			4	
<b>N3</b>	Boca Sn. Carlos	Ricardo Rodriguez	10	47.02	84	10.61	<i>D. panamensis</i>			9	
<b>N30</b>	Cureñita	Gerardo Miranda (Curena)	10	43.73	84	4.53	<i>D. panamensis</i>	544 R	125.00	4	17.00
<b>N31</b>	Cureñita/Sn. Juan	Octavino Morera	10	45.08	84	4.60	<i>D. panamensis</i>	987	95.00	4	15.00
<b>N33</b>	La Marina	Enrique Rojas	10	40.13	84	3.98	<i>D. panamensis</i>	6	111.00	5	22.00
<b>N34</b>	La Unión	Adán Tenorio					<i>D. panamensis</i>				
<b>N35</b>	Las Marias	Pto. Viejo					<i>D. panamensis</i>				
<b>N36</b>	Los Almendros	Israel Vargas					<i>D. panamensis</i>				
<b>N3</b>	Los Angeles	ABAS 1	10	41.	83	58.8	<i>D.</i>	546	152.	4	33.00

<b>7</b>				59		8	<i>panamensis</i>	R	00		
<b>N38</b>	Los Angeles	ABAS 2	10	41.62	83	58.82	<i>D. panamensis</i>	545	120.00	4	23.00
<b>N39</b>	Los Angeles	ABAS 3					<i>D. panamensis</i>				
<b>N4</b>	Boca Tapada	Ecaforestal	10	39.04	84	14.23	<i>D. panamensis</i>	284	145.00	5	
<b>N40</b>	Los Angeles	William Leon					<i>D. panamensis</i>				
<b>N41</b>	Los Angeles	William Mendoza					<i>D. panamensis</i>				
<b>N43</b>	Ojoche	Manuel Alvarado	10	35.80	84	14.52	<i>D. panamensis</i>	713	159.00	4	15.00
<b>N44</b>	Pangola	Banco Nac. (Piñera)	10				<i>D. panamensis</i>				
<b>N45</b>	Pangola	Hnos. Murillo A	10	34.67	84	4.92	<i>D. panamensis</i>			4	21.00
<b>N46</b>	Pangola	Hnos. Murillo B	10	34.08	84	5.51	<i>D. panamensis</i>	69	150.00	7	33.00
<b>N47</b>	Patastillo	Claudio Arce Rojas (Cullo)	10	42.17	84	16.25	<i>D. panamensis</i>	507	115.00	5	35.00
<b>N48</b>	Pejibayera	Didre	10	40.99	84	9.53	<i>D. panamensis</i>	522	148.00	8	35.00
<b>N49</b>	Piedra Alegre	Marcos Rojas	10	34.47	84	18.06	<i>D. panamensis</i>	712	114.00	8	
<b>N5</b>	Boca Tapada	Eduardo Artavia	10	40.53	84	12.19	<i>D. panamensis</i>			4	
<b>N50</b>	Q. Ojoche	Marvin Quesada	10	37.51	84	16.02	<i>D. panamensis</i>	707	140.00	4	15.00
<b>N51</b>	Q. Ojoche	Marvin Quesada					<i>D. panamensis</i>	300	110.00	2	
<b>N52</b>	Rancho Alegre	Pata de Sapo					<i>D. panamensis</i>				
<b>N5</b>	San Marcos	Jose Luis	10	37.	84	22.4	<i>D.</i>	983	161.	8	14.00

