

**UNIVERSIDAD NACIONAL AGRARIA
LA MOLINA**

FACULTAD DE CIENCIAS FORESTALES



**"DISTRIBUCIÓN POTENCIAL DE ESPECIES
FORESTALES AMENAZADAS Y ENDÉMICAS
RARAS MEDIANTE MODELOS DE NICHOS
ECOLÓGICOS"**

Presentado por:

Corina Erika Navarrete Macedo

TESIS PARA OPTAR EL TÍTULO DE
INGENIERO FORESTAL

Lima - Perú
2019

ACTA DE SUSTENTACIÓN DE TESIS

Los Miembros del Jurado que suscriben, reunidos para calificar la sustentación del Trabajo de Tesis, presentado por la ex-alumna de la Facultad de Ciencias Forestales, Bach. **CORINA ERIKA NAVARRETE MACEDO**, intitulado “**"DISTRIBUCIÓN POTENCIAL DE ESPECIES FORESTALES AMENAZADAS Y ENDÉMICAS RARAS MEDIANTE MODELOS DE NICHO ECOLÓGICO"**”.

Oídas las respuestas a las observaciones formuladas, lo declaramos:

.....

con el calificativo de

En consecuencia queda en condición de ser considerada APTA y recibir el título de **INGENIERO FORESTAL**.

La Molina, 19 de julio de 2018

.....
Mg. Sc. **Victor Barrena Arroyo**
Presidente

.....
Ing. **Juan Carlos Ocaña Canales**
Miembro

.....
Mg. Sc. **Fernando Bernardo Regal**
Gastelumendi
Miembro

.....
Ing. **Carlos Vargas Salas**
Asesor

Dr. Jose Luis Marcelo Peña
Coasesor

DEDICATORIA

A Dios, por permitirme recuperar mi salud.

A mi Madre, por haberme sacado adelante en base a su incansable esfuerzo y amor. Trabajar día y noche por mí. Espero poder brindarte el amor y calidad de vida que mereces, este trabajo con mucho cariño va en tu honor mami. Tu eres la verdadera ingeniera.

A mis abuelitos, ese amor bonito que jamás olvidaré.

A mi Padre y hermanos, por su confianza y aliento.

A pitufina y pepita, por iluminar mis días con su ternura.

A él, que capturó mi corazón en la universidad y fue mi mejor amigo. Deseo que siempre seas feliz y mantengas tus ilusiones. No dudes de ti, eres brillante.

A mis amigas/os, su noble corazón me orienta hacía ustedes. Son un tesoro para mí.

AGRADECIMIENTOS

A mi alma mater la Universidad Nacional Agraria La Molina, especialmente a la Facultad de Ciencias Forestales por ser parte de esta investigación.

Un agradecimiento muy especial al Ing. Carlos Vargas Salas por ser mi asesor, por brindarme sus valiosos consejos, aportes y observaciones para el desarrollo de la tesis.

Un agradecimiento muy especial al profesor PhD. José Luis Marcelo-Peña por brindarme su confianza, apoyo, consejos y las facilidades para el desarrollo de esta investigación.

RESUMEN

En el presente estudio se determinó la distribución potencial de cinco especies forestales representantes de bosques andinos y bosques secos, mediante modelos correlativos con base ecológica. Asimismo, se evaluó las variables ambientales que aportaron mayor contribución en el modelado y los parámetros óptimos para obtener el mejor modelo posible. Los registros de ocurrencias de las especies fueron obtenidos de herbarios físicos y digitales, así como las variables ambientales se obtuvieron de bases digitales climáticas y topográficas. Para el modelamiento se utilizó el algoritmo *maxent* con una combinación de 35 configuraciones en parámetros y se evaluó mediante estadísticos de rendimiento y pruebas de significancia, de acuerdo al número de registros que ingresaron en el modelamiento. Según los resultados obtenidos, los mejores modelos para las especies fueron obtenidos con parámetros de configuración distintos a los que el algoritmo *maxent* otorga por defecto, así como las variables que tuvieron mayor aporte fueron la precipitación en el periodo más seco y la isothermalidad. Los modelos de *Buddleja incana* y *Polylepis racemosa* fueron considerados como de alta predictibilidad. Sin embargo, las distribuciones obtenidas para *Cedrela kuelapensis*, *Tecoma rosaefolia* y *Esenbeckia cornuta* deben considerarse como primeras aproximaciones geográficas a su distribución. Para optimizar el modelamiento de distribución de especies, se debe realizar futuros modelos basados en características ambientales y espectrales.

Palabras claves: Biogeografía, geomática, Árboles forestales, Métodos de optimización.

ÍNDICE GENERAL

	Página
I. Introducción	1
II. Revisión de Literatura	3
1. Bases conceptuales para la modelización del nicho ecológico	3
1.1. Definición de nicho ecológico	3
1.2. Diagrama BAM.....	4
1.3. Modelado de nicho ecológico (MNE) y Modelado de distribución de especies (MDEs)	5
1.4. Dualidad de Hutchinson.....	6
1.5. Aplicaciones del modelado de nicho ecológico	6
2. Modelado de nicho ecológico	8
2.1. Tipos de datos	10
2.1.1. datos de presencia.....	10
2.1.2. Ausencia estricta	13
2.1.3. Pseudoausencia.	13
2.1.4. "Fondo" o Background.	14
2.2. Variables ambientales	15
2.2.1. Selección de variables ambientales	16
2.2.2. Fuente de datos ambientales	16
2.3. Método de modelado	17
2.3.1. Selección del método adecuado.	18
2.4. Calibración y predicción del modelo.....	22
2.5. Evaluación del modelo	22
2.5.1. Evaluadores de rendimiento.....	23
2.6. Umbrales de corte	27
2.6.1. Determinación subjetiva de umbrales	28
2.6.2. Determinación de umbrales.....	29
3. Softwares libres	30
3.1. Rstudio y R.....	30
3.2. QGIS.....	30
4. Especies forestales de interés	31
4.1. Especies forestales amenazas.....	31
4.1.1. <i>Polylepis racemosa</i> Ruiz & Pav.	31
4.1.2. <i>Buddleja incana</i> Ruiz & Pav.	32
4.2. Especies endémicas.....	33
4.2.1. <i>Cedrela kuelapensis</i> T.D.Penn. & Daza	33
4.2.2. <i>Tecoma rosaefolia</i> kunth	33
4.2.3. <i>Esenbeckia cornuta</i> Engl.	34
III. Materiales y Métodos	35
1. Área de estudio	35
2. Equipos y materiales	35
3. Especies de interés	36
4. Colección de datos de ocurrencias y variables ambientales	36
5. Metodología	38
5.1. Modelado de especies con registros mayores a 20	38
5.1.1. Flujo de trabajo	38
5.1.2. Selección de variables ambientales	39
5.1.3. Preparación de datos de ocurrencia	39

5.1.4.	Calibración y predicción.....	40
5.1.5.	Evaluación de rendimiento en modelos y selección parámetros óptimos	41
5.1.6.	Modelado de parámetros óptimos y aplicación de umbral.....	41
5.1.7.	Diseño de mapa de distribución potencial	41
5.2.	Modelado de especies con registros menores a 20.....	41
5.2.1.	Flujo de trabajo.	42
5.2.2.	Selección de variables ambientales.....	42
5.2.3.	Preparación de datos de ocurrencia.	42
5.2.4.	Calibración y predicción.....	43
5.2.5.	Recodificación de valores y aplicación de umbrales.	43
5.2.6.	Prueba de significancia.	43
5.2.7.	Modelo final y segunda aplicación de umbral.....	43
5.2.8.	Diseño de mapa de distribución potencial.	43
IV.	Resultados y discusión	45
1.	Colección de ocurrencias	45
2.	Selección de variables ambientales.....	45
3.	Distribución potencial de especies	46
3.1.	<i>Buddleja incana</i> Ruiz & Pav.	46
3.2.	<i>Polylepis racemosa</i> Ruiz & Pav.	51
3.3.	<i>Cedrela Kuelapensis</i> T.D.Penn. & Daza	55
3.4.	<i>Tecoma rosaefolia</i> Kunth.....	58
3.5.	<i>Esenbeckia cornuta</i> Engl.....	62
V.	Conclusiones	67
VI.	Recomendaciones	69
VII.	Referencias bibliográficas	71
VIII.	Anexos.....	81

Índice de tablas

	Página
Tabla 1: Variables bioclimáticas de WorldClim.....	17
Tabla 2: Estadísticos derivados de matriz de confusión.....	24
Tabla 3: Especies seleccionadas de interés.....	36
Tabla 4: Fuente de Registros para las Especies de Interés.....	36
Tabla 5: Registros colectados para las especies en estudio.....	45
Tabla 6: Evaluadores de rendimiento para <i>Buddleja incana</i>	47
Tabla 7: Evaluadores de rendimiento para modelo final de <i>Buddleja incana</i>	49
Tabla 8: Evaluadores de rendimiento para <i>Polylepis racemosa</i>	52
Tabla 9: Evaluadores de rendimiento para modelo final de <i>Polylepis racemosa</i>	53
Tabla 10: Resultado de predicción de registro excluido en <i>Cedrela kuelapensis</i>	56
Tabla 11: Resultado prueba de significancia Pearson para <i>Cedrela kuelapensis</i>	56
Tabla 12: Resultado de predicción de registro excluido en <i>Tecoma rosaefolia</i>	59
Tabla 13: Resultado prueba de significancia Pearson para <i>Tecoma rosaefolia</i>	59
Tabla 14: Resultado de predicción de registro excluido en <i>Esenbeckia cornuta</i>	63
Tabla 15: Resultado de prueba de significancia Pearson para <i>Esenbeckia cornuta</i>	63

Índice de figuras

	Página
Figura 1: Representación de nicho fundamental y nicho realizado	4
Figura 2: Diagrama BAM	4
Figura 3: Espacio geográfico y espacio ecológico para una misma región.....	6
Figura 4: Esquema de flujo de trabajo para la realización y validación de modelos de nicho ecológicos y distribución de especies.	9
Figura 5: Crecimiento de registro de datos de biodiversidad en el portal GBIF.....	11
Figura 6: AUC versus tamaño de muestra para 12 algoritmos	13
Figura 7: Matriz de confusión.....	24
Figura 8: Curva ROC	26
Figura 9: Proceso de modelado de distribución potencial para especies con registros mayores a 20.....	38
Figura 10: Proceso de elección de variables.....	39
Figura 11: Preparación de grupos de entrenamiento y evaluación para especies con registros mayores a 20.....	40
Figura 12: Proceso de modelado de distribución potencial para especies con registros menores a 20.....	42
Figura 13: Cluster de asociación entre variables.....	46
Figura 14: Distribución de registros de <i>Buddleja incana</i>	47
Figura 15: Valores de $\Delta AICc$ versus multiplicadores de regularización en <i>Buddleja incana</i>	48
Figura 16: Importancia de variables en el modelo de <i>Buddleja incana</i>	49
Figura 17: Distribución potencial de <i>Buddleja incana</i>	50
Figura 18: Distribución de Registros de <i>Polylepis racemosa</i>	51
Figura 19: Valores de $\Delta AICc$ versus multiplicadores de regularización en <i>Polylepis racemosa</i>	52
Figura 20: Importancia de variables en el modelo de <i>Polylepis racemosa</i>	53
Figura 21: Distribución potencial de <i>Polylepis racemosa</i>	54
Figura 22: Distribución de Registros de <i>Cedrela kuelapensis</i>	55
Figura 23: Contribución de variables en el modelo de <i>Cedrela kuelapensis</i>	57
Figura 24: Distribución potencial de <i>Cedrela kuelapensis</i>	57
Figura 25: Distribución de Registros de <i>Tecoma rosaefolia</i>	59
Figura 26: Contribución de variables en el modelo de <i>Tecoma rosaefolia</i>	60

Figura 27:	Distribución potencial de <i>Tecoma rosaefolia</i>	61
Figura 28:	Distribución de Registros de <i>Esenbeckia cornuta</i>	62
Figura 29:	Contribución de variables en el modelo de <i>Esenbeckia cornuta</i>	64
Figura 30:	Distribución potencial de <i>Esenbeckia cornuta</i>	65

Índice de anexos

	Página
Anexo 1 Mapa de distribución potencial de <i>Buddleja Incana</i> Ruiz & Pav.	81
Anexo 2 Mapa de distribución potencial de <i>Polylepis racemosa</i> Ruiz & Pav.	82
Anexo 3 Mapa de distribución potencial de <i>Cedrela kuelapensis</i> T.D. Penn. & Daza	83
Anexo 4 Mapa de distribución potencial de <i>Tecoma rosaefolia</i> Kunth.....	84
Anexo 5 Mapa de distribución potencial de <i>Esenbeckia cornuta</i> Engl.	85
Anexo 6 Combinaciones de configuraciones de parámetros para <i>Buddleja incana</i> y <i>Polylepis racemosa</i>	86
Anexo 7 Evaluación de rendimiento de modelos para <i>Buddleja incana</i>	87
Anexo 8 Evaluación de rendimiento de modelos para <i>Polylepis racemosa</i>	88
Anexo 9 Script desarrollado en r studio para el modelado de <i>Buddleja incana</i>	89
Anexo 10 Script desarrollado en RStudio para el modelado de <i>Polylepis racemosa</i>	91
Anexo 11 Script desarrollado para el modelado de <i>Cedrela odorata</i>	93
Anexo 12 Script desarrollado en RStudio para el modelado de <i>Tecoma rosaefolia</i>	97
Anexo 13 Script desarrollado en RStudio para el modelado de <i>Esenbeckia cornuta</i>	104

I. INTRODUCCIÓN

El incremento de la población humana, la deforestación, la explotación excesiva y el creciente impacto climático, representan una creciente amenaza para la biodiversidad (Peterson *et al.* 2011). En el Perú especies como *Buddleja incana* y *Polylepis racemosa* se encuentran consideradas como amenazadas debido a su explotación irracional (IUCN 2016). Por otro lado, tenemos a especies endémicas poco conocidas como *Cedrela kuelapensis*, *Tecoma rosaefolia*, *Esenbeckia cornuta*, las cuales necesitan ser documentadas, pero debido al desconocimiento de sus distribuciones no pueden ser recolectadas con mayor frecuencia.

Los modelos de nichos ecológicos son de principal interés porque proporcionan una base predictiva para nuevas inferencias sobre la biodiversidad y su distribución en el espacio y tiempo. Una manera en que esta capacidad predictiva puede ser aprovechada es la de anticipar la distribución de nuevos elementos de la biodiversidad (poblaciones y especies) que aún no son ampliamente conocidas ni se documentan (Peterson *et al.* 2011). Esta aplicación puede clasificarse como uno de los usos más interesantes de estos modelos ya que se consolidan como una herramienta que permite un mayor descubrimiento y documentación de la biodiversidad, así como proporcionar potenciales áreas de reforestación para especies amenazadas.

En el Perú el desarrollo de investigaciones relativas a distribuciones potenciales de especies forestales son pocas. El Organismo de Supervisión de los Recursos Forestales y de Fauna Silvestre (OSINFOR), publicó las distribuciones potenciales de especies amazónicas para la región de Loreto; utilizando el algoritmo *maxent* y su configuración por defecto (OSINFOR, 2013; OSINFOR 2016). Särkinen *et al.* (2013) a través de modelos de nichos ecológicos descubrió una nueva especie de *Solanum* en los Andes Peruanos. Laurente (2015) modeló el efecto del cambio climático en *Cedrela odorata* L., para su idoneidad de hábitat; utilizó *maxent* y abarcó como área de estudio todo el territorio Peruano. Mejía (2014), realizó una propuesta metodológica para la cartografía del Género *Polylepis*, en el cual combinó técnicas de modelado de nicho ecológico, interpretación visual de imágenes satelitales y muestreo en campo; el área de estudio abarcó el departamento de Arequipa y Lima. Utilizó cuatro

algoritmos (Domain, Bioclim, Garp, Maxent) y tres diferentes bases de datos, donde obtuvo como resultado que los modelos con mejores predicciones fueron los que utilizaron *maxent* y técnicas de ensamblado (promedio de modelos).

Zutta *et al.* (2012) desarrollaron la distribución potencial de *Polylepis sericea* y *Polylepis besseri* para Sudamérica; utilizaron el algoritmo *maxent*, logrando elevados rendimientos de predicción.

Sin embargo, prevalece el vacío de información de estos estudios a escala nacional para especies andinas y de regiones secas.

El presente trabajo de investigación tiene como objetivo determinar la distribución potencial de cinco especies forestales amenazadas y endémicas raras, a escala nacional, mediante modelos correlativos con base ecológica. Así mismo, los objetivos específicos consisten en establecer las variables ambientales de mayor importancia en la distribución potencial en el modelado de las especies objeto de estudio y comparar las distintas combinaciones de complejidad del algoritmo *maxent* en los modelos, mediante la evaluación de su capacidad predictiva.