<b>3</b>		Rodriguez		00		2	<i>panamensis</i>		00		
<b>N54</b>	Santa Delia	Vicente Miranda	10	34.53	84	2.16	<i>D. panamensis</i>		92.00	4	
<b>N55</b>	Santa Delia	Carlos Paniagua	10	34.94	84	2.84	<i>Enterolobium schomburgkii</i>	537 R	108.00	4	
<b>N56</b>	Santa Delia	COOCIQUE A/santander	10	34.19	84	2.24	<i>D. panamensis</i>	538	112.00	1	
<b>N57</b>	Tambor						<i>D. panamensis</i>			4	
<b>N58</b>	Boca Tapada	Tres Amigos	10	38.92	84	15.30	<i>D. panamensis</i>	993	115.00	7	30.00
<b>N59</b>	Yucatan	Juan Durán I	10	39.02	84	8.60	<i>D. panamensis</i>	968	126.00	4,8	18.96
<b>N6</b>	Boca Tapada	El Chino (WorldCo)	10	41.25	84	15.34	<i>D. panamensis</i>	898	153.00	5	
<b>N60</b>	Yucatan	Juan Durán II	10	39.54	84	8.94	<i>D. panamensis</i>	115	113.00	4	25.00
<b>N62</b>	Boca Río Cuarto	Efrain Avila (Info. Pepin)	10				<i>D. panamensis</i>				
<b>N63</b>	Pataste	Cullo	10	41.52	84	16.55	<i>D. panamensis</i>				
<b>N64</b>	Cureña	Alex Alfaro C	10	42.76	84	5.65	<i>D. panamensis</i>	86	94.00	2	15.00
<b>N65</b>	Santa Delia	COOCIQUE B	10	34.48	84	2.03	<i>D. panamensis</i>		68.00	4	
<b>N66</b>	Boca Tapada	Luz Mirio Alvarez (potrero de adentro)	10	38.66	84	14.71	<i>D. panamensis</i>	708	173.00	4	
<b>N67</b>	Des. Río Toro	Eudelia Alfaro	10	38.73	84	1.19	<i>D. panamensis</i>				
<b>N68</b>	Camino al Mirador	Juan 'Pelado' (Finca Francés)	10	42.07	84	3.79	<i>D. panamensis</i>	897	118.00	7	10.00

<b>N69</b>	Tamborcito	Walter Gonzalez Olivarez	10	43.03	84	3.23	<i>Vochysia ferruginea</i>	540 R	110.00	5	20.00
<b>N7</b>	Boca Tapada	Joho A	10	39.95	84	13.01	<i>D. panamensis</i>	113	128.00	5	
<b>N70</b>	Tamborcito	Alberto Quiroz	10	43.99	84	2.94	<i>D. panamensis</i>	541 R	150.00	4	25.00
<b>N71</b>	Tamborcito	Alberto Quiroz	10	43.97	84	2.98	<i>D. panamensis</i>	542 R	115.00	4	25.00
<b>N72</b>	Tamborcito	Alberto Quiroz	10	43.99	84	3.05	<i>D. panamensis</i>	543 R	140.00	4	
<b>N73</b>	El Desnudo	Carlos Araya	10	39.03	84	8.64	<i>D. panamensis</i>	525	126.00	4,8	16.00
<b>N74</b>	Boca Tapada	Luz Mirio (abejas)	10	38.43	84	14.74	<i>D. panamensis</i>	576	190.00	4	33.00
<b>N75</b>	La Mona	Alfredo Peralta	10	43.02	84	12.74	<i>Vochysia ferruginea</i>	606	105.00	2,8	
<b>N76</b>	Boca Tapada	Antonio Vargas	10	36.68	84	16.71	<i>D. panamensis</i>	577	161.00	4	18.00
<b>N77</b>	Finca Churrias 2	Rafael	10	42.39	84	6.66	<i>Carapa nicaraguensis</i>	609	78.00	4,9	16.00
<b>N78</b>	El Flaco	Cuidador: Pollo finca de Jose Luis Ferrer	10	43.13	84	5.44	<i>Enterolobium schomburgkii</i>	655	65.00	4,9	11.00
<b>N79</b>	Caño Boca Tapada (Pozo Azul)	Eduardo Brenes: cuida Don Cullo	10	42.23	84	14.69	<i>Vochysia ferruginea</i>	515	118.30	6,8	13.00
<b>N8</b>	Boca Tapada	Joho B	10	39.41	84	12.88	<i>D. panamensis</i>	71	300.00	6	
<b>N80</b>	Yucatan (Cangrejera)	Alvaro Vargas: cuida Eliecer Jimenez	10	39.06	84	10.60	<i>D. panamensis</i>	373	159.40	4,2	17.00
<b>N81</b>	Boca Tapada (Ulises)	Ecaforestal (Ulises)	10	38.91	84	14.08	<i>D. panamensis</i>	389	178.30	5	35.00
<b>N82</b>	La Legua	David Webb	10	31.64	84	17.89	<i>D. panamensis</i>	213		4	