II. REVISIÓN DE LITERATURA

1. BASES CONCEPTUALES PARA LA MODELIZACIÓN DEL NICHOS ECOLÓGICO

1.1. DEFINICIÓN DE NICHOS ECOLÓGICO

Johnson (1910), utilizó por primera vez la palabra nicho como un término ecológico, dándole de manera intuitiva el significado que ahora tiene. Según Grinnell (1917), el nicho ecológico de una especie se puede definir como la combinación de condiciones ecológicas que permiten mantener la población sin inmigración.

Hutchinson (1957) manifiesta que, «el termino nicho se define como la suma de todos los factores que actúan en un organismo; así el nicho se define como un hiperespacio n - dimensional». Enfatizando que estos factores pueden ser físicos y biológicos, en este concepto el nicho es propiedad de la especie; no del ambiente. Así, propone que los nichos cambian, son *mutables*, evolucionan. Este concepto de nicho se puede cuantificar, ya que, variables ambientales se pueden relacionar con el desempeño de una especie; por lo cual es posible *graficarlo* en el hiperespacio.

Este último autor propone dos tipos de nichos ecológicos: (1) *el nicho fundamental*: todos los aspectos del hiperespacio de n -dimensiones donde la especie puede vivir; (2) *el nicho realizado*: subconjunto del nicho fundamental en el cual las especies están restringidas por sus interacciones inter-específicas (ver la figura 1).

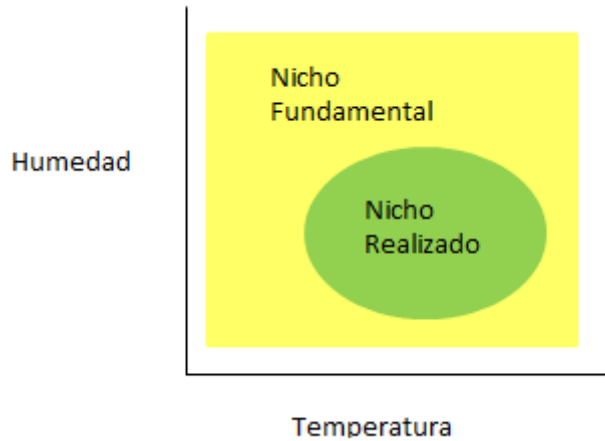


Figura 1: Representación de nicho fundamental y nicho realizado

FUENTE: Adaptado de Martínez-Meyer 2014.

Existen diversos conceptos de nicho ecológico manifestado por los investigadores en ecología; sin embargo, en el marco de esta investigación adoptaremos el de Hutchinson (1957).

1.2. DIAGRAMA BAM

El área de distribución de una especie, es el resultado de una combinación de factores en un tiempo dado (Grinnell 1917, Good 1931, Cain 1944, Udvardy 1969). Soberón y Peterson (2005) presentaron el diagrama BAM (Figura 2) donde se manifiesta la relación de tres factores primordiales en el espacio geográfico (G):

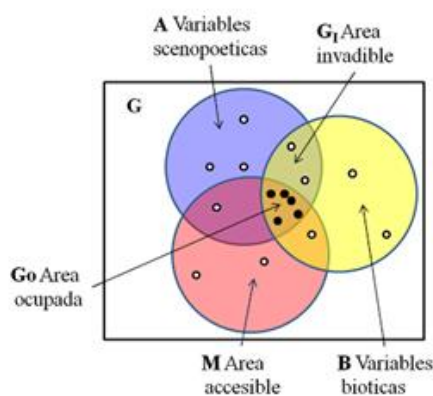


Figura 2: Diagrama BAM

FUENTE: Adaptado de Soberón y Peterson 2005.

- La región del mundo donde se encuentran las condiciones ambientales que permiten la sobrevivencia de la especie (ese conjunto de condiciones se llama *nicho fundamental* de la especie). Denotado con la letra A en el diagrama BAM, se calcula usando variables *scenopoéticas* (variables abióticas que no interaccionan con la especie).
- La región del mundo que contiene el ambiente biótico adecuado para la especie. Esto implica interacción con otros organismos que favorezcan a la especie. Se denota con la letra B en el diagrama BAM. Este tipo de información es escaso.
- El conjunto de sitios que la especie ha podido “explorar” desde una fecha dada, se denota con la letra M.

Finalmente, Soberón y Peterson (2005) manifiestan que estos conceptos permiten definir dos importantes regiones: el área ocupada (Go) y el área invadible (GI). El área ocupada Go es la parte del mundo que cuenta con condiciones ambientales y bióticas favorables y que ha estado accesible a la especie. El área invadible GI son regiones que cuentan con condiciones ambientales y bióticas favorables, pero no han sido accesibles a la especie (existencia de barreras).

En el modelado del nicho ecológico (MNE) y modelado de distribución de especies (MDEs) es muy común confundir lo que se obtiene, si Go, GI o A. Si se realiza el trabajo con registros de presencia y ausencias verdaderas, mediante algoritmos de regresión, se está aproximando a Go. Sin datos de ausencia verdadera, los diferentes algoritmos están produciendo estimaciones de A (Jiménez-Valverde *et al.* 2008, Soberón y Nakamura 2009).

1.3. MODELADO DE NICHOS ECOLÓGICOS (MNE) Y MODELADO DE DISTRIBUCIÓN DE ESPECIES (MDES)

En la literatura existe un debate, hasta cierto grado semántico (Elith y Leathwick 2009, Franklin 2009), sobre el nombre correcto para la actividad de modelar áreas de distribución. Estos dos términos no son equivalentes y percibirlos como tales lleva a confusión y problemas de interpretación de resultados. Soberón *et al.* (2017) manifiestan que cuando se intenta modelar el área ocupada (implica factores bióticos, abióticos y área invadible), lo natural es hablar de MDEs. Si se intenta modelar las áreas potenciales, lo cual esencialmente implica localizar en la geografía las regiones favorables a la especie (en función de factores abióticos) entonces lo lógico sería usar el término MNE.

1.4. DUALIDAD DE HUTCHINSON

La idea fundamental tanto en MDEs como en MNE, es que existen dos espacios. El geográfico denotado por G y el espacio ecológico denotado por E, este último abstracto. El espacio ecológico y el espacio geográfico se relacionan directamente, ya que, en un tiempo dado; a cada punto de G corresponde una combinación de variables ambientales en E. Esta combinación ser o no única, dependiendo de la resolución de la información ambiental, del número de variables usadas, y de su precisión (CONABIO 2014).

Según Hutchinson (1978), el espacio ecológico (E) está compuesto por variables *scenopoéticas*, es decir, variables que establecen las condiciones para que una especie pueda simplemente existir.

La Figura 3, se muestra un ejemplo de espacio geográfico, específicamente la extensión del continente americano. A la derecha, se muestra el espacio ecológico correspondiente a la misma región usando tres variables medioambientales, donde el número de celdillas del espacio geográfico es igual al número de puntos del espacio ecológico. Sin embargo, esto no es una regla, ya que, a un punto en el espacio ecológico le pueden corresponder varias celdillas del espacio geográfico, sin embargo, una celdilla en el espacio geográfico con un número de variables ambientales constantes solo le puede corresponder un punto en el espacio ecológico (CONABIO 2014).



Figura 3: Espacio geográfico y espacio ecológico para una misma región.

FUENTE: Adaptado de CONABIO 2014.

1.5. APLICACIONES DEL MODELADO DE NICHO ECOLÓGICO

Un modelo de nicho ecológico es básicamente una caracterización de condiciones ambientales adecuadas para las especies a partir de una muestra de presencia, que ayuda a identificar dónde están espacialmente ubicados los sitios que cumplen con los requisitos adecuados para las especies en estudio (Pearson *et al.* 2007a).

Entre las diversas aplicaciones que tiene el modelado de nicho ecológico (MNE), se pueden nombrar las siguientes:

a) Conservación y restauración ecológica

El MNE combinado con herramientas SIG ayuda en la generación de mapas de distribución potencial de especies que se encuentran en peligro, especies poco conocidas o especies nativas, lo que resulta fundamental tanto para evaluar el estado de conservación de las especies, como para evaluar potenciales sitios para desarrollar iniciativas de conservación y restauración. En el caso de las especies poco conocidas permitirá desarrollar mayor investigación sobre ellas para su posterior documentación (Morales 2012).

b) Cambio climático

Factores climáticos controlan los límites de distribución de especies, por lo que estos límites son determinados utilizando métodos centrados en las relaciones entre especies y clima. Tal situación requiere una estimación cuidadosa y completa de todos los límites del nicho ecológico. Se puede formular varias generalidades sobre la forma en que el cambio climático probablemente generará efectos sobre las especies. Resultados de modelado de nicho ecológico sugieren que climas cambiantes tendrán efectos espaciales dramáticos para la idoneidad del hábitat (Peterson *et al.* 2011).

El modelado de nichos ecológicos permitirá predecir cómo variará la distribución potencial de una especie ante cambios de temperatura, humedad, precipitación, etc. Por ejemplo, se podría evaluar cómo se vería afectada la producción frutícola al ver restringida su área de distribución o, al contrario, al expandirse su potencial distribución a causa de un aumento de la temperatura atmosférica (Morales 2012).

c) Invasión de especies

Morales (2012) menciona que, el conocimiento de distribuciones potenciales de especies, con base en modelos de nichos ecológicos puede ser utilizado para determinar zonas más propensas a plagas. Esta información será de utilidad para el diseño de muestreos preventivos, enfocándose a zonas específicas, logrando con ello un mayor nivel de eficiencia y, por consiguiente, ahorro de recursos económicos También señala que es posible estimar dónde puede iniciarse una invasión y los patrones espaciales de distribución que seguirá. Este

enfoque de aplicación es fundamental para el control de invasiones y establecimiento de medidas preventivas ante este fenómeno.

2. MODELADO DE NICHOS ECOLÓGICOS

Peterson *et al.* (2011) indican que el modelado de nicho ecológico (MNE), es un proceso que consiste en caracterizar cada celda dentro de una región definida (área de estudio), en términos de valores cuantitativos relacionados con su idoneidad de hábitat en función de las condiciones ambientales presentadas en esa celda. Esta caracterización se dará a partir de puntos de presencia conocida. La caracterización de los valores de idoneidad es resultado de una función matemática, estimada entre puntos de presencia y los valores ambientales de cada celda. Dicha función puede producir diferentes resultados, dependiendo del algoritmo utilizado. Una consideración importante aquí es la finalidad por la que se está desarrollando el modelo (Peterson 2006). Así, una consideración inicial básica es si el modelo está destinado a fines de predicción o de explicación (Araujo y Guisan 2006).

Los modelos de nichos orientados a la predicción pueden utilizar métodos y datos que producen predicciones óptimas, pero que generan poca información para interpretar la relación entre las variables ambientales y la idoneidad de hábitat obtenida. Por otra parte, los modelos de nichos orientados a la explicación de variables ambientales pueden generar amplia información que manifieste la interacción entre variables, aunque no necesariamente producirán las mejores predicciones geográficas (Peterson *et al.* 2011). Sin embargo, algunos modelos pueden cumplir ambos fines de manera satisfactoria.

El proceso de MNE se basa en una serie de premisas que guían el flujo de trabajo. Mateo *et al.* (2011) plantean un esquema (Figura 4) para el modelado del nicho ecológico y de la distribución de especies.

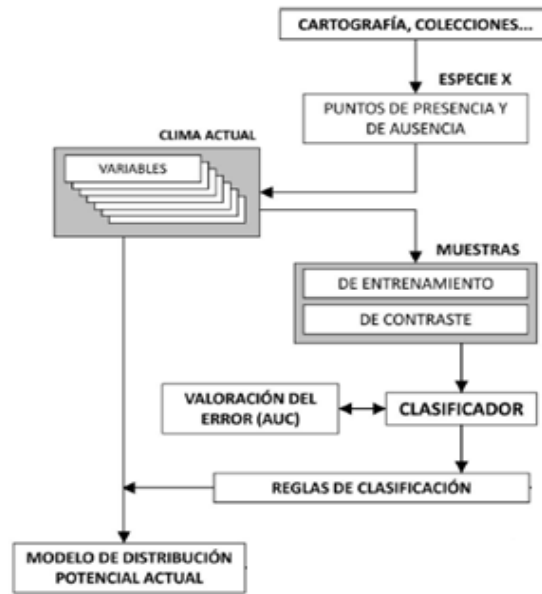


Figura 4: Esquema de flujo de trabajo para la realización y validación de modelos de nicho ecológicos y distribución de especies.

FUENTE: Adaptado de Mateo *et al.* 2011.

Como se observa en la Figura 4, se parte de datos de ocurrencia que pueden representar solo presencia, o presencia – ausencia, siendo estos últimos más difíciles de encontrar. Las variables ambientales pueden ser de clima o diversas como fisiografía, altitud, pendiente, etc. El clasificador está constituido por el algoritmo a utilizar; la valoración de error se realiza a través de diversos estadísticos explicados con más detalle en las siguientes páginas. Finalmente, las reglas de clasificación son conocidas como umbrales de decisión, siendo estas no obligatorias y a criterio del investigador, ya que se puede presentar mapas continuos de distribución potencial o mapas binarios de acuerdo al objetivo y al tamaño de muestra de presencias.

Lógicamente el proceso anterior es muy general y es necesario hacer explícito cada paso desde los datos a trabajar hasta la evaluación del modelo y umbral de decisión (Mateo *et al.* 2011).

Londoño (2012) define como fortalezas del MNE: (1) potencializa el uso de registros geográficos de especies, (2) es posible modelar distribuciones potenciales con pocos registros de presencia, (3) es posible generar hipótesis espaciales en escenarios geográficos distintos en tiempo y espacio. Como limitaciones señala: (1) la representatividad de los datos de entrada, (2) la incertidumbre de algoritmos para proyecciones en espacio y tiempo, (3) no incorporar

otros procesos importantes como barreras geográficas, capacidad de dispersión, interacciones bióticas, procesos históricos.

2.1. TIPOS DE DATOS

Según Graham *et al.* (2004), existen tres enfoques básicos para el montaje y uso de datos para la modelización: (1) utilizar solo datos de presencia, (2) usar datos de presencia y ausencia cuando estén disponibles, (3) utilizar datos de presencia y una muestra de *background* o datos de *pseudoausencia* en lugar de ausencia

Es necesaria una cuidadosa consideración de los datos de ausencia para que el MNE basado en esta información sea rigurosa, exitosa y no repercuta en sesgos adicionales. Datos de ausencia u otros, son requeridos por varios enfoques de modelado que utilizan datos de presencia y algún tipo de ausencia, *pseudoausencia* o *background* (Hirzel *et al.* 2002, Elith *et al.* 2006).

2.1.1. DATOS DE PRESENCIA

CONABIO (2014) indica que son los datos que registran la ocurrencia de la especie en un determinado lugar, los cuales deben ser datos primarios (observaciones de la especie), los datos secundarios (atlas) no deben ser usados en MNE/MDEs. También señala que el control de calidad de los datos de ocurrencia debe cumplir los siguientes requerimientos: (1) la nomenclatura debe ser consistente respecto a un concepto taxonómico, (2) la georeferenciación debe de hacerse de acuerdo a un sistema internacional estandarizado, (3) el radio de incertidumbre de los puntos georeferenciados se debe filtrar para retener solo los registros que cumplan un nivel de precisión adecuado al problema y razonablemente coincidentes con la resolución de los datos ambientales.

Empezando en la década de 1990, se iniciaron esfuerzos en varias regiones para ofrecer un acceso integrado y eficaz a datos de Biodiversidad (Soberón 1999). De acuerdo a Peterson *et al.* (2011), las primeras tecnologías han dado paso a eficaces protocolos de transferencia de información que se han desarrollado específicamente para la transmisión de información sobre biodiversidad. La Infraestructura de Información en Biodiversidad (en inglés GBIF), instituida en el año 2001, cuenta a la fecha cerca de 728 millones en registros de ocurrencia de especies. La disponibilidad de registros digitales de ocurrencia primaria ha ido en aumento rápidamente a través del tiempo. En la Figura 5, se puede observar cómo se han incrementado

los datos con el transcurso de los años, según los registros de que se dispone en el GBIF una de las más importantes bases de datos para el estudio de la biodiversidad.



Figura 5: Crecimiento de registro de datos de biodiversidad en el portal GBIF

FUENTE: Adaptado de Soberón y Peterson 2008.

En las últimas décadas se han producido mejoras en los sistemas de acopio de datos y el aumento de los mismos se ha producido de forma exponencial, redundando en una mayor disponibilidad de información sobre la biodiversidad. Lo que antes era un reto al recolectar información en formato analógico, hoy se ha convertido en grandes bases de datos digitales ampliamente disponibles y de manera rápida. Sin embargo, en la emoción de esta “nueva era” de acceso a los datos, es importante tener en cuenta la consideración de procedencia, calidad y posibles sesgos que los datos pueden llevar. Es de suma importancia que los datos de presencia además de sus coordenadas y nombre de la especie, lleven información documentada referente a la entidad recolectora, el año de recolección, la precisión del equipo con el que se tomó el dato y la localidad de recolección, siendo estos lo que se denomina “metadata” (Peterson *et al.* 2011). Como fuente de datos primarios muy importantes además de GBIF encontramos a *Lifemapper*, *SpeciesLink*, *VertNet*, etc.