<b>N8 3</b>	Pangola	El Retorno	10	34. 17	84	4.45	<i>D. panamens is</i>			1
<b>N8 4</b>	Pangola	Jose Antonio	10	34. 90	84	4.85	<i>D. panamens is</i>			5
<b>N8 5</b>	Yucatan	Alvaro Vargas: cuida Eliecer Jimenez	10	38. 90	84	10.5 9	<i>D. panamens is</i>			4
<b>N8 6</b>	Boca Tamborcito	Miguel Trejos	10	44. 00	83	58.7 0	<i>Prioria Copaifera</i>	220	186. 00	2,4
<b>N8 7</b>	La Marina	Amadeo	10	38. 31	84	4.67	<i>D. panamens is</i>	226		4
<b>N8 8</b>	San Marcos	Alexis Cervantes	10	37. 00	84	23.0 8	<i>D. panamens is</i>			4
<b>N8 9</b>	Camino a C. Lata	Monterrey					<i>D. panamens is</i>			
<b>N9</b>	Boca Tapada	Luz Mirio Alvarez	10	38. 55	84	14.4 9	<i>D. panamens is</i>	999	135. 00	4
<b>N9 0</b>	Cureña (Yuc. III)	Juan Durán					<i>D. panamens is</i>			4
<b>N9 1</b>	Yuc.(Cangrej era)	Álvaro Vargas	10	39. 49	84	10.3 5	<i>D. panamens is</i>	313		4
<b>N9 2</b>	La Unión	Jaime Viquez	10	37. 58	84	3.13	<i>D. panamens is</i>	319	300. 00	4
<b>N9 3</b>	Los Angeles	Fabio Vargas					<i>D. panamens is</i>			4
<b>N9 4</b>	Boca Cureña	Carlos Espinoza					<i>D. panamens is</i>			4
<b>N9 5</b>	Caño San Antonio	Jorge Morales					<i>D. panamens is</i>			
<b>N9 8</b>	Cureñita	Octavino Morera	10	44. 43	84	4.58	<i>Vochysia ferruginea</i>			
<b>N9 9</b>	Santa Delia	Desconoci do	10	35. 25	84	3.43	<i>D. panamens is</i>			
<b>N1 00</b>	Santa Delia	Vicente Miranda	10	34. 33	84	1.55	<i>D. panamens is</i>			