Benito (2015) indica que los enfoques basados en solo-presencia exhiben como ventaja: una alta disponibilidad (colecciones virtuales), mientras que sus desventajas son: (1) no permiten calcular probabilidad de presencia, (2) no informan sobre sesgos en el muestreo, (3) requieren mayor precisión en coordenadas geográficas.

a) Numero de datos de ocurrencia

La cuestión de los tamaños mínimos de muestra requerida para producir predicciones útiles permanece aún entre los temas difíciles de establecer. Pearson *et al.* (2007) desarrollaron y evaluaron modelos para 13 especies en Madagascar con tamaños de muestra que variaban entre 4 a 23 registros de presencia. La predicción se basó sobre 20 capas ambientales (en formato raster de resolución de 1 km), utilizando los métodos de *maxent* y GARP. Obtuvieron como resultado altas tasas de éxito con tamaños de muestra tan bajo como 5 registros de presencia cuando se utilizó *maxent*. Los resultados que obtuvieron para GARP a tamaños de muestras menores a diez fueron menos exitosos. Este estudio concluyó que los modelos desarrollados con un enfoque de pequeñas muestras, deben interpretarse como regiones identificadas con condiciones ambientales similares aquellas prevalecientes en donde se registró la presencia y no como la predicción de límites reales de una especie. Finalmente, indicaron que su análisis demostró que «las predicciones geográficas a partir de pequeños tamaños de muestras pueden ser de gran valor para acelerar el descubrimiento de poblaciones y especies desconocidas».

Por otro lado, Wisz *et al.* (2008) evaluaron la predicción de 12 algoritmos para 46 especies de seis regiones del mundo con tres tamaños de muestra (100, 30 y 10 registros). El objetivo era evaluar la sensibilidad de los algoritmos ante los tamaños de muestra. Evaluaron la calidad de predicción de los modelos a través del área bajo la curva (AUC por sus siglas en ingles). Determinaron que *maxent* fue el algoritmo menos sensible al tamaño de muestra, teniendo la mejor potencia predictiva en todos los tamaños de muestra en relación a otros algoritmos. Sin embargo, concluyeron que la tendencia presentada en la mayoría de algoritmos fue que a menor tamaño de muestra había un deterioro en el rendimiento de la predicción. En la Figura 6, se observa el AUC evaluado versus el tamaño de muestra para cada algoritmo que fue utilizado.

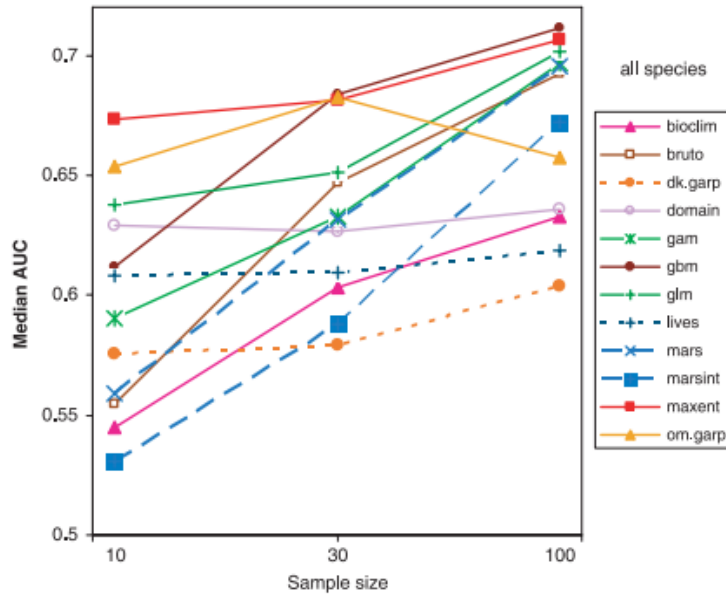


Figura 6: AUC versus tamaño de muestra para 12 algoritmos

FUENTE: Tomado de Wisz et al. 2008

2.1.2. AUSENCIA ESTRICTA

Son las observaciones confiables (generalmente de un inventario derivado de campo) sobre la no presencia de una especie en un sitio. Estos datos permiten el uso de modelos de regresión para estimar áreas ocupadas; sin embargo, lo más frecuente es que estos datos no estén disponibles y su interpretación no es inmediata, ya que no es igual “ausencia del área potencial” para modelos de nichos ecológico que “ausencia del área ocupada” para modelos de distribución de especies (CONABIO 2014).

Benito (2015) indica que los enfoques que trabajan con presencia-ausencia manifiestan como ventajas que: (1) permiten calcular probabilidad de presencia, (2) tienen en cuenta el sesgo en el muestreo. Asimismo, indica como desventajas: (1) baja disponibilidad, (2) que la ausencia en realidad sea una falsa ausencia (producida por intervención humana como deforestación, cambio uso de suelo, falta de dispersión, etc.).

2.1.3. PSEUDOAUSENCIA.

Dada la escasez de datos de ausencia, los investigadores suelen recurrir a la generación de *pseudoausencia* por numerosos métodos (Fielding y Bell 1997, Stockwell y Peterson 2002). Las *pseudoausencias* son datos generados para simular las ausencias, tomando una muestra al azar de puntos en el área de distribución o especificando una región conocida hipotetizada

como desfavorable para la especie. Estas ausencias simuladas deben ser utilizadas con mucho cuidado ya que pueden generar un gran sesgo en el modelo, no obstante, si se tiene conocimiento ecológico amplio de la especie se puede determinar con mayor precisión las regiones desfavorables donde se puede generar un muestreo aleatorio de *pseudoausencias* (CONABIO 2014).

Según Peterson *et al.* (2011), a pesar que la generación de *pseudoausencias* tiene como intención imitar las ausencias, tales datos sufren típicamente de la inclusión de pixeles provenientes de áreas que en realidad la especie podría ocupar, lo que dificulta que se logre lo pretendido inicialmente. En consecuencia, cuando se trabaja con *pseudoausencia* o *background* los análisis resultantes deben interpretarse de manera diferente de aquellos basados en ausencia reales.

Benito (2015) indica que los enfoques que trabajan con presencia-*pseudoausencia* manifiestan como ventaja: (1) *pseudoausencias* son fáciles de generar. Asimismo, indica como desventaja que: (1) no son ausencias reales, (2) no tienen en cuenta sesgo en el muestreo, (3) el criterio ecológico para generarlas es discutible.

2.1.4. “FONDO” O BACKGROUND.

De acuerdo con Elith y Phillips (2011), son pixeles muestra tomados al azar del área de referencia completa para cada variable ambiental. *Maxent* los utiliza para calcular el modelo de nicho ecológico. Los datos de *background*, que en *maxent* se usan para caracterizar el ambiente, y los datos de *pseudoausencias*, que en regresión se usan como substitutos de ceros, son en concepto totalmente distintos, y no se deben confundir ni comparar (Elith y Phillips 2011).

Peterson *et al.* (2011) señalan que, aunque los términos *pseudoausencia* y *background* se utilizan a veces indiscriminadamente, el muestreo de *background* en su conjunto se puede utilizar para caracterizar las condiciones ambientales presentes a través de la región de estudio (es decir, creando una muestra del "fondo" que potencialmente incluye sitios donde se ha observado la presencia de la especie). En contraste, el muestreo por *pseudoausencia* implica la selección de áreas o sitios donde la especie no ha sido detectada (o a veces sólo del subconjunto de no detección de sitios donde se ha realizado el muestreo de presencia).

Benito (2015) indica que los enfoques que trabajan con presencia-*background* manifiestan como ventajas: (1) las muestras de *background* son fáciles de generar, (2) no hay problema si

solapan con las presencias. Asimismo, indica como desventajas que: (1) no funciona bien para métodos como *Random Forest*, *Support Vector Machines*, (2) al trabajar con métodos de regresión, el *background* debe ponderarse.

2.2. VARIABLES AMBIENTALES

Según Peterson *et al.* (2011), el modelado de nicho ecológico (MNE) requiere dos fuentes de datos de entrada: (1) los puntos de ocurrencia conocidos de la especie y (2) predictores ambientales en capas de tipo raster (formato tradicional de sistemas de información geográfica para representar información de manera espacial). La importancia de estas variables ambientales radica en la relación que tienen con la especie y en cómo ello afecta su distribución. Las variables macro climáticas influyen en las distribuciones más gruesas, mientras que las características del paisaje (por ejemplo, cubierta vegetal) actúan a meso escala, y las características específicas del hábitat y las interacciones bióticas tienen las influencias más fuertes a escala local, esto último enfocado cuando se hace MNE para animales (Mackey y Lindenmayer 2001, Pearson y Dawson 2003).

Una pregunta clave es, por tanto: ¿cuáles y cuántas variables son necesarias para modelado del nicho de una especie?, las respuestas dependen claramente de la escala a la que se aborda el problema, los conocimientos disponibles sobre la auto-ecología de la especie en cuestión y la disponibilidad de datos de alta calidad (Peterson *et al.* 2011).

Gracias al crecimiento de las técnicas de modelización espacial, el sensoramiento remoto y los sistemas de información geográfica (GIS), se dispone cada vez más de datos ambientales digitales de alta calidad. Estos datos se encuentran en diversas extensiones y resoluciones espaciales desde global a local con tamaños de píxeles desde aproximaciones a un kilómetro (Peterson *et al.* 2011). La mayoría de estas variables pertenecen al conjunto de variables *scenopoéticas* (clima, topografía, cobertura de suelo) perteneciente al diagrama BAM detallado en la página 4, las cuales son las utilizadas en el MNE.

Algunas variables ambientales son más estáticas que otras, en el sentido de que sólo cambian lentamente (por ejemplo, topografía y clima), mientras que son más dinámicas (por ejemplo, cobertura de la tierra), especialmente dada la influencia generalizada de los seres humano sobre el medio ambiente (Peterson *et al.* 2011).

2.2.1. SELECCIÓN DE VARIABLES AMBIENTALES

Gran parte de los investigadores prefieren utilizar variables preseleccionadas, que no se encuentren fuertemente correlacionadas y respondan a reglas fisiológicas de la especie, tales como temperatura y estrés hídrico (Huntley *et al.* 1995, Pearson *et al.* 2002, Huntley *et al.* 2004, Baselga y Araújo 2009). Por otro lado, cuando no existe conocimiento previo sobre factores claves de la especie se han usado grandes conjuntos de datos (Stockwell 2006), aunque con fuertes críticas (Peterson *et al.* 2007b). Ambos extremos tienen sus inconvenientes: cuando se utilizan pocas variables se tiende a la subcaracterización del nicho; mientras, la sobre-dimensionalidad (exceso de variables) puede conducir a modelos de nichos excesivamente complejos, que pueden complicar el desarrollo del modelo y su caracterización para el objeto de estudio (Peterson *et al.* 2007b).

Como consecuencia, para evitar estos problemas, se han propuesto varias técnicas para explorar la interrelación entre las variables y reducir su dimensionalidad antes de iniciar el modelado.

El *análisis de correlación múltiple* ha sido ampliamente utilizado por investigadores para identificar los conjuntos menos correlacionados de las variables (Baselga y Araújo 2009). Otros han reducido la dimensionalidad de variables a través de *análisis de componentes principales*, el cual genera nuevas variables (en menor número que las originales) con la información sintetizada de las originales (Manel *et al.* 2001, Hirzel *et al.* 2002, Peterson *et al.* 2007b). Puede resultar que una variable sea producto de la combinación lineal de dos variables, para lo cual se sugiere usar *Variance Inflation Factor* (VIF) con el fin de disminuir la colinealidad entre variables (Benito 2015).

Benito (2015) define los requerimientos técnicos para que las variables ambientales (en formato raster) puedan ingresar al modelado: (1) matrices idénticas (igual resolución, igual extensión, igual celdas con valores nulos); (2) baja colinealidad entre ellas.

CONABIO (2014) sugiere que en general se debe evitar utilizar menos de diez variables o menos variables que puntos de ocurrencia.

2.2.2. FUENTE DE DATOS AMBIENTALES

Según CONABIO (2014), existen diferentes fuentes digitales para la obtención de variables ambientales digitales, una de las más conocidas es *WorldClim*. Esta base de datos contiene data de capas climáticas, con resoluciones desde 1 km, obtenidas de la interpolación de

registros de estaciones climáticas entre 1950 – 2000, procedentes de diversas fuentes a nivel global, regional, nacional y local. Las 19 variables bioclimáticas (Tabla 1) son derivadas de la temperatura media anual, precipitación multianual, temperatura mínima anual y temperatura máxima anual (Hijmans *et al.* 2006). Para productos de datos topográficos y derivados se encuentra *Earth Resources Observation and Science (EROS)* del servicio geológico americano. Si se desea trabajar con datos derivados de percepción remota como índices de vegetación, la plataforma de *Global Land Cover Facility* es una buena opción (CONABIO 2014).

Tabla 1: Variables bioclimáticas de WorldClim

Código	Variables bioclimáticas
BIO1	Temperatura promedio anual
BIO2	Promedio del rango diario (Promedio mensual (T° máx. - T° min.))
BIO3	Isotermalidad (BIO2/BIO7)* 100
BIO4	Temperatura estacional (Desviación estándar *100)
BIO5	Temperatura máxima del periodo más caliente
BIO6	Temperatura mínima del periodo más frío
BIO7	Rango anual de temperatura (BIO5-BIO6)
BIO8	Temperatura media en el trimestre más lluvioso
BIO9	Temperatura promedio en el trimestre más seco
BIO10	Temperatura promedio en el trimestre más caluroso
BIO11	Temperatura promedio en el trimestre más frío
BIO12	Precipitación anual
BIO13	Precipitación en el periodo más lluvioso
BIO14	Precipitación en el periodo más seco
BIO15	Estacionalidad de la precipitación (Coeficiente de variación)
BIO16	Precipitación en el trimestre más lluvioso
BIO17	Precipitación en el trimestre más seco
BIO18	Precipitación en el trimestre más caluroso
BIO19	Precipitación en el trimestre más frío

FUENTE: *Elaboración propia*

2.3. MÉTODO DE MODELADO

Mateo *et al.* (2011) realizó una síntesis de los métodos más conocidos en el modelado de nicho ecológico, clasificándolos en tres tipos de técnica de clasificación:

a) Técnicas discriminantes

Son aquellas que necesitan datos de presencia como de ausencia (*pseudopresencia* o *background*), para construir el clasificador. Algunos ejemplos son: *random forest*, *boosted regression trees*, redes neuronales, *multivariate adaptive regression splines* (MARS),

modelos lineales generalizados (GLM), modelos aditivos generalizados (GAM), estadística bayesiana y *maxent*. Este último genera sus propias “ausencias” denominadas *background*, para caracterizar las condiciones ambientales generales, mientras que en las otras técnicas estos puntos representarían condiciones en las que no vive el organismo. En este último caso, la selección se torna crítica, ya que pueden inflar el número de falsos negativos.

b) Técnicas descriptivas

Estas técnicas, que solo requieren datos de presencia, fueron las primeras utilizadas en modelización. Algunos ejemplos son: superficies geográficas, superficies ambientales (BIOCLIM, ANUCLIM, BIOMAP), métodos de distancias matemáticas (ENFA, DOMAIN, Biomapper)

c) Técnicas mixtas

Se emplean varias reglas entre ellas descriptivas y otras discriminantes, a la vez que generan sus propias *pseudoausencias*. Desktop-GARP (Stockwell y Peters 1999) y OM-GARP (Elith *et al.* 2006) son los más ampliamente conocidos. El “ensamble de modelos” definido como una combinación de varios modelos generados con los mismos datos, surge como una nueva opción para reducir la incertidumbre de los modelos (Marmion *et al.* 2009).

2.3.1. SELECCIÓN DEL MÉTODO ADECUADO.

La elección del algoritmo suele ser un reto hasta la actualidad. A pesar que se han realizado una diversidad de estudios para comparar algoritmos, no existe una guía sólida para el proceso de elección (Elith *et al.* 2006, Guisan *et al.* 2007, Ortega-Huerta y Peterson 2008, Wisz *et al.* 2008). Sin embargo, a pesar de que existen muchas opciones, los requisitos de un algoritmo útil incluyen lo siguiente: (1) si los datos disponible son solo presencia se debe utilizar un algoritmo que trabaje con estos datos (*Domain*, *Bioclim*, *Garp*) o con datos de presencia-background (*maxent*); (2) el método debe tener el potencial para replicar respuestas complejas de las especies a las variables ambientales; siendo *Domain* y *bioclim* enfoques extremadamente simples, no son recomendables; en cambio, *maxent* y *Garp* sí pueden replicar respuestas complejas; (3) se debe tomar en cuenta la sensibilidad del algoritmo al tamaño de muestra, pues algunos algoritmos requieren grandes cantidades de datos para funcionar eficazmente (Wisz *et al.* 2008); (4) la elección debe tomar en cuenta el objetivo de estudio, ya que algunos algoritmos son mejores para predecir áreas, mientras que otros son mejores para testear hipótesis (Benito 2015).

Se presenta una breve revisión de los algoritmos más utilizados:

a) Máxima entropía: *maxent*

Maxent es un algoritmo cuya teoría está sumamente documentada y que en la práctica ha demostrado dar buenos resultados. Está basado en una aproximación estadística llamada “máxima entropía” que permite hacer predicciones utilizando información incompleta, en este caso datos de presencia u ocurrencia de la especie (Phillips *et al.* 2006). *Maxent* se basa en el principio de que la distribución estimada de la especie debe coincidir con la distribución deducida a partir de las condiciones ambientales donde ha sido observada, evitando hacer cualquier suposición que no sea soportada por los datos (Ferrer 2012). Sin embargo, es un método que tiende a sobre-ajustar los datos, cuando los parámetros de configuración están por defecto. Para generar modelos suavizados, *maxent* permite al usuario configurar los “features” y “multiplicador de regularización” por el usuario. Los “features” son las transformaciones y combinaciones posibles que se le da a las variables. La “regularización” evita un ajuste demasiado rígido a los datos; a su vez selecciona los nuevos predictores derivados de los “features” que más contribuyen al ajuste del modelo reduciendo el coeficiente de los demás a cero (Benito 2015).

Según Silver (2012), se habla de sobreajuste cuando un algoritmo de aprendizaje puede adaptarse muy bien a los datos de entrenamiento, pero presenta un mal desempeño en nuevos datos, Esto quiere decir que ha fracasado en generalizar y captar la señal o esencia de los datos sobre el fenómeno estudiado. Una forma útil de determinar si el modelo está sobreajustado, es evaluar si el rendimiento del modelo en su conjunto de entrenamiento es mucho mejor que en su conjunto de validación (evaluadores como AUC). El sobreajuste está relacionado con la complejidad del modelo, mientras más complejidad del modelo mayor tendencia al sobreajuste. Para reducir el sobreajuste se disminuye la complejidad del modelo, con lo que se ganaría generalización. Sin embargo, si se reduce la complejidad del modelo en exceso se podría crear un modelo infraajustado con un desempeño muy bajo, con un alto nivel de sesgo. Se debe buscar un término medio de complejidad en función del ajuste de parámetros (Silver 2012).

Como se mencionó, los “features” son las formas como van a ser transformadas las variables predictoras, las cuales pueden ser: (1) *features linear*, que utilizan coeficientes lineales simples para cada uno de los predictores, (2) *features quadratic*, que utilizan valores de

predicción al cuadrado, (3) *features product*, describen la interacción de pares entre predictores, (4) *features threshold* implica funciones simples de escalera o paso (5) *features hinge*, combina funciones lineales y de paso. El multiplicador de regularización por defecto es uno, aumentando el valor si se desea modelos menos ajustados y disminuyéndolo en caso contrario (Phillips y Dudík 2008).

Anderson y González (2011), compararon diferentes combinaciones de parámetros de *maxent*, para determinar la más óptima que minimice el sobreajuste. Su resultado expuso que en varios casos el multiplicador de regularización óptimo no era el configurado por defecto. Otros estudios también han demostrado que los “features” y multiplicador de regularización por defecto, no son necesariamente los más apropiados (Warren y Seifert 2011, Shcheglovitova y Anderson 2013). Considerando que cada vez hay más pruebas que el uso de parámetros por defecto de *maxent* no genera siempre el mejor modelo posible, se destaca la importancia de evaluar las mejores combinaciones de estos parámetros (“features” y multiplicador de regularización) antes de decidir sobre el mejor modelo (Merow *et al.* 2013).

Morales *et al.* (2016), realizaron un estudio donde se revisó 244 artículos de modelado de nicho ecológico procedentes de 142 revistas, entre el 2013 y 2015, para evaluar si los investigadores seguían recomendaciones de evitar la configuración predeterminada de parámetros, o si utilizaban *maxent* como una herramienta de caja negra. Los resultados mostraron que solo 16 por ciento de artículos analizados evaluaron los mejores “features”, el 6.9 por ciento evaluaron los mejores multiplicadores de regularización y un escaso 3.7 por ciento evaluó ambos parámetros antes de producir el modelo definitivo. Finalmente concluyeron que estos resultados son preocupantes, dado que modelos demasiados complejos o demasiados simplistas pueden socavar la aplicabilidad del modelo.

El producto del modelado en *maxent* es un raster que contiene en sus celdas el valor “raw output” que es una medida de similitud de los ambientes en cada pixel a los pixeles donde la especie ha sido observada. Podría considerarse como una aproximación a la distribución potencial. La suma de los valores del “raw output” sobre todas las celdas vale uno y, por lo tanto, en retículas grandes, tiene valores muy pequeños. La salida “logistic output” de *maxent* es esencialmente un intento de producir una probabilidad sobre cada pixel x, pero no incluye nueva información sobre la presencia de la especie. Simplemente transforma los valores de “raw output” a una nueva escala, sin incluir nueva información (CONABIO 2014).

A pesar de que existe una amplia gama de modelos similares, *maxent* posee dos ventajas por sobre los demás: (1) requiere solo datos de presencia, versus otros modelos que requieren datos de presencia y ausencia, (2) en forma consistente presenta un buen desempeño, en comparación con otros métodos similares especialmente cuando existe un número pequeño de muestras (Elith *et al.* 2006, Phillips *et al.* 2006, Hernández *et al.* 2008).

b) Algoritmo genético basado en reglas: GARP

Peterson *et al.* (2011), manifiestan que es un algoritmo genético que crea un modelo de nicho ecológico para una especie, en el cual se necesitan datos de observaciones de presencia de la especie y variables ambientales. GARP trata, de forma iterativa, de encontrar las correlaciones entre las presencias y puntos no muestreados de la especie con los parámetros ambientales, utilizando una serie de reglas diferentes. Cada regla implementa un método diferente para construir los modelos de predicción de la especie. GARP trata los puntos de ausencia y los no muestreados como conceptos equivalentes. Todos los puntos de no presencia se consideran puntos de ausencia de la especie (Liras 2008). GARP puede errar por el lado de la producción de resultados excesivamente amplios. Este algoritmo presenta una mayor sensibilidad al tamaño de muestra (Wisz *et al.* 2008).

c) Modelo lineal generalizado: GLM

Los datos ecológicos a menudo no se ajustan a los supuestos del modelo lineal, los Modelos Lineales Generalizados (GLMs) permiten modelar respuestas no lineales, donde los residuos pueden seguir distintas distribuciones de probabilidad, por lo que requieren datos de ausencia y presencia, y permiten la configuración de distintos parámetros, como la complejidad de la curva e interacciones entre variables (Peterson *et al.* 2011). Necesita al menos cinco presencias y cinco ausencias por cada variable ambiental, así como presencias adicionales por cada término polinomial. Es un algoritmo que necesita trabajar con una gran cantidad de datos para generar buenos resultados (Benito 2015).

d) Modelos aditivos generalizados (GAMs).

Los modelos aditivos generalizados son un enfoque flexible y automático para describir relaciones no lineales entre predictores y respuestas. Difieren de los GLMs en su capacidad para caracterizar la naturaleza de la función de respuesta (Yee y Mitchell 1991). Los GAMs son extensiones no paramétricas del GLM (Guisan *et al.* 2002). Los GAMs requiere de datos

de ausencia y presencia, así como tamaño de muestras de datos grandes, inclusive mayores a los GLMs (Benito 2015).

2.4. CALIBRACIÓN Y PREDICCIÓN DEL MODELO

Pliscoff y Fuentes-Castillo (2011), señalan que la calibración del modelo hace énfasis a los análisis estadísticos asociados a buscar el ajuste del modelo para posteriormente realizar la predicción en el espacio geográfico, la calibración dependerá del método (algoritmo) que se haya elegido; en esta etapa se seleccionaran los parámetros y configuraciones en el método a utilizar. Además, manifiestan que nos proporciona herramientas muy útiles para analizar los resultados del entrenamiento del modelo, una de las cuales es el de las curvas de respuestas, que son gráficos que muestran cómo se distribuyen las presencias sobre los valores de las variables descriptoras seleccionadas. Rykiel (1996) define la calibración como «la estimación y ajuste de parámetros del modelo y constantes para mejorar el acuerdo entre la salida del modelo y un conjunto de datos». En términos sencillos, la calibración a través de variables independientes (valores de variables ambientales en casos de ocurrencias extraídos para los fines de la calibración) y variables dependientes, las cuales son valores de presencias y ausencias -1 y 0 correspondientes - genera una “fórmula o modelo ajustado” para poder ser aplicado posteriormente.

Una vez que las respuestas de la especie (curva de respuesta o perfil ecológico) se obtiene procedente del calibrado, su distribución potencial en el área a modelar se puede predecir (Guisan y Zimmermann 2000). Es decir, su predicción geográfica es el siguiente paso una vez calibrado (entrenado) el modelo. Se trata de aplicar la “formula” o “modelo ajustado” al espacio geográfico de interés (área de estudio). Como ya se había mencionado en la página 6, al espacio geográfico le corresponde su espacio ecológico caracterizado por las variables ambientales seleccionadas. Dichas variables (en formato raster) actuarán como variables independientes y al aplicar el “modelo ajustado” se obtendrá un mapa raster con valores de 0 a 1 en la mayoría de casos, representando la idoneidad de hábitat, o -en algunos casos (cuando se trabaja con presencia y ausencia)- probabilidad de presencia.

2.5. EVALUACIÓN DEL MODELO

La forma correcta para evaluar el rendimiento del modelo es disponer de un conjunto de datos de presencia independiente del que se ha usado para calibrar el modelo. Esto es difícil de obtener cuando se tienen datos de presencia limitados, y se resuelve realizando una partición

del conjunto de datos en dos sets: uno de calibración y otro de evaluación (Guisan y Zimmermann 2000).

Esta situación conlleva a dos enfoques para evaluar el poder predictivo de un modelo:

a) Evaluación desde un único set de datos

Cuando se tiene un único set de datos para calibrar y evaluar el modelo, se emplean técnicas apropiadas como: *k-fold*, *bootstrap*, *data splitting*, *leave-one-out* (Benito 2015). *Data splitting* consiste en separar un conjunto de datos para calibrar el modelo y otro para evaluar. *K-fold* separa los datos en “n” grupo, se calibra con n-1 y se evalúa con 1 grupo. *Bootstrap* es una partición iterativa de los datos, para calibrar con unos y evaluar con otros. *Leave-one-out* o *Jacknife* para muestras pequeñas: cada registro se elimina una vez del conjunto de datos y el modelo se calibra utilizando los restantes n - 1 registros. Por lo tanto, para especies con n localidades observadas, se construyeron n modelos separados para las pruebas. El desempeño predictivo de cada modelo se evalúa mediante la capacidad de cada modelo para predecir la única localidad excluida del conjunto de datos de entrenamiento (Wisz *et al.* 2008). Cuando el tamaño de muestra es pequeño se recomienda usar *bootstrap* o *leave-one-out*.

Por lo general, se usa el 70 por ciento del conjunto de datos para entrenamiento y el 30 por ciento para datos de prueba (Ferrer 2012).

b) Evaluación desde 2 set de datos independientes

Cuando se utilizan dos conjuntos independientes de datos el primero se utiliza para ajustar el modelo (calibración), mientras que el segundo se utiliza para evaluar el rendimiento del modelo, obteniéndose un solo valor para el estadístico evaluador del rendimiento, por ejemplo, AUC (Guisan y Zimmermann 2000). Cuando la proporción de presencias es mayor en los datos de calibración del modelo (50 por ciento de los datos o más), mejora la caracterización de idoneidad de hábitat (Cramer 2003).

2.5.1. EVALUADORES DE RENDIMIENTO

Existen muchas medidas cuantitativas para evaluar el rendimiento del modelo, se dividen en dos categorías: (1) las diseñadas para su uso con predicciones binarias (mapas binarios resultantes del algoritmo); (2) las diseñadas para predicciones no binarias (mapas continuos como los resultantes de *maxent*). A menudo también son denominadas “dependiente del umbral” e “independiente del umbral” respectivamente (Peterson *et al.* 2011).

a) Evaluación en predicciones binarias

La mayoría de las medidas de rendimiento para modelos binarios se derivan de elementos de una matriz de confusión (Fielding y Bell 1997). La matriz de confusión (Figura 7), contiene filas que resumen predicción de presencia y ausencia, así como columnas que indican el verdadero estado de presencias y ausencias. Las cuatro celdas de la matriz indican así combinaciones distintas de la predicción frente a la realidad (Peterson *et al.* 2011).

		Estado Actual	
		Presencia	Ausencia
Estado Predicho	Presencia	a	b
	Ausencia	c	d

Figura 7: Matriz de confusión

FUENTE: Peterson *et al.* 2011

La matriz (Fig. 7) permite el cálculo de estadísticos para los modelos, como sensibilidad (fracción de presencias predichas correctamente o verdaderos positivos) y especificidad (porcentaje de ausencias predichas correctamente o verdaderos negativos). Así como tasas de comisión (fracción de ausencias predichas como presencia o falsos positivos) y tasas de omisión (fracción de presencias predichas como ausencia o falsos negativos).

Tabla 2: Estadísticos derivados de matriz de confusión

Estadísticos de matriz de confusión	Formula
Sensibilidad	$\frac{a}{a + c}$
Especificidad	$\frac{d}{b + d}$
Tasa de comisión	$\frac{b}{b + d}$
Tasa de omisión	$\frac{c}{c + a}$
Estadístico de Kappa(k)	$\frac{[(a + d) - \frac{(a + c)(a + b) + (b + d)(c + d)}{n}]}{[n - \frac{(a + c)(a + b) + (b + d)(c + d)}{n}]}$

FUENTE: Elaboración propia

Según Peterson *et al.* (2011), el estadístico de Kappa estima la exactitud del modelo, usando todos los valores de la matriz de confusión, sin embargo, resulta inadecuado usarlo cuando el objetivo es obtener distribuciones potenciales (debido a que se calcula ambientes apropiados

aunque la especie esté ausente). Cuando el modelo ha usado solo datos de presencia se puede calcular sensibilidad y tasa de omisión.

Freeman y Moisen (2008) manifiestan que los modelos de nichos ecológicos se desarrollan a menudo para usarse con más de un propósito. Por ejemplo, para diversas aplicaciones en de gestión de tierras. Por lo tanto, es preferible desarrollar mapas de probabilidad continua y permitir a los usuarios potenciales aplicar umbrales para producir mapas binarios, de acuerdo a su criterio.

b) Evaluación en predicciones continuas

Cuando un modelo es un mapa con valores continuos de probabilidades de presencia, la forma más básica es reclasificar el mapa de valores continuos, basados en un valor de corte o umbral determinado, para luego analizar cómo los puntos conocidos de las presencias se distribuyen en este nuevo mapa reclasificado en forma binaria (Liu *et al.* 2005, Jiménez-Valverde y Lobo 2007, Freeman y Moisen 2008). Valores sobre el umbral determinado se estiman como presencia, y valores por debajo como ausencia en el área de estudio (Pliscoff y Fuentes-Castillo 2011). Sin embargo, los estadísticos de la matriz de confusión aplicada a este mapa binario serán sensibles del método que se utilizó para elegir el umbral (Ferrer 2012).

Fielding y Bell (1997), manifiestan que en las predicciones binarias, la evaluación del modelo no tiene en cuenta toda la información proporcionada por el modelo, por ello, es aconsejable la utilización de un estadístico que de una sola medida de evaluación del poder predictivo a través de todo el rango de posibles umbrales. Esto se consigue con el AUC: Área bajo la curva ROC (característica operativa del receptor).

La curva ROC (Fig. 8), se obtiene de enfrentar a “1-especificidad” a lo largo de todos los posibles umbrales. Se utiliza la sensibilidad y especificidad por tener en cuenta estos estadísticos todos los elementos de la matriz de confusión. Se utiliza ”1-especificidad” porque con la sensibilidad son directamente proporcionales para diferentes valores del umbral (Ferrer 2012).

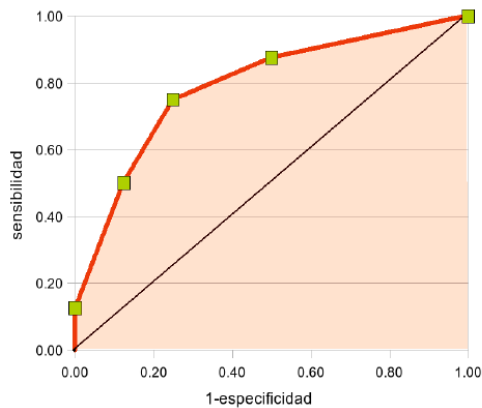


Figura 8: Curva ROC

FUENTE: Benito 2015.