<b>N1 01</b>	Santa Delia	Vicente Miranda	10	34. 34	84	10.3 6	<i>D. panamensis</i>
<b>N1 02</b>	Camino a BT	Javier Rojas	10	37. 35	84	16.0 1	<i>D. panamensis</i>
<b>N1 03</b>	Constructor a Belén	Álvaro Vargas	10	38. 46	84	10.1 2	<i>D. panamensis</i>
<b>N1 04</b>	Boca del Toro Unión	Alonso Funes	10	38. 27	84	1.28	<i>Vochysia ferruginea</i>
<b>N1 05</b>	La Unión	Adán Tenorio	10	37. 53	84	2.59	<i>D. panamensis</i>
<b>N1 06</b>	Colpachí- Tambor	Víctor Gómez/Mi guel Trejos	10	43. 50	84	58.1 9	<i>D. panamensis</i>
<b>N1 07</b>	Tambor	Víctor Gómez/Mi guel Trejos	10	44. 12	84	58.4 1	<i>Cativo</i>
<b>N1 08</b>	Rancho Alegre	Jorge Mora	10	39. 23	84	2.33	<i>Desconoci do</i>
<b>N1 09</b>	Cureña	Rafael Urias	10	42. 30	84	6.28	<i>Vochysia ferruginea</i>
<b>N1 12</b>	Boca Tapada	Sama	10	40. 29	84	12.1 4	<i>D. panamensis</i>
<b>N1 17</b>	Boca Tapada	Lagarto Lodge/La Belén	10	40. 53	84	10.8 0	<i>D. panamensis</i>
<b>N1 23</b>	La Unión	Martin Corrales	10	38. 03	84	1.49	<i>D. panamensis</i>
<b>N1 24</b>	Boca Río Toro	Carlos Chaves	10	40. 22	84	0.06	<i>D. panamensis</i>
<b>N1 25</b>	La Trinidad	Eulalio Sosa	10	43. 17	83	56.4 7	<i>D. panamensis</i>
<b>N1 26</b>	Copalchi	Borola	10	43. 28	84	58.1 9	<i>Cativo</i>
<b>N1 27</b>	Boca Tambor	Miguel Trejos	10	43. 28	84	58.1 5	<i>D. panamensis</i>
<b>N1 28</b>	Casa Lata	Andrés López	10	43. 21	84	9.30	<i>D. panamensis</i>
<b>N1 29</b>	Boca Tapada	Luz Mirio	10	38. 11	84	14.3 9	<i>D. panamensis</i>
<b>N1</b>	Boca	Lagarto	10	41.	84	10.3	<i>D.</i>

<b>30</b>	Tapada	Lodge		29		8	<i>panamensis</i>
<b>N1</b>	Boca	Marcelo	10	42.	84	10.0	<i>D.</i>
<b>31</b>	Tapada	Villalobos		13		4	<i>panamensis</i>
<b>N1</b>	Santa Delia	Cascales	10	35.	84	3.42	<i>D.</i>
<b>32</b>				24			<i>panamensis</i>

Para esta investigación solo se utilizaron los nidos en almendros amarillos y además se cambió el formato de georeferenciación de grados, minutos y segundos a un formato digital.

## 15.2 PROGRAMA HECHO EN JAVA PARA CREAR LA CAPA DE RANGO DE INFLUENCIA

Para construir la capa de rango de influencia y los experimentos para la extrapolación se construyó un programa hecho en *Java*. El mismo puede leer ocurrencias en formato .csv descargados directamente de GBIF o en el formato que utiliza *Maxent*. La capa de rango de influencia usa como base una de las capas del *WorldClim* que se utilizaron durante los experimentos.

Para la realización de los experimentos de extrapolación existe un método que excluye las ocurrencias siguiendo los parámetros que se le indican, así se pueden ir diseñando los distintos experimentos.

A continuación se muestra el código fuente de dicho programa:

```
import java.io.FileWriter;
import java.io.IOException;
import java.nio.charset.Charset;
import java.nio.file.Files;
import java.nio.file.Paths;
import java.util.ArrayList;
import java.util.HashMap;
import java.util.Iterator;
import java.util.List;
import java.util.regex.Matcher;
import java.util.regex.Pattern;

public class BuildInfluenceRaster {
    private int _numberOfColumns = 0, _numberOfRows = 0, _influenceRange = 5;
    private double _xCorner = 0, _yCorner = 0, _cellSize = 0;
    private String _outputFileName = "base_influence_grid";
    private String _inputFileName = "ara_ambiguous_from_gbif";
    private String _workingFolder = "D:/TEC/Thesis/Species/";
    private static final int NO_DATA_VALUE = -9999;

    public static void main(String[] args) {
        BuildInfluenceRaster raster = new BuildInfluenceRaster();

        List<Ocurrence> occurrences = null;

        try {
            RasterCell[][] matrix = raster.buildRasterMatrixFromTemplate();

            occurrences = raster.readOccurrences(true, true);

            raster.updateMatrixUsingOccurrences(matrix, occurrences);

            raster.writeAsciiGrid(matrix);

            raster.writeOccurrencesExcludingFiftyPercentOfOccurrences(occurrences,
true, false);