El AUC se ha utilizado ampliamente en la literatura de modelado de nicho ecológico. Este mide la capacidad de un modelo para discriminar entre los sitios donde está presente una especie, en comparación con aquellos en los que está ausente (Hanley y McNeil 1982). El rendimiento de un modelo es mayor cuando la proporción de verdaderos positivos (sensibilidad) es mayor a la de falsos positivos (“1-especificidad”), lo que se expresa gráficamente en una curva (Fig. 8) que mientras mayor sea su precisión, se acercara más a la esquina izquierda del gráfico, generando un “área de curva” mayor. Se estima que valores de AUC mayores a 0,9 dan modelos con alta precisión y valores entre 0,7 y 0,9 indican modelos con buena precisión (Thullier *et al.* 2009), pero se debe tener en cuenta que el valor del AUC depende del número de presencias y del tamaño total del área de estudio (Lobo *et al.* 2007).

A pesar de ser un estadístico de uso casi universal en modelado del nicho ecológico, el AUC tiene serios cuestionamientos (Austin 2007, Lobo *et al.* 2007, Peterson *et al.* 2008). Entre otros:

- En teoría, requiere de datos de ausencias verdaderas. En el caso de que no se cuente con ausencias verdaderas, la abscisa de la ROC no es la proporción de "falsos positivos", sino la proporción de área predicha por el modelo.
- Que presupone que la predicción genera todos los valores del rango 0 a 1, lo cual no siempre es cierto.

Benito (2015), manifiesta que el AUC resultante de datos de evaluación presencia-ausencia debe interpretarse como la «probabilidad de que, seleccionando al azar una presencia y una

ausencia, el modelo clasifique con un valor de idoneidad mayor a la presencia que a la ausencia». El mismo autor manifiesta que para AUC derivados de datos de evaluación presencia-background el AUC será interpretado como «la probabilidad de que un punto de presencia seleccionado al azar tenga un valor de idoneidad más alto que el de un punto aleatorio seleccionado al azar».

A pesar de la controversia que gira en torno del AUC, es una buena herramienta para comparar modelos para la misma especie y área de trabajo (Benito 2015).

El coeficiente de correlación también se ha usado como evaluador de precisión “independiente del umbral”, y es denominado como correlación biserial en el caso en que una de las variables es binario (Elith *et al.* 2006, Elith y Graham 2009). La correlación biserial mide el grado en que las predicciones varían linealmente con los datos de la prueba (presencia-ausencia), la ventaja es que tiene en cuenta la magnitud de diferencia entre predicción y los datos de prueba (Franklin 2009).

Las medidas de rendimiento predictivo descritas, evalúan la capacidad de un modelo para discriminar la presencia de ausencia. Sin embargo, todas estas medidas pueden aplicarse a la presencia-pseudoausencia o presencia-background, aunque deben interpretarse de manera diferente (Franklin 2009).

Un caso particular es la evaluación de muestras pequeñas (menores a 20 presencias), para la cual evaluar el rendimiento es difícil con AUC, ya que se debería utilizar el mayor número de muestras en la calibración del modelo. Pearson (2007), propuso una *prueba de significancia p-value* para muestras pequeñas, dividiendo cada muestra a través de leave-one-out, evaluando si cada punto dejado fuera de la data de entrenamiento era capaz de ser predicho por el modelo. Así, se denominaría éxito si el modelo lo predice (asignándole valor 1) y fracaso si no lo hace (asignándole valor 0). El *p-value* de Pearson, utiliza estos valores de fracaso y éxito con la proporción de área predicha para cada modelo del área total de estudio. El resultado es el *p-value* interpretado de acuerdo al nivel de significancia elegido. De ser significativamente estadístico se interpreta que los modelos de predicción para las observaciones de presencia no utilizado son mejores que un resultado aleatorio.

2.6. UMBRALES DE CORTE

Freeman y Moisen (2008), manifiestan que los modelos de nichos ecológicos se desarrollan a menudo para usarse para más de un propósito, por ejemplo, para diversas aplicaciones en

gestión de tierras. Por lo que es preferible desarrollar mapas de probabilidad continua y permitir a los usuarios potenciales aplicar umbrales para producir mapas binarios, de acuerdo a su criterio.

El umbral se selecciona para maximizar la concordancia entre la distribución observada y la modelada (Liu *et al.* 2005).

Según Benito (2015), se debe tener en cuenta criterios importantes antes de aplicar un umbral: (1) la elección del umbral depende del objetivo del modelo, no hay una norma fija; (2) los modelos de especies con pocas presencias son muy sensibles a la elección del umbral, (3) no hay obligación de aplicar un umbral a un modelo salvo que el análisis concreto lo requiera, (4) siempre es mejor usar la versión continua de un modelo.

Existe muchos enfoques para la determinación de umbrales, que se dividen en dos categorías: subjetivo y objetivo.

2.6.1. DETERMINACIÓN SUBJETIVA DE UMBRALES.

Entre los principales se encuentran:

a) Valor fijado

Se determina un valor arbitrario. Un representante de la primera categoría está tomando 0,5 como el umbral, que se utiliza ampliamente en la ecología (Manel *et al.* 2001, Luck 2002, Stockwell y Peterson 2002). A veces 0,3 y 0,5 también se utilizan como umbrales. Estas opciones son muy arbitrarias y carecen de cualquier base ecológica (Osborne *et al.* 2001).

b) Valor predicho más bajo

El valor predicho más bajo en base a las ocurrencias observadas, no se elige directamente el umbral, pero el investigador aplica su criterio al determinar que sea el más bajo observado para un registro de presencia (Ferrer 2012).

c) Sensibilidad fijada

El umbral en el cual se alcanza una sensibilidad fijada arbitrariamente. Por ejemplo 95 por ciento, de la sensibilidad (Cantor *et al.* 1999). Por lo tanto, el umbral correspondiente se puede encontrar.

2.6.2. DETERMINACIÓN DE UMBRALES.

Entre los principales se encuentran:

- a) Maximización de capas

El umbral en el que estadístico de kappa es máximo (Ferrer 2012).

- b) Maximización de éxito global de predicción

Umbral en el que el estadístico de éxito global de predicción se maximiza (Liu *et al.* 2005)

- c) Enfoque de prevalencia

El umbral se determina tomando la prevalencia de los datos de construcción de modelos (Cramer 2003).

- d) Probabilidad media de idoneidad

El umbral se determina a partir del valor de idoneidad medio de los datos de construcción del modelo (Cramer 2003).

- e) Punto medio de probabilidad

Punto medio entre las probabilidades medias de idoneidades de la especie para los sitios ocupados y no ocupados (Fielding y Haworth 1995).

- f) Maximización de suma sensibilidad-especificidad

Umbral en el que la suma de sensibilidad y especificidad se maximiza (Cantor *et al.* 1999).

- g) Igualdad de sensibilidad-especificidad

El umbral en el que la sensibilidad y especificidad se igualan (Ferrer 2012).

- h) Enfoque basado en el gráfico ROC

El umbral corresponde al punto en la curva ROC (sensibilidad frente a 1- valor de especificidad), que tiene la distancia más corta a la esquina superior izquierda en la línea ROC (Cantor *et al.* 1999).

Liu *et al.* (2005) realizaron el mayor estudio comparativo de métodos para elección de umbrales (12 métodos), trabajando con dos especies en Europa y redes neuronales artificiales. Como resultado obtuvieron que el método de maximización de kappa y valor fijado fueron

los peores. Sin embargo, el enfoque de prevalencia, probabilidad media de idoneidad, maximización de suma sensibilidad-especificidad, igualdad de sensibilidad-especificidad y el enfoque basado en el gráfico ROC dieron buenos resultados al evaluarlos con distintos estadísticos.

3. SOFWARES LIBRES

3.1. RSTUDIO Y R

Hijmans (2016) manifiesta que Rstudio es un entorno de desarrollo e integrado para R (lenguaje de programación). Incluye una consola, editor de sintaxis y herramientas para la gestión de datos. Rstudio posee una interface gráfica más amigable que la consola de R, por lo cual es utilizado para análisis de datos de diversa índole, así como su integración al proceso de datos de sistemas de información geográfica y percepción remota. Actualmente Rstudio y R es utilizado en el modelamiento de nichos ecológicos, ya que, cuenta con la posibilidad de integrar complementos en forma de “paquetes” que permiten realizar el modelado. Algunos de los paquetes más destacados son “biomod”, “biomod2”, “dismo”, “evaluate”, “raster”, “rgdal”, “sp”, “ENMeval” los cuales permiten llevar a cabo todo el modelado incluyendo evaluación de rendimiento por diferentes estadísticos y aplicación de umbrales.

3.2. QGIS

Según Jung (2014), para el análisis de datos geográficos, generación de mapas y trabajar con herramientas estadísticas a estos datos, se hace uso de herramientas de sistemas de información geográfica (GIS). QGIS es un software libre de GIS, creado por Gary Sherman en el 2002 y que actualmente está integrado por una comunidad de desarrolladores de diversos países.

Permite el manejo y análisis de datos geográficos y se caracteriza por su amplia interoperabilidad con otros softwares libres y sus “plugins” o complementos generados por una comunidad de usuarios. Actualmente QGIS cuenta con el complemento QSDM (QGIS Species Distribution Model), el cual permite modelar nichos ecológicos y distribución de especies a través del algoritmo *maxent*, sin embargo, en un futuro el complemento espera incorporar algoritmos como *Random Forest* y regresiones logísticas para el modelado (Jung, 2014).

4. ESPECIES FORESTALES DE INTERÉS

4.1. ESPECIES FORESTALES AMENAZAS

4.1.1. *Polylepis racemosa* Ruiz & Pav.

En los Andes Peruanos se registra un total de 19 especies del género *Polylepis*, que representa más del 70 por ciento de las 27 especies registradas en todos los Andes. Los bosques de *Polylepis* son recursos vitales para la conservación de la biodiversidad y funciones hidrológicas, la cual se verá alterada por el cambio climático a nivel mundial desafiando la sostenibilidad de las comunidades locales. Estos ecosistemas andinos de gran altitud son cada vez más vulnerables por la presión antropogénica que se viene dando, generando fragmentación y deforestación. La importancia para predecir la distribución de bosques nativos ha aumentado para contrarrestar los efectos negativos del cambio climático a través de la conservación y la reforestación (Zutta *et al.* 2012).

Polylepis racemosa es endémica de Perú y Bolivia, introducida en Ecuador. Son arboles de 4 a 15 metros de alturas, su corteza externa es ritidoma membranáceo o papiráceo, de color rojizo a marrón amarillento que se desprende de forma continua en capas delgadas (León 1988, Reynel 1988). Sus tallos y ramas son encorvados y torcidos, posee hojas compuestas imparipinnadas (Lao *et al.* 1990).

Actualmente está considerada como especie amenazada en estado vulnerable (IUCN 2016.). Según Brako y Zarucchi (1993) *Polylepis racemosa* ha sido reportada en el Perú con rango altitudinal de 1800 – 5000 m.s.n.m (ecorregiones de la serranía esteparia y puna) en diez departamentos del Perú: Ancash, Apurímac, Ayacucho, Cajamarca, Cusco, Huancavelica, Junín, Lima, La libertad, Pasco. También ha sido registrada en Puno (Simpson 1979). Específicamente se distribuye en los bosques interandinos del norte al centro del Perú (Lao *et al.* 1990).

Por lo general vive en laderas poco expuestas, con presencia de neblina. En el caso de bosques residuales lo usual es que sean masas puras y ocupen suelos con alto contenido de materia orgánica formada de los mismos árboles. Es capaz de continuar su función biológica a temperatura tan baja como 4°C en el suelo. Su rango de distribución es amplio desde 250 a 2000 mm anuales lo que hace a la especie resistente a la sequía. Posee rusticidad para desarrollarse en suelos superficiales hasta en suelos profundos de valles y quebradas. Prefiere suelos ácidos de textura media (Lao *et al.* 1990).

Reynel y León (1990) manifiestan que tolera bien diversos tipos de suelos, aunque crece con más rapidez en aquellos con texturas francas o franco-arcillosas; crece bien en suelos pocos profundos con alto porcentaje de pedregosidad. Requiere niveles de humedad constantes, entre medios a altos. *Polylepis racemosa* es posiblemente la especie del género con más rápido ritmo de crecimiento en la Sierra Central y Sur del Perú.

4.1.2. *Buddleja incana* Ruiz & Pav.

Según IUCN (2016), *Buddleja incana* se encuentra entre las especies forestales andinas amenazadas del país.

Según Arica (2003), el “Quishuar” es una especie de alto potencial, orientada a estabilizar taludes, riberas de quebradas y zonas de deslizamiento. Generalmente esta especie, establece pequeños rodales y sistemas combinados de vegetación como agroforestería. Protege al suelo del proceso de erosión eólica, disminuyendo la velocidad del aire en zonas deforestadas, manteniendo la estabilidad geomorfológica del suelo e incrementando los niveles de retención de humedad.

Arica (2003), describe a la especie como un árbol de porte pequeño hasta mediano, de 3 a 12 metros de altura, el tronco a menudo es recto y robusto. Su corteza externa es agrietada y de color marrón cenizo, sus hojas son subcoriáceas a coriáceas, simples, opuestas y alargadas.

Buddleja incana se encuentra en Bolivia, Colombia, Ecuador y Perú. En el Perú se le encuentra en los departamentos de Amazonas, Ancash, Cajamarca, Cusco, Huánuco, La Libertad, Lima, Junín, Pasco y Puno. El rango de distribución va de 1400 a 4200 m.s.n.m., (pertenecientes a eco regiones de la Serranía esteparia y la Puna), en formaciones de bosque seco a bosque subhúmedo (Reynel y Marcelo 2010).

Costales (2005) señala que esta especie se desarrolla óptimamente entre los 2300 a 3400 m.s.n.m., con requerimientos de temperatura máxima de 18°C y una temperatura mínima de 2°C. Esta especie prefiere los suelos ligeramente alcalinos a neutros y con texturas francas a franco-arenosas; sin embargo, es una especie que se adapta con facilidad a otras texturas. Tolerancia a la alta pedregosidad; sus requerimientos de humedad son moderados. Requiere poca agua para su desarrollo, soportar bien las heladas o largos períodos de bajas temperaturas.

4.2. ESPECIES ENDÉMICAS

4.2.1. *Cedrela kuelapensis* T.D.Penn. & Daza

Pennington y Muellner (2010) describen a la especie como un «árbol de 10 m de altura y 30 cm de diámetro aproximadamente. Corteza externa marrón pálido, lenticelada, corteza interna de color rosa-verdoso, con olor característico. Rama terminal de sección circular, lenticelada. Hojas compuestas, pinnadas, 25-35 cm de longitud, folíolos opuestos y subopuestos, 6-7 pares de pinas, ovadas y ampliamente lanceoladas, ápice estrechamente acuminado, base asimétrica, borde entero, venación eucamptódroma, con 9-12 pares de nervios secundarios. Inflorescencia en panículas, terminales, multifloras. Flores unisexuales (monoica), cáliz en forma de copa o cónico, irregularmente dentado, con 5 pétalos oblongos o elípticos, de 7- 8,5 cm de ancho y 1,5-2 mm de longitud, androceo con 5 estambres, libres, ovario de 1-2 mm de longitud. Frutos tipo cápsula, péndulas, ampliamente elipsoidal, con ápice redondo y base aguda, lenticeladas, valvas leñosas aladas. Ramas y hojas con olor característico parecido al ajo».

Pennington y Muellner (2010), manifiestan que es una especie endémica de los Bosques Secos del Valle Marañón, registrada en los Departamentos de Amazonas, Cajamarca y La Libertad; de uso maderable y con una distribución altitudinal desde 600 hasta 2500 m.s.n.m.

4.2.2. *Tecoma rosaefolia* kunth

Marcelo-Peña *et al.* (2010) presentan una breve revisión sobre la especie «*Tecoma rosaefolia* es un arbusto de 1-3 m de alto. Ramitas terminales lenticeladas, pubérulas. Hojas imparipinadas, opuestas, decusadas, con 2-4 pares de folíolos, las láminas de 3-4 cm de largo, 1,5-1,8 cm de ancho, elípticas o elíptico-ovadas, con ápice obtuso a levemente acuminado, base cuneada a redonda, borde dentado; venación eucamptódroma, con 4-7 pares de nervios secundarios, las hojas pubescentes. Inflorescencias en racimos terminales. Flores hermafroditas, anaranjadas o rojizas; cáliz con 5 dientes diminutos; corola estrechamente tubular-campanulada, de 3-5 cm de largo, con 5 lóbulos; estambres 4, dos más largos. Fruto silicua, con cáliz persistente, de 8-15 cm de largo, glabro o con algunas escamas inconspicuas». Además, menciona que es una especie endémica del norte del Perú, con poblaciones distribuidas en los departamentos de Amazonas, Cajamarca, La Libertad y Piura. Con distribución altitudinal desde 300 hasta 2800 m.s.n.m.