```

```

    } catch (IOException e) {
        // TODO Auto-generated catch block
        e.printStackTrace();
    }

    System.out.println("The asc grid has been generated.");
}

public void showLimits(RasterCell[][] matrix){
    for(int x = 0; x < _numberOfColumns; x++){
        matrix[( _numberOfRows - 140) / 2][x].setValue(1000);
    }

    for(int y = 0; y < _numberOfRows; y++){
        matrix[y][(( _numberOfColumns + 0) / 2)].setValue(1000);
    }
}

public RasterCell[][] buildRasterMatrixFromTemplate() throws IOException{
    Charset charset = Charset.forName("ISO-8859-1");
    List<String> lines =
Files.readAllLines(Paths.get("D:/TEC/Thesis/AsciiLayers", "annualmeantemperature.asc"),
charset);

    _numberOfColumns = Integer.parseInt(getNumberFromLine(lines.get(0)));
    _numberOfRows = Integer.parseInt(getNumberFromLine(lines.get(1)));
    _xCorner = Double.parseDouble(getNumberFromLine(lines.get(2)));
    _yCorner = Double.parseDouble(getNumberFromLine(lines.get(3)));
    _cellSize = Double.parseDouble(getNumberFromLine(lines.get(4)));

    RasterCell[][] matrix = new RasterCell[_numberOfRows][_numberOfColumns];
    double yLevelValue = _yCorner;
    double xLevelValue = _xCorner;

    int lineIndex = lines.size() - 1;
    int matrixIndex = _numberOfRows - 1;

    while(!lines.get(lineIndex).contains("NODATA_value")){
        String line = lines.get(lineIndex);
        String[] cells = line.split(" ");

        for(int c = 0; c < _numberOfColumns; c++){
            int value = Integer.parseInt(cells[c]);

            if(value != NO_DATA_VALUE){
                value = 0;
            }
            RasterCell cell = new RasterCell(xLevelValue, xLevelValue +
_cellSize, yLevelValue, yLevelValue + _cellSize, value);
            matrix[matrixIndex][c] = cell;
            xLevelValue += _cellSize;
        }

        xLevelValue = _xCorner;
        yLevelValue += _cellSize;
        lineIndex--;
        matrixIndex--;
    }

    return matrix;
}

public String getNumberFromLine(String line){
    Pattern p = Pattern.compile("-?\\d+\\.?\\d*");
    Matcher m = p.matcher(line);

    if(m.find()){
        return m.group();
    }else{
        return null;
    }
}
}

```

```

public RasterCell[][] buildRasterMatrix(){
    RasterCell[][] matrix = new RasterCell[_numberOfRows][_numberOfColumns];
    double yLevelValue = _yCorner;
    double xLevelValue = _xCorner;

    // Fill the matrix with the RasterCell including coordinates
    // I start by the left inferior corner to fill out the matrix
    for(int y = _numberOfRows - 1; y >= 0; y--){
        for(int x = 0; x < _numberOfColumns; x++){
            RasterCell cell = new RasterCell(xLevelValue, xLevelValue +
            _cellSize, yLevelValue, yLevelValue + _cellSize, NO_DATA_VALUE);
            matrix[y][x] = cell;
            xLevelValue += _cellSize;
        }

        xLevelValue = _xCorner;
        yLevelValue += _cellSize;
    }

    return matrix;
}

/**
 * This method reads the occurrences of an species, the valid formats for this
 * methods are the one used by Maxent (Species, Longitude, Latitude).
 * Or the csv that uses GBif for share occurrences. The format used to read the file
 * depends on the parameter isFromGBif passed to the method.
 * In the case that the input file is from GBif, then this method writes a version
 * usable in Maxent and if the parameter after2000 is set to true, then
 * the occurrences previous to 2000 are excluded.
 * @param isFromGBif
 * @param after2000
 * @return
 * @throws IOException
 */
public List<Occurrence> readOccurrences(boolean isFromGBif, boolean after2000) throws
IOException{
    Charset charset = Charset.forName("ISO-8859-1");
    List<String> lines = Files.readAllLines(Paths.get(_workingFolder,
    _inputFileName + ".csv"), charset);
    List<Occurrence> occurrences = new ArrayList<Occurrence>();
    boolean firstLine = true;

    for(String line : lines){

        // Skip the header
        if(firstLine) {
            firstLine = false;
            continue;
        }
        if(!isFromGBif){
            String[] components = line.split(",");
            occurrences.add(new
            Occurrence(Double.parseDouble(components[2]), Double.parseDouble(components[1]),
            components[0]));
        }else{
            String[] components = line.split(" ");
            String specie = components[9];
            String latitudeStr = components[16];
            String longitudeStr = components[17];
            String yearStr = components[25];

            // Check that the latitude, longitude and species fields are
            not in blank
            if(specie != null && !specie.isEmpty() && latitudeStr != null
            && !latitudeStr.isEmpty() && longitudeStr != null && !longitudeStr.isEmpty() && yearStr !=
            null && !yearStr.isEmpty()){
                Double latitude = Double.parseDouble(latitudeStr);
                Double longitude = Double.parseDouble(longitudeStr);
                int year = Integer.parseInt(yearStr);

```

```

        // Check that they are not 0
        if(latitude != 0 && longitude != 0){
            boolean addOccurrence = true;
            if(after2000){
                if(year < 2000){
                    addOccurrence = false;
                }
            }

            if(addOccurrence){
                occurrences.add(new Occurrence(latitude,
longitude, specie));
            }
        }
    }

    // Write the occurrences in MaxEnt format
    if(isFromGBif){
        FileWriter fw = new FileWriter(_workingFolder + _inputFileName +
"_MaxentFormat.csv");
        fw.write("species,dd long,dd lat\n");

        for(Occurrence ocu: occurrences){
            fw.write(ocu.getSpecie() + "," + ocu.getLongitude() + "," +
ocu.getLatitude() + "\n");
        }

        fw.close();
    }

    return occurrences;
}

/**
 * This method exclude the occurrences that do not satisfy the criteria, if the
parameter useDiagonal is true,
 * then the study area is split in 4 parts, and just the superior left corner and
the inferior right corner.
 * If the parameter useSuperiorArea is set to true, then the study area is split in
two parts, superior and inferior, and just the upper part of the area is taken into account.
 * If the parameter useDiagonal and useSuperiorArea are set to false, then the study
area is split in two parts, and just the occurrences
 * in the right part are included in the final occurrences lists
 * @param occurrences
 * @param useDiagonal
 * @param useSuperiorArea
 * @throws IOException
 */
public void writeOccurrencesExcludingFiftyPercentOfOccurrences(List<Occurrence>
occurrences, boolean useDiagonal, boolean useSuperiorArea) throws IOException{
    double middleX = _xCorner + (_cellSize * ((_numberOfColumns + 0) / 2));
    double middleY = _yCorner + (_cellSize * ((_numberOfRows + 140) / 2));
    List<Occurrence> processedOccurrences = new ArrayList<Occurrence>();

    // Add just the occurrences that are right superior corner and left inferior
corner
    for(Occurrence ocu : occurrences){
        if(useDiagonal){
            if((ocu.getLongitude() < middleX && ocu.getLatitude() >
middleY) ||
            (ocu.getLongitude() > middleX &&
ocu.getLatitude() < middleY)){
                processedOccurrences.add(ocu);
            }
        } else if (useSuperiorArea){
            if(ocu.getLatitude() > middleY){
                processedOccurrences.add(ocu);
            }
        }
    }
}