Marcelo-Peña *et al.* (2010), sobre su estado de conservación, sugiere considerarla como vulnerable, por hallarse en un área pequeña sujeta a una fuerte deforestación. Destaca su uso

potencialmente ornamental y utilizable para la estación de taludes. Asimismo, manifiesta que la especie es «poco frecuente; observable en terrenos ondulados o sobre taludes de carreteras; prefiere suelos franco arenosos a pedregosos. Asociado con *Mimosa pectinatipinna*, *Croton thurifer*, *Cyathostegia mathewsii*».

4.2.3. *Esenbeckia cornuta* Engl.

Marcelo-Peña *et al.* (2010) presenta una breve revisión sobre la especie «Árbol de hasta 8 m de alto. Corteza externa lisa o lenticelada, gris claro. Ramitas terminales cilíndricas, con pubescencia ceniza. Hojas simples, alternas esparcidas, de 10-12 cm de largo y unos 7 cm de ancho, las láminas oblongo-elípticas, con ápice cortamente acuminado, base obtusa, borde entero, venación eucamptódroma, 8-12 pares de nervios secundarios, glabras por el haz, glabrescentes por el envés. Inflorescencias en cimas multifloras, terminales o axilares. Flores hermafroditas, blancas; cáliz con 5 sépalos cortos; corola con 5 pétalos ovados a ovado-elípticos; estambres 5, alternos con los pétalos; disco anular más largo que el ovario. Fruto una cápsula con 5 carpelos leñosos, 2-3 cm de largo, 1.5-2 cm de ancho, dorsalmente en forma de cuerno, de hasta 1,5 cm de largo».

Marcelo-Peña *et al.* (2010) manifiesta que «*Esenbeckia cornuta* es una especie apreciada por su madera de uso estructural, endémica de los Bosques Tropicales estacionalmente secos. Ha sido observada en el distrito de Jaen, en el sector de sharango. Con distribución altitudinal de 500 a 700 m y crece en suelos franco arenosos hasta pedregosos, asociada a *Croton thurifer*, *Tretrasida chachapoyensis*, *Cordia iguaguana*, *Guapira* sp. Solamente en el sector de Shanango se pueden encontrar poblaciones densas de esta especie, de hasta 760 individuos por ha».

Marcelo-Peña (2008) sugiere categorizar a esta especie en peligro crítico, por su rango geográfico restringido, la severa fragmentación de su hábitat, y la eminente desaparición de sus poblaciones por la demanda de su madera.

III. MATERIALES Y MÉTODOS

1. ÁREA DE ESTUDIO

La extensión del área donde se desarrolló el modelado de distribución potencial de las especies, abarca todo el territorio continental del Perú. La extensión del área de estudio en coordenadas geográficas es de: Latitud 0°2'13''S hasta 18°21'2''S y longitud 68°39'9''O hasta 81°21'46''O. Las cuales han sido obtenidos de un archivo shapefile del Geoportal de Servicio Nacional de Áreas Naturales protegidas (<http://geo.sernanp.gob.pe>).

2. EQUIPOS Y MATERIALES

- Computadora portátil Lenovo, de características Intel i7 y memoria RAM de 12GB.
- Softwares: RStudio (www.rstudio.com), y R (<https://www.r-project.org>) Qgis (www.qgis.org).
- Registros de especies del Herbario MOL UNALM.
- Base digital Herbario Austral Americano, Museo de Historia Natural – Chicago, Jardín Botánico de Missouri, Jardín Botánico de New York.
- Base de datos digital climática WORLDCLIM.
- Base de datos digital topográfica del Centro de ciencia y observación de recursos de la tierra (EROS).
- Shapefile de Límites sugeridos para los Bosques Tropicales estacionalmente secos (Marcelo-Peña *et al.* 2016)
- Shapefile de Fisiografía del Perú (ONERN)
- Modelo de Elevación Digital SRTM a 90 m de resolución (EROS).

3. ESPECIES DE INTERÉS

Se eligió especies de principal interés: dos especies andinas amenazadas por encontrarse en riesgo sus poblaciones y tres especies raras de bosques secos interandinos, estas últimas con gran necesidad de ser documentadas por la ciencia.

Tabla 3: Especies seleccionadas de interés

Nombre científico	Descripción	Fuente
<i>Polylepis racemosa</i> Ruiz & Pav	Especie amenazada	IUCN (2016)
<i>Buddleja incana</i> Ruiz & Pav	Especie amenazada	IUCN (2016)
<i>Cedrela kuelapensis</i> T.D.Penn. & Daza	Especie endémica rara	Pennington y Muellner (2010).
<i>Tecoma rosaefolia</i> Kunth	Especie endémica rara	Marcelo-Peña <i>et al.</i> (2010).
<i>Esenbeckia cornuta</i> Engl.	Especie endémica rara	Marcelo-Peña (2008).

4. COLECCIÓN DE DATOS DE OCURRENCIAS Y VARIABLES AMBIENTALES

De acuerdo a su disponibilidad se obtuvo coordenadas para las especies de interés del Museo de Historia Natural – Chicago, Jardín Botánico Missouri, Jardín Botánico de New York, Herbario Austral Americano y Herbario MOL Forestal – UNALM. La tabla 4 detalla la ocurrencia de registros y sus fuentes.

Tabla 4: Fuente de Registros para las Especies de Interés

Nombre científico	Fuente de registros
<i>Polylepis racemosa</i> Ruiz & Pav	Museo de Historia Natural – Chicago, Jardín Botánico Missouri, Jardín Botánico de New York, Herbario Austral Americano y Herbario MOL Forestal – UNALM
<i>Buddleja incana</i> Ruiz & Pav	Museo de Historia Natural – Chicago, Jardín Botánico Missouri, Jardín Botánico de New York, Herbario Austral Americano y Herbario MOL Forestal – UNALM
<i>Cedrela kuelapensis</i> T.D.Penn. & Daza	Herbario MOL Forestal – UNALM
<i>Tecoma rosaefolia</i> Kunth	Herbario MOL Forestal – UNALM
<i>Esenbeckia cornuta</i> Engl.	Herbario MOL Forestal – UNALM

Las variables ambientales obtenidas fueron procedentes de la base WorldClim 2 (<http://worldclim.org/version2>), 19 variables bioclimáticas y la variable de radiación solar multianual, a resolución espacial de 1km. Adicionalmente se obtuvieron las variables altitud y pendiente de la base del Centro de ciencia y observación de recursos de la tierra (EROS, <https://eros.usgs.gov>).

Las variables obtenidas fueron:

- a) Bio1: Temperatura media anual en °C
- b) Bio2: Promedio del rango diario (Promedio mensual (T° máx. - T° min.)) en °C
- c) Bio3: Isotermalidad (Bio1/Bio7) * 100
- d) Bio4: Temperatura estacional (Desviación estándar *100) en °C
- e) Bio5: Temperatura máxima del mes más caliente en °C
- f) Bio6: Temperatura mínima del mes más frío en °C
- g) Bio7: Rango anual de temperatura (Bio5-Bio6) en °C
- h) Bio8: Temperatura media en el trimestre más lluvioso en °C
- i) Bio9: Temperatura promedio en el trimestre más seco en °C
- j) Bio10: Temperatura promedio en el trimestre más caluroso en °C
- k) Bio11: Temperatura promedio en el trimestre más frío en °C
- l) Bio12: Precipitación anual en mm
- m) Bio13: Precipitación en el periodo más lluvioso en mm
- n) Bio14: Precipitación en el periodo más seco en mm
- o) Bio15: Estacionalidad de la precipitación (Coeficiente de variación)
- p) Bio16: Precipitación en el trimestre más lluvioso en mm
- q) Bio17: Precipitación en el trimestre más seco en mm
- r) Bio18: Precipitación en el trimestre más caluroso en mm
- s) Bio19: Precipitación en el trimestre más frío en mm
- t) Radiación solar multianual en $\text{kJ m}^{-2} \text{ día}^{-1}$

- u) Altitud en metros
- v) Pendiente en grados

5. METODOLOGÍA

Se presenta el proceso de modelado dividido para especies menores a 20 registros y mayores a este, ya que, conllevan diferentes consideraciones, las cuales son descritas a continuación.

5.1. MODELADO DE ESPECIES CON REGISTROS MAYORES A 20

Los procesos descritos fueron aplicados para el modelamiento de *Buddleja incana* Ruiz & Pav y *Polylepis racemosa* Ruiz & Pav.

5.1.1. FLUJO DE TRABAJO

Para una mejor comprensión de la metodología, se presenta un flujograma (Figura 9) que resume los pasos que se siguieron para cumplir los objetivos. Este flujo se aplicó a cada especie. A continuación, se describen cada uno de ellos.

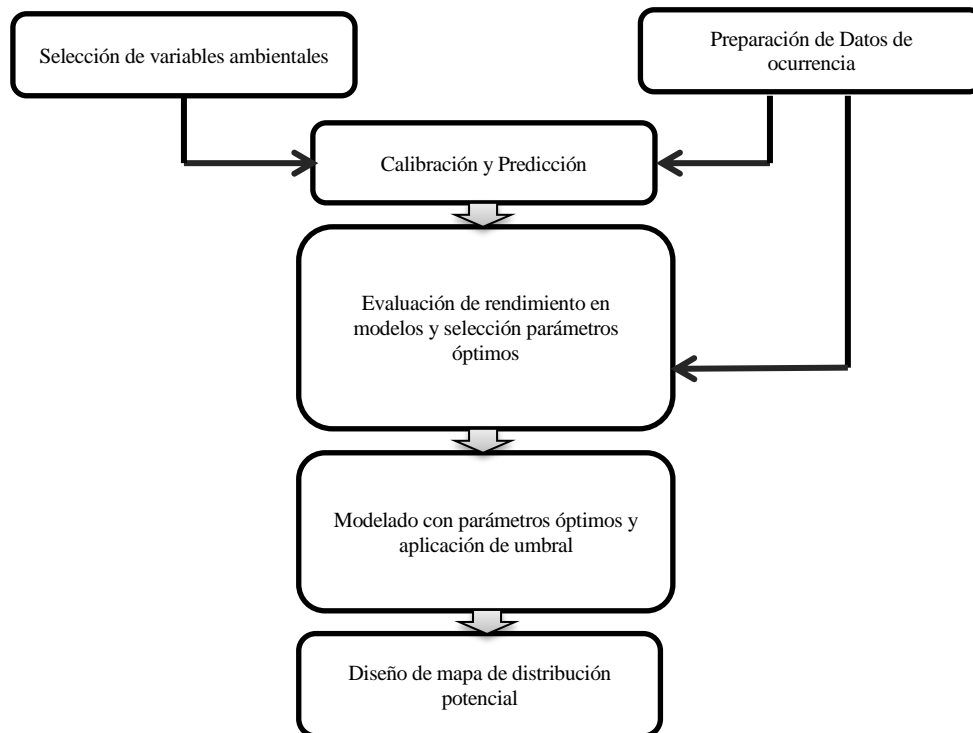


Figura 9: Proceso de modelado de distribución potencial para especies con registros mayores a 20.

5.1.2. SELECCIÓN DE VARIABLES AMBIENTALES

Las 22 variables recolectadas fueron alineadas y enmascaradas en RStudio (paquete “raster”), para que obtengan la misma extensión espacial (1 km) y se enmascaren los valores nulos de igual forma en todos los raster (representación espacial de las variables). Para seleccionar las variables a ingresar en el modelo, se realizó un análisis de correlación múltiple en RStudio (paquete “base”) generando un árbol de correlación, donde aquellos casos de variables con correlación mayor a 0,7 fueron agrupados y se seleccionó de dicho grupo la variable que apporto una mayor interpretación en el modelado. Las variables con correlación menor a 0,7 fueron seleccionadas directamente en el modelo.

Filtradas las variables procedentes del análisis de correlación, se realizó un análisis de factor de inflación de la varianza (VIF) en RStudio, para eliminar aquellas variables que pueden ser resultado de la regresión de otras variables, lo que indica una información redundante. Las Variables con VIF mayor a 5, se analizaron y eliminaron progresivamente hasta lograr un VIF menor a 5 en todas las variables (Benito, 2015). Finalmente se obtuvo las variables seleccionadas a ingresar en el modelo. El proceso se realizó en el software RStudio (interface gráfica de R).



Figura 10: Proceso de elección de variables.

5.1.3. PREPARACIÓN DE DATOS DE OCURRENCIA

Previamente colectada la data para cada especie. Se eliminó aquellos registros duplicados, así como se consideró un registro por celda (perteneciente a variables ambientales finales). Posteriormente se dividió los registros en dos grupos (75 por ciento y 25 por ciento). El 75 por ciento se procedió a dividir en grupos de entrenamiento y evaluación, a través de la técnica de re muestreo k-fold, dividiendo en cuatro grupos los datos (75 por ciento para entrenamiento y 25 por ciento para evaluación), lo que permitió generar combinaciones diferentes de grupo, donde cada grupo tuvo la oportunidad de evaluar los modelos para cada combinación de parámetros de complejidad, mientras los restantes tenían el rol de entrenar el modelo.

El proceso se realizó en el software RStudio (paquete “ENMeval”). El 25 por ciento de registros separados inicialmente se utilizó para evaluar el modelo final.

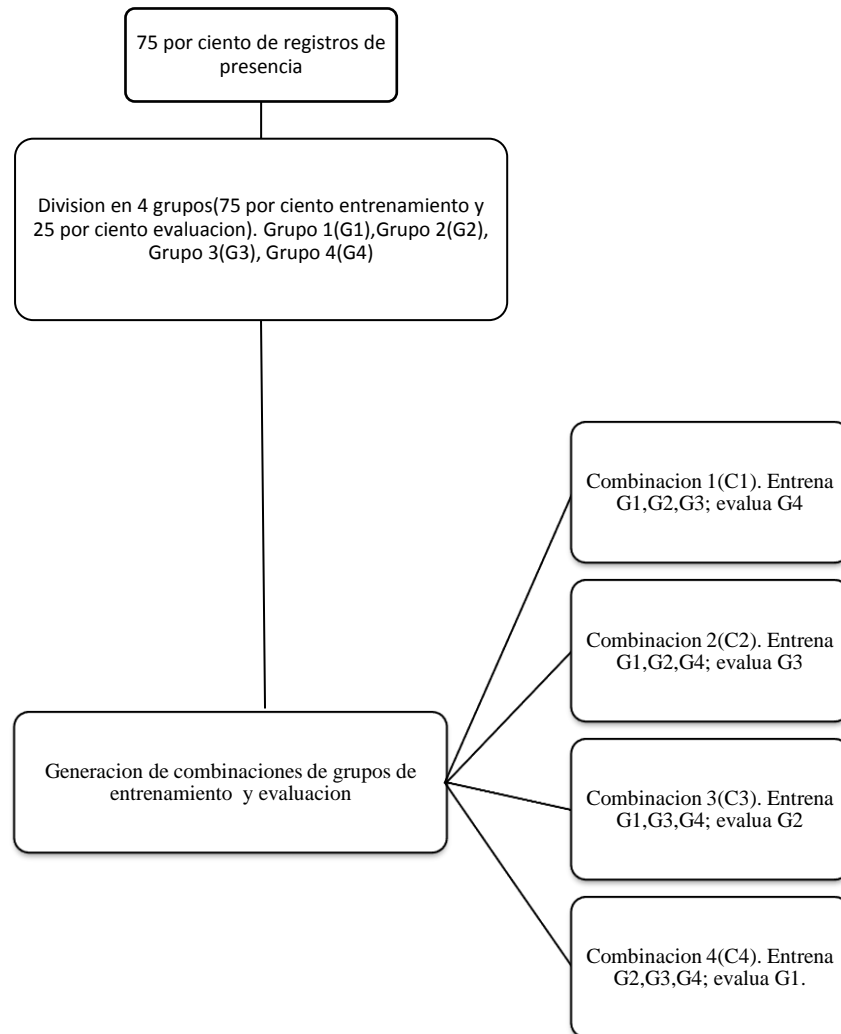


Figura 11: Preparación de grupos de entrenamiento y evaluación para especies con registros mayores a 20.

5.1.4. CALIBRACIÓN Y PREDICCIÓN

Obtenidas las combinaciones de grupos, así como las variables seleccionadas de los pasos precedentes, el siguiente paso fue entrenar el modelo. Para calibrar (entrenar) el modelo se configuro los parámetros de complejidad del modelo (multiplicador de regularización y *features*). Para el multiplicador de regularización se usó cinco valores posibles (0,5; 1; 1,5; 2; 2,5), así como para *features* se generó siete configuraciones (linear, quadratic, *hinge*, *linear-quadratic*, *linear-hinge*, *hinge-features*, *linear-quadratic-features*). Las combinaciones de *features* y multiplicadores de regularización dieron 35 combinaciones posibles de complejidad. Cada una de estas combinaciones de complejidad fue aplicada en el algoritmo *maxent*. Obteniéndose así un total de 35 modelos. A su vez se configuro en el algoritmo *maxent* que todos los modelos con formato de salida “raw” para cada modelo. El proceso se realizó en el software RStudio (paquetes “raster” y “ENMeval”).