```

```

        } else{
            if(ocu.getLongitude() < middleX){
                processedOcurrences.add(ocu);
            }
        }
    }

    // Write the new file
    FileWriter fw = new FileWriter(_workingFolder + _inputFileName +
    "_MaxentFormat_fiftyPercent.csv");
    fw.write("species,dd long,dd lat");

    for(Ocurrence ocu : processedOcurrences){
        fw.append("\n");
        fw.append(ocu.getSpecie() + "," + ocu.getLongitude() + "," +
    ocu.getLatitude());
    }

    fw.close();
}

public void updateMatrixUsingOcurrences(RasterCell[][] matrix, List<Ocurrence>
ocurrences){

    // I traverse the matrix and ask for each occurrence in order to see if the
    occurrence is inside the particular cell
    for(int y=0; y < _numberOfRows; y++){
        for(int x=0; x < _numberOfColumns; x++){
            // Go throw the list of occurrences in order to see if I found
            the cell where it is located at
            for(Ocurrence ocu : ocurrences){
                if(ocu.getLongitude() >= matrix[y][x].getXInitial() &&
    ocu.getLongitude() <= matrix[y][x].getXFinal() &&
    ocu.getLatitude() >=
    matrix[y][x].getYInitial() && ocu.getLatitude() <= matrix[y][x].getYFinal()){
                    HashMap<String, RasterCell> cells = new
    HashMap<>();

                    getCellOfExpandedSelection(cells,
    _influenceRange, y, x);

                    // Iterate the hashMap that contains the cells
                    Iterator<String> keySetIterator =
    cells.keySet().iterator();

                    while(keySetIterator.hasNext()){
                        String key = keySetIterator.next();
                        RasterCell value = cells.get(key);
                        updateCell(matrix,
    value.getMatrixRowPosition(), value.getMatrixColumnPosition());
                    }
                }
            }
        }
    }

}

/**
 * The occurrence will be in a certain Cell, but depending on the influence range
the selection has to be expanded
 * to other cells, so this method do that, from the start point that passed by
parameter it expand the selection in
 * a ring recursively and returns the pairs of coordinates of the matrix that has to
be updated
 * @param cells
 * @param depth
 * @param row
 * @param column
 * @return
 */

```

```

    public HashMap<String, RasterCell> getCellOfExpandedSelection(HashMap<String,
RasterCell> cells, int depth, int row, int column){
    // If cell out of the limits of the matrix then return or the depth of the
level already reach 0
    if(row < 0 || row >= _numberOfRows || column < 0 || column >=
_numberOfColumns || depth < 0){
        return cells;
    }

    RasterCell currentCell = new RasterCell(row, column);
    String currentKey = row + "," + column;

    // Check if the cell was already added
    if(!cells.containsKey(currentKey)){
        cells.put(currentKey, currentCell);
    }

    // Add the rest of cells that surround this one
    getCellOfExpandedSelection(cells, depth - 1, row - 1, column - 1); //Up
    getCellOfExpandedSelection(cells, depth - 1, row - 1, column);
//Up
    getCellOfExpandedSelection(cells, depth - 1, row - 1, column + 1); //Up
    getCellOfExpandedSelection(cells, depth - 1, row + 1, column - 1); //Down
    getCellOfExpandedSelection(cells, depth - 1, row + 1, column);
//Down
    getCellOfExpandedSelection(cells, depth - 1, row + 1, column + 1); //Down
    getCellOfExpandedSelection(cells, depth - 1, row, column - 1);
//Left
    getCellOfExpandedSelection(cells, depth - 1, row, column + 1);
//Right

    return cells;
}

public void updateCell(RasterCell[][] matrix, int row, int column){
    if(matrix[row][column].getValue() != NO_DATA_VALUE){
        matrix[row][column].setValue(matrix[row][column].getValue() + 1);
    }else{
        matrix[row][column].setValue(1);
    }
}

public void writeAsciiGrid(RasterCell[][] matrix) throws IOException{
    FileWriter fw = new FileWriter(_workingFolder + _outputFileName + ".asc");
    fw.write("ncols      " + _numberOfColumns + "\n");
    fw.write("nrows       " + _numberOfRows + "\n");
    fw.write("xllcorner    " + _xCorner + "\n");
    fw.write("yllcorner    " + _yCorner + "\n");
    fw.write("cellsize     " + _cellSize + "\n");
    fw.write("NODATA_value " + NO_DATA_VALUE + "\n");

    for(int y=0; y < _numberOfRows; y++){
        for(int x=0; x < _numberOfColumns; x++){
            fw.append(" " + matrix[y][x].getValue());

            // It's not the last
            if(x < _numberOfColumns - 1){
                fw.append(" ");
            }else{
                fw.append("\n"); // line break since is the last
one
            }
        }
    }

    fw.close();

    // Write the projection information
    FileWriter fwProjection = new FileWriter(_workingFolder + _outputFileName +
".prj");
    fwProjection.write("Projection    GEOGRAPHIC\n");

```

```

        fwProjection.write("Datum           WGS84\n");
        fwProjection.write("Spheroid      WGS84\n");
        fwProjection.write("Units       DD\n");
        fwProjection.write("Zunits      METERS\n");
        fwProjection.write("Parameters  \n");

        fwProjection.close();
    }
}

public class RasterCell {
    private double _xInitial;
    private double _xFinal;
    private double _yInitial;
    private double _yFinal;
    private int _matrixRowPosition;
    private int _matrixColumnPosition;
    private int _value;

    public RasterCell(int row, int column){
        setMatrixRowPosition(row);
        setMatrixColumnPosition(column);
    }

    public RasterCell(double xInitial, double xFinal, double yInitial, double yFinal,
int value){
        setXInitial(xInitial);
        setXFinal(xFinal);
        setYInitial(yInitial);
        setYFinal(yFinal);
        setValue(value);
    }

    public double getXInitial() {
        return _xInitial;
    }

    public void setXInitial(double _xInitial) {
        this._xInitial = _xInitial;
    }

    public double getXFinal() {
        return _xFinal;
    }

    public void setXFinal(double _xFinal) {
        this._xFinal = _xFinal;
    }

    public double getYInitial() {
        return _yInitial;
    }

    public void setYInitial(double _yInitial) {
        this._yInitial = _yInitial;
    }

    public double getYFinal() {
        return _yFinal;
    }

    public void setYFinal(double _yFinal) {
        this._yFinal = _yFinal;
    }

    public int getValue() {
        return _value;
    }

    public void setValue(int _value) {

```

```

        this._value = _value;
    }

    public int getMatrixRowPosition() {
        return _matrixRowPosition;
    }

    public void setMatrixRowPosition(int _matrixRowPosition) {
        this._matrixRowPosition = _matrixRowPosition;
    }

    public int getMatrixColumnPosition() {
        return _matrixColumnPosition;
    }

    public void setMatrixColumnPosition(int _matrixColumnPosition) {
        this._matrixColumnPosition = _matrixColumnPosition;
    }
}

public class Occurrence {
    private double _latitude;
    private double _longitude;
    private String _specie;

    public Occurrence(double latitude, double longitude, String specie){
        setLatitude(latitude);
        setLongitude(longitude);
        setSpecie(specie);
    }

    public double getLatitude() {
        return _latitude;
    }
    public void setLatitude(double _latitude) {
        this._latitude = _latitude;
    }

    public double getLongitude() {
        return _longitude;
    }

    public void setLongitude(double _longitude) {
        this._longitude = _longitude;
    }

    public String getSpecie() {
        return _specie;
    }

    public void setSpecie(String _specie) {
        this._specie = _specie;
    }
}

```