5.1.5. EVALUACIÓN DE RENDIMIENTO EN MODELOS Y SELECCIÓN PARÁMETROS ÓPTIMOS

Para los 35 modelos se obtuvo los indicadores promedios de evaluación del rendimiento (AUC) y evaluación de parámetro óptimo de complejidad ($\Delta AICc$). Se seleccionó el modelo con $\Delta AICc = 0$, ya que, indica la distancia entre el mejor modelo y el modelo evaluado en cuestión; así un valor de 0 indica que el modelo que se está evaluando es el mejor. El proceso se realizó en el software RStudio (paquete “ENMeval”).

5.1.6. MODELADO DE PARÁMETROS ÓPTIMOS Y APLICACIÓN DE UMBRAL

Seleccionado el modelo con parámetros óptimos, se procedió a obtener un nuevo modelo con estos parámetros óptimos y con los cuatro grupos de registros en el modelado, para la evaluación de este modelo final (AUC) se utilizó el 25 por ciento de registros separados inicialmente. Se configuró en el algoritmo *maxent*, que realice un análisis de importancia variables según su contribución al modelo (análisis “Jackknife”). El modelo final, presentaba valores continuos de idoneidad para cada pixel, que indica la probabilidad de similitud ambiental a las condiciones ambientales donde se ubican los registros de la especie. Estos se re escalaron de 0 a 1, para mejorar la interpretación de idoneidad de hábitat. Seguidamente, para determinar la zona de distribución potencial se aplicó un umbral basado en el método de “Valor predicho más bajo”, lo que produjo un mapa binario. Donde el valor igual a 1 representa la distribución potencial (valores mayores o iguales al valor predicho más bajo) y valor 0 indica la ausencia de distribución potencial. El proceso se realizó en el software RStudio (paquetes “raster” y “ENMeval”).

5.1.7. DISEÑO DE MAPA DE DISTRIBUCIÓN POTENCIAL

Por último, para la representación cartográfica se vectorizó el mapa binario, solo los valores de 1 que representan la distribución potencial, posteriormente con el archivo vectorial se procedió a realizar una intersección con las ecorregiones de Serranía Esteparia y Puna, adicionalmente se interseccionó con el rango altitudinal según la literatura para cada especie. Para el análisis se utilizó la capa de fisiografía a nivel de zona para obtener aquellas que se encontraba enmarcada en la distribución potencial de la especie. El diseño del mapa con sus respectivos elementos fue realizado en el software QGIS.

5.2. MODELADO DE ESPECIES CON REGISTROS MENORES A 20

Los procesos descritos fueron aplicados a *Cedrela kuelapensis*, *Tecoma rosaefolia*, *Esenbeckia cornuta*.

5.2.1. FLUJO DE TRABAJO.

Se presenta un flujo de trabajo (Fig. 12) que manifiesta un panorama general, cada especie paso por dicho proceso, a diferencia del flujo de trabajo visto anteriormente para las especies con registros menores a 20, se observa que no se realiza una evaluación del rendimiento por tener un registro muy bajo, sin embargo, se permite hacer una prueba de significancia para verificar si los modelos son mejores a los realizados por azar.

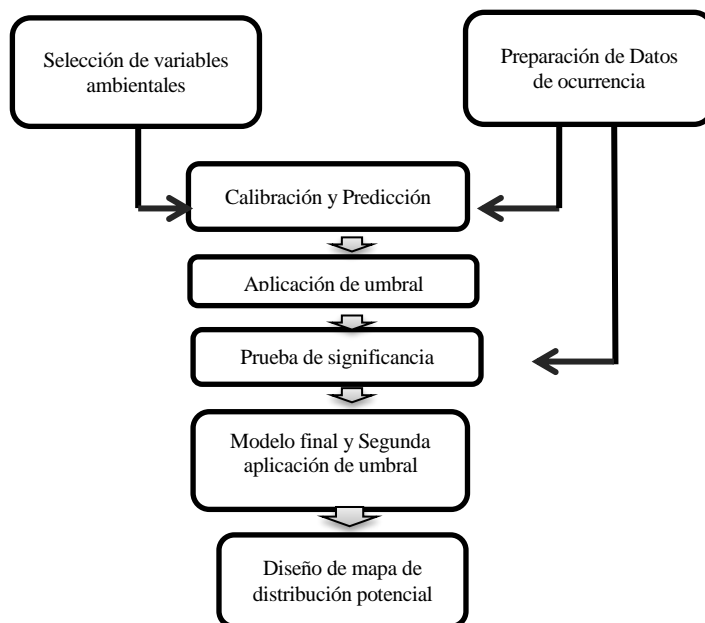


Figura 12: Proceso de modelado de distribución potencial para especies con registros menores a 20

5.2.2. SELECCIÓN DE VARIABLES AMBIENTALES.

Las variables finales para el ingreso del modelo fueron las mismas que en el primer caso, ya que, la extensión del área de estudio (Perú) es igual.

5.2.3. PREPARACIÓN DE DATOS DE OCURRENCIA.

Previamente colectada la data para cada especie. Se eliminó registros duplicados, así como se consideró un registro por celda (perteneciente a variables ambientales finales). Posteriormente se utilizó la técnica de muestreo “leave-one-out”, donde se deja un registro afuera para verificar el éxito de la predicción y los n-1 registros entrenan el modelo. Generando n combinaciones posibles y por lo tanto n modelos para cada especie. El proceso fue realizado en RStudio (paquete “dismo”).

5.2.4. CALIBRACIÓN Y PREDICCIÓN.

Para la configuración en el algoritmo *maxent* se eligió como *features*: *Linear* y *quadratics*. Con un multiplicador de regularización igual a 1, para los n modelos. Dado que registros reducidos no permiten evaluar combinaciones de complejidad, por lo que se optó por un enfoque conservador. A su vez se configuro en el algoritmo *maxent* que todos los modelos tengan el formato de salida “raw”. El proceso fue realizado en RStudio (paquete “dismo”).

5.2.5. RECODIFICACIÓN DE VALORES Y APLICACIÓN DE UMBRALES.

Generado los n modelos para cada especie, se aplicó un umbral basado en el método “valor más bajo predicho” de idoneidad de hábitat en base a los registros de entrenamiento, generando modelos binarios con valor de 1 para las zonas de distribución potencial. El proceso fue realizado en RStudio (paquetes “dismo” y “raster”).

5.2.6. PRUEBA DE SIGNIFICANCIA.

Con el registro que se excluyó para cada modelo, se verifico el éxito o fracaso de la predicción, si coincidía en el área de distribución tomaba valor de 1(éxito) en caso contrario 0(fracaso). Los valores de éxito o fracaso para cada modelo, complementado con su proporción de área predicha fueron ingresados a la prueba de significancia p-value (Pearson, 2007), a un nivel de significancia de 0,05. El proceso fue realizado en el ejecutable p-value Pearson de libre descarga.

5.2.7. MODELO FINAL Y SEGUNDA APLICACIÓN DE UMBRAL.

Evaluated la significancia del conjunto de modelo para cada especie, se procedió a obtener un modelo final configurando en algoritmo *maxent* los *features linear-quadratic*, con parámetro de regularización 1 e incluyendo todos los registros. Al modelo obtenido se reescalo los valores a 0 -1. Se aplicó un umbral basado en el “valor más bajo predicho” para determinar la zona de distribución potencial. El proceso fue realizado en RStudio (paquetes “dismo” y “raster”) se realizó una intersección con la capa digital de Bosques Tropicales Estacionalmente Secos y el rango altitudinal correspondiente a la especie según literatura.

5.2.8. DISEÑO DE MAPA DE DISTRIBUCIÓN POTENCIAL.

Por último, para la representación cartográfica se vectorizo el mapa binario, solo los valores de 1 que representan la distribución potencial, posteriormente con el archivo vectorial se procedió a realizar una se realizó una intersección con la capa digital de Bosques Tropicales Estacionalmente Secos y el rango altitudinal correspondiente a la especie según literatura.

Para el análisis se utilizó la capa de fisiografía a nivel de zona para obtener aquellas que se encontraba enmarcada en la distribución potencial de la especie. El diseño del mapa con sus respectivos elementos fue realizado en el software QGIS.

IV. RESULTADOS Y DISCUSIÓN

1. COLECCIÓN DE OCURRENCIAS

Se obtuvieron 228 registros en total para las cinco especies de estudio: *Buddleja incana*, *Polylepis racemosa*, *Cedrela kuelapensis*, *Tecoma rosaefolia*, *Esenbeckia cornuta*. La Tabla 5 muestra la cantidad de registros según la especie:

Tabla 5: Registros colectados para las especies en estudio

<i>Especie</i>	<i>Nº de registros</i>
<i>Buddleja incana</i>	94
<i>Polylepis racemosa</i>	99
<i>Cedrela kuelapensis</i>	7
<i>Tecoma rosaefolia</i>	17
<i>Esenbeckia cornuta</i>	11

La cantidad de registros observados en la Tabla 5, no son los registros finales que entran al modelado de la especie, dado que, pasan por un proceso de preparación de estos datos (eliminar registros duplicados, seleccionar un registro por celda, etc.), posteriormente se detalla la cantidad de registros que ingresaron al modelado de cada especie.

2. SELECCIÓN DE VARIABLES AMBIENTALES

En la Figura 13, se muestra el valor absoluto de correlación realizado en RStudio, las variables enmarcadas en borde rojo, fueron seleccionadas como representativas de cada agrupamiento. Posteriormente ingresaron al análisis VIF, en el cual no hubo variables con VIF mayor a 5, por lo cual todas las variables del análisis VIF ingresaron al modelado, las cuales fueron: Isotermalidad (Bio3), Rango anual de Temperatura en °C (Bio7), Precipitación en el periodo más lluvioso en mm (Bio13), Precipitación en el periodo más seco en mm (Bio14), Estacionalidad de la precipitación (Bio15), Radiación y pendiente.

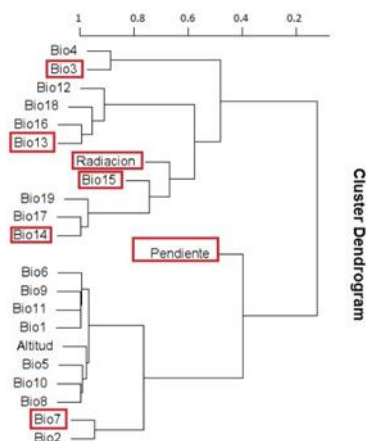


Figura 13: Cluster de asociación entre variables.

3. DISTRIBUCIÓN POTENCIAL DE ESPECIES

3.1. *Buddleja incana* Ruiz & Pav.

De los 94 registros colectados en la fase inicial, quedaron 48, una vez realizada la eliminación de registros duplicados y selección de un registro por celda. Estos 48 registros fueron los que integraron el entrenamiento (75 por ciento) y la evaluación final (25 por ciento). Se obtuvo 35 configuraciones de parámetros de complejidad, las cuales se aplicó a las *Buddleja incana* generando modelos diferentes. En la Figura 14 se visualiza la distribución de los registros tiene mayor concentración en los Departamentos de Cusco y Ancash, pero también abarca Junín, Apurímac, La Libertad y Cajamarca. Acorde a Reynel y Marcelo (2010), *Buddleja incana* ha sido reportada también en Amazonas, Huánuco, Lima, Pasco y Puno sin embargo no se obtuvo registros para estos Departamentos.

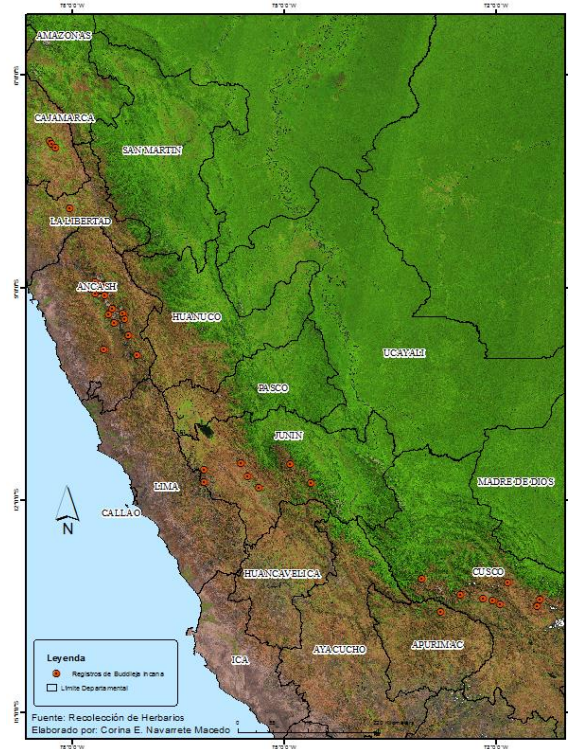


Figura 14: Distribución de registros de *Buddleja incana*.

En la Tabla 6 se muestran los evaluadores de rendimiento que se obtuvo: el AUCtest (AUC promedio con grupos entrenamiento), AUCval (AUC promedio con los grupos de evaluación) y el $\Delta AICc$ (delta del Índice de criterio de Akaike corregido) los 3 “mejores” modelos en función del $\Delta AICc$.

Tabla 6: Evaluadores de rendimiento para *Buddleja incana*

Modelos	Parámetros de configuración	AUCtest	AUCval	$\Delta AICc$
Modelo 24	Features: <i>Linear, quadratic</i> . Multiplicador de regularización: 2,5	0,912	0,909	0
Modelo 7	Features: <i>Linear, quadratic, hinge</i> . Multiplicador de regularización: 2,5	0,921	0,901	2,861
Modelo 6	Features: <i>Linear</i> . Multiplicador de regularización: 2	0,914	0,876	3,146

Según los resultados de la Tabla 6, el mejor modelo es el 24 con un $\Delta AICc=0$ y parámetros *linear-quadratic* con un multiplicador de regularización igual a 2,5. Estos parámetros difieren de los configurados por defecto para registros menores a 80 (*linear-quadratic-hinge* y multiplicador de regularización 1), lo que refuerza lo sustentado por Anderson y González

(2011) que los parámetros configurados por defecto no brindan el resultado más óptimo. El modelo 7 muestra valores cercanos entre el AUC de entrenamiento y evaluación, sin embargo, el modelo 6 empieza a mostrar marcadas diferencias entre dichos AUC lo que implicaría que esta sobre ajustado a la data de entrenamiento, presentando un desempeño inferior en la data de evaluación, esto va acorde a lo mencionado por Silver (2012).

En la Figura 15 se muestra la tendencia en los valores de $\Delta AICc$ para los diferentes valores de multiplicador de regularización según la combinación de features.

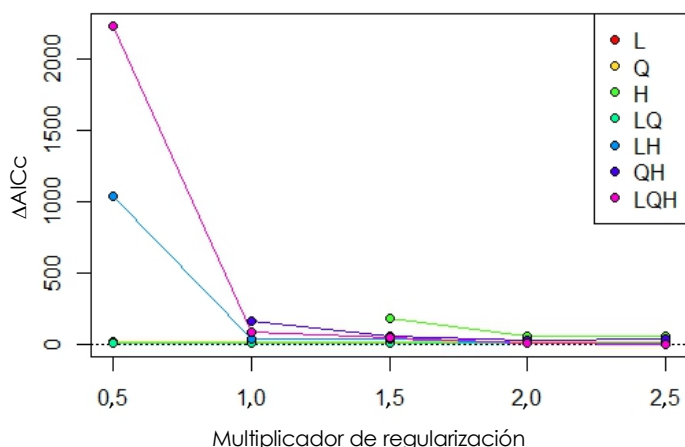


Figura 15: Valores de $\Delta AICc$ versus multiplicadores de regularización en *Buddleja incana*.

Se observa que el valor para $\Delta AICc$ con la combinación de *features linear-quadratic* (LQ) para diferentes multiplicadores de regularización se mantiene bajo, en contraste con la combinación *linear-quadratic-hinge* (LQH) que a mayores multiplicadores de regularización obtiene menores valores de $\Delta AICc$. A su vez se observa que la combinación *Hinge* con multiplicadores de regularización no presenta valores para multiplicadores de regularización 0,5 y 1, esto se debe a que la cantidad de variables transformadas por el *features* supera el número de registros y en ese caso el valor de $\Delta AICc$ es sin datos, dado que, el número de variables transformadas no puede superar el número de registros acorde a CONABIO (2014).

Obtenida la mejor configuración de complejidad (parámetros óptimos), se obtuvo un nuevo modelo con los 36 registros y se evaluó con los 12 registros separados inicialmente. La Tabla 7 muestra los valores de AUC para entrenamiento y validación respectivamente.

Tabla 7: Evaluadores de rendimiento para modelo final de *Buddleja incana*.

Modelo	Parámetros de configuración	AUC entrenamiento	AUC evaluación
Modelo final	Features: Linear, quadratic. Multiplicador de regularización: 2,5	0,913	0,905

Se observa que los valores de AUC no son muy distantes, lo que implica que se ha conseguido un modelo no-sobre ajustado, la cual es la finalidad de los parámetros óptimos. En este caso nos interesa el valor de AUC de evaluación que corresponde a aquellos registros que no uso el modelo para entrenar, con un valor de 0,905 el modelo se considera de alta precisión en concordancia a lo manifestado por Thullier *et al.* (2009).

En la Figura 16, se muestra la importancia de las variables de importancia a través de su contribución en el modelo.

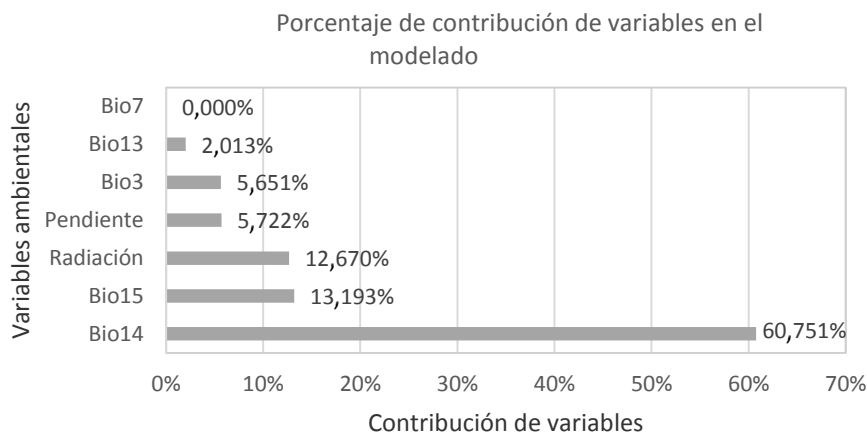


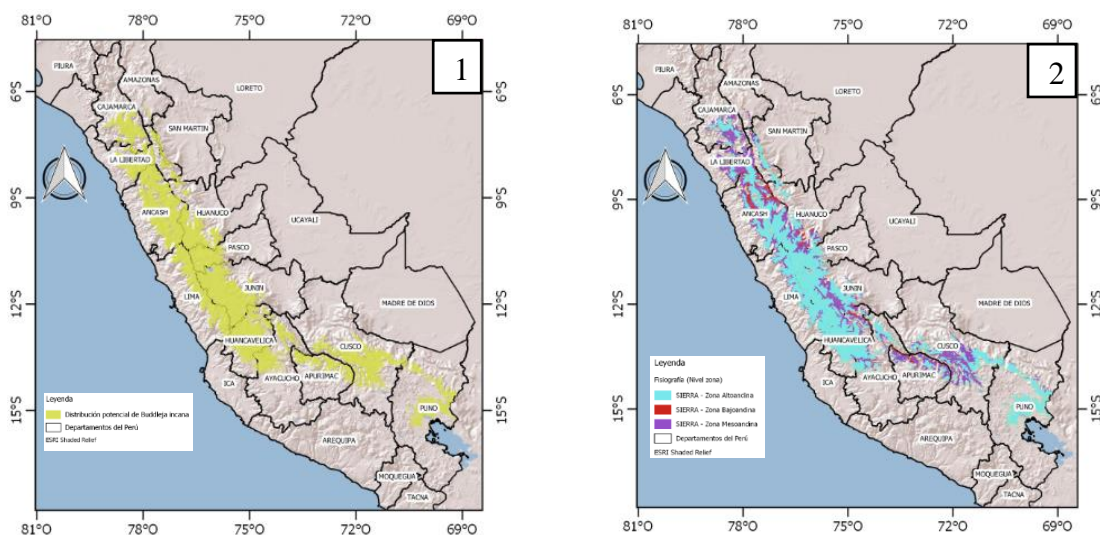
Figura 16: Importancia de variables en el modelo de *Buddleja incana*.

Se observa que en el caso de *Buddleja incana* la variable que obtuvo mayor contribución al modelo fue Precipitación en el período más seco (Bio14), siendo está una característica fundamental señalada por Costales (2005), la tolerancia a escasas precipitaciones por parte de la especie. Además de ser una variable que diferencia muy bien ambientes desérticos de ambientes húmedos. La variable que no aportó al modelo fue el rango anual de temperatura en °C (Bio7). Esto podría deberse a que no se tuvo patrones similares ya que los valores para los registros en esta variable fue de mínimo 9°C y 15°C.

Los valores de idoneidad del modelo para los registros de entrenamiento variaron de 0,4 a 0,8; por lo cual se estableció el umbral a partir del menor valor. Finalmente, a esta distribución potencial establecida por el umbral se intersectó con la capa de ecorregiones (Puna y Serranía

Esteparia) y limitando a rangos de altitud entre 1400 a 4200 según lo mencionado por Reynel y Marcelo (2010).

La Figura 17 muestra la distribución potencial de *Buddleja incana* obtenida (imagen 1) y la fisiografía a nivel de zona enmarcada en esa distribución (imagen 2).



1) Distribución potencial por Departamentos; 2) Distribución potencial en zonas fisiográficas.

Figura 17: Distribución potencial de *Buddleja incana*.

La distribución potencial interpretada está como una idoneidad de hábitat, donde la especie puede estar presente o no, pero alcanza las condiciones idóneas para su establecimiento recorre desde Cajamarca hasta Puno. Se observa en la Figura 17 (imagen 1) como Departamentos como Huánuco, Lima, Pasco y Puno a pesar de no haber tenido registros iniciales, el modelo ha logrado encontrar zonas con condiciones idóneas a partir del grupo de registros que ingresaron en el modelado. En la figura 17 (imagen 2) podemos observar como la Sierra - Zona altoandina es la que alcanza el 81 por ciento del área de distribución potencial, seguido de la Sierra - Zona mesoandina el 14 por ciento y la Sierra - Zona bajoandina el 5 por ciento.

Actualmente las poblaciones naturales, actualmente relictas son muy reducidas e incluso espaciadas a lo largo de quebradas y laderas de escasa continuidad geográfica. Esto se da para la mayoría de especies de árboles de la Región Andina, las cuáles se encuentran vulnerables al cambio climático y sujetas a que las especies puedan conservar su nicho ecológico o sufrir un dinamismo en él.

3.2. *Polylepis racemosa* Ruiz & Pav.

De los 99 registros de la fase inicial, quedaron 50 registros producto del pre procesamiento. El 75 por ciento (38 registros) para el entrenamiento y evaluación de parámetros óptimos del algoritmo *maxent* y el 25 por ciento para evaluación final (12 registros). Se obtuvo 35 configuraciones de parámetros de complejidad, las cuales se aplicó a *Polylepis racemosa*.

En la Figura 18 se visualiza la distribución de los registros tiene mayor concentración en los Departamentos de Cajamarca, Cusco y Junín, pero también abarca Ancash, Ayacucho y Puno. Acorde a Brako y Zarucchi (1993) *Polylepis racemosa* ha sido reportada también en Ayacucho, Lima, Pasco y Puno sin embargo no se obtuvo registros para estos Departamentos.

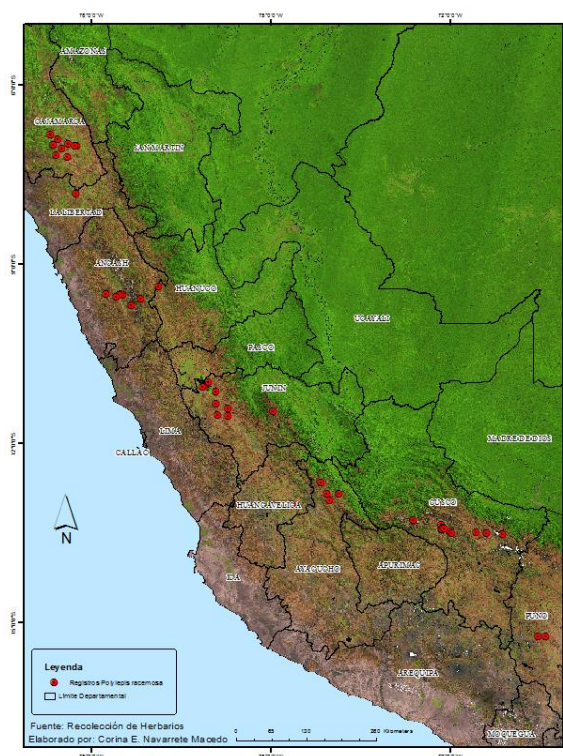


Figura 18: Distribución de Registros de *Polylepis racemosa*.

En la Tabla 8 se muestran los evaluadores de rendimiento que se obtuvo el AUCtest (AUC promedio con grupos entrenamiento), AUCval (AUC promedio con los grupos de evaluación) y el $\Delta AICc$, el orden de los “mejores” modelos se encuentra en función del $\Delta AICc$.

Tabla 8: Evaluadores de rendimiento para *Polylepis racemosa*.

Modelos	Parámetros de configuración	AUCtest	AUCval	$\Delta AICc$
Modelo 33	Features: <i>linear, quadratic</i> . Multiplicador de regularización: 1	0,907	0,898	0
Modelo 4	Features: <i>linear, hinge</i> . Multiplicador de regularización: 2	0,906	0,897	0,002
Modelo 7	Features: <i>quadratic, hinge</i> . Multiplicador de regularización: 2	0,951	0,921	0,869

Según los resultados de la Tabla 8, el mejor modelo es el 33, con un $\Delta AICc=0$ y parámetros *linear-quadratic* con un multiplicador de regularización igual a 1, este resultado vuelve a reforzar que los parámetros configurados por defectos no producen los modelos más óptimos, acorde a Anderson y González (2011). El modelo 4 muestra valores de AUC cercanos entre los datos de entrenamiento y evaluación, así como un $\Delta AICc$ bajo. El modelo 7 muestra valores de AUC de entrenamiento y evaluación muy distantes, por lo que se asume un ligero sobreajuste a los datos de entrenamiento.

En la Figura 19 se muestra la tendencia en los valores de $\Delta AICc$ para los diferentes valores de multiplicador de regularización según la combinación de *features*.

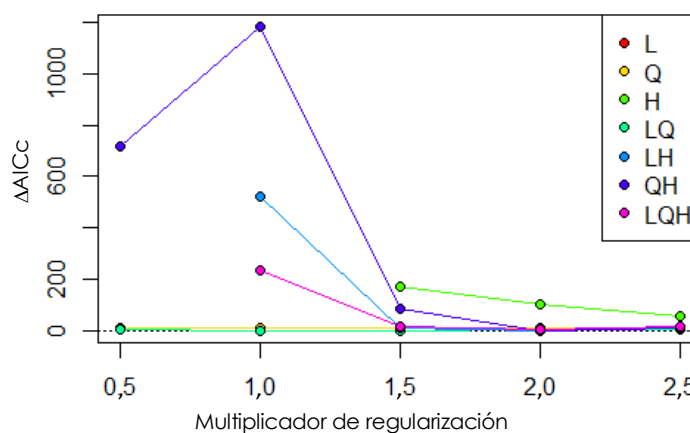


Figura 19: Valores de $\Delta AICc$ versus multiplicadores de regularización en *Polylepis racemosa*.

Se observa que los valores de $\Delta AICc$ para la combinación de *features Linear-quadratic* (LQ) a los diferentes valores de multiplicador de regularización permanece bajo, esto implica que es una combinación idónea de *features* para elegir parámetros óptimos. La Figura 19 nos muestra una particularidad, las combinaciones de *features hinge* con multiplicador de regularización 1 y *linear-quadratic-hinge* (LQH), *quadratic-hinge* (QH) y *hinge* con multiplicador de regularización 0,5 no muestran un valor, esto se debe que al combinarse los *features*, generan un nuevo número de variables transformadas y cuando estas exceden al

número de registro, el $\Delta AICc$ ya no adquiere valor, dado que, el número de variables no puede ser superior al número de registros, esto va acorde a lo sugerido por CONABIO (2014).

Obtenida la mejor configuración de complejidad (parámetros óptimos), se obtuvo un nuevo modelo con los 38 registros y se evaluó con los 12 registros separados inicialmente. La Tabla 9, muestra los valores de AUC para entrenamiento y validación respectivamente.

Tabla 9: Evaluadores de rendimiento para modelo final de *Polylepis racemosa*.

<i>Modelo</i>	<i>Parámetros de configuración</i>	<i>AUC entrenamiento</i>	<i>AUC evaluación</i>
Modelo final	Features: linear, quadratic. Multiplicador de regularización: 1	0,928	0,911

Según los resultados de la Tabla 9, el valor de AUC con la data de evaluación alcanza el 0,911 por lo cual el modelo se considera de alta precisión en concordancia a lo manifestado por Thullier *et al.* (2009).

En la Figura 20, se muestra la importancia de las variables de importancia a través de su contribución en el modelo.

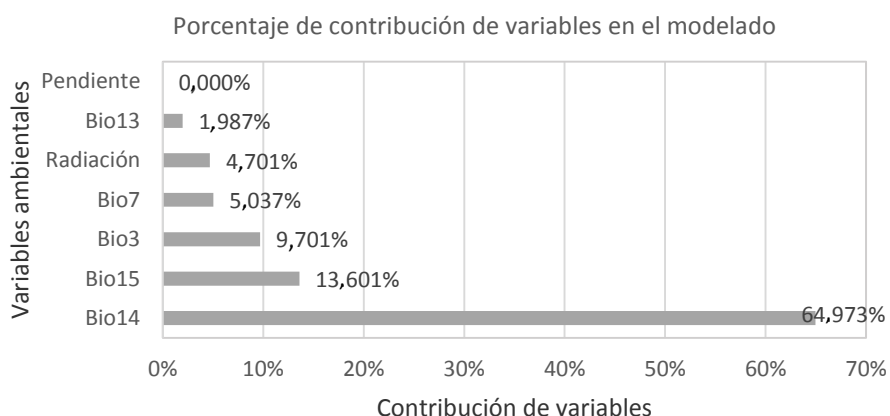


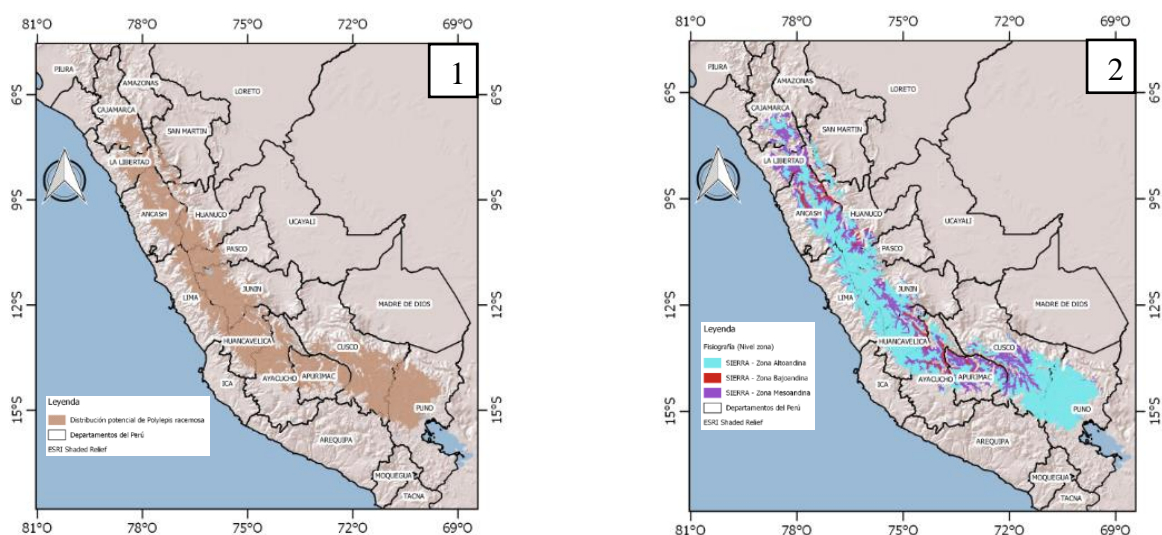
Figura 20: Importancia de variables en el modelo de *Polylepis racemosa*.

Según la Figura 20, la precipitación en el periodo más seco en mm (Bio14) seguido de estacionalidad de la precipitación (Bio15), son las variables de mayor contribución en el modelo. Según Lao *et al.* (1990), la especie posee un amplio rango de distribución lo que la hace resistente a la sequía, lo cual explicaría porque la variable de precipitación en el periodo más seco es la que contribuye en mayor porcentaje al modelo. Además, cabe destacar que la variable pendiente no contribuye en el modelo, como un patrón determinante, esto puede

deberse a que la especie es de rápido crecimiento acorde a Reynel y León (1990) y por su rusticidad puede crecer tanto en valles como en quebrada según Lao *et al.* (1990), por lo cual el modelo no encontró un patrón común en dicha variable.

Los valores de idoneidad del modelo para los registros de entrenamiento variaron de 0,56 a 0,77; por lo cual se estableció el umbral a partir del menor valor. Finalmente, a esta distribución potencial establecida por el umbral se intersecto con la capa de ecorregiones (Puna y Serranía Esteparia) y limitando a rangos de altitud entre 1800 a 5000 según lo mencionado por Brako y Zarucchi (1993).

La Figura 21 muestra la distribución potencial de *Polylepis racemosa* obtenida (imagen 1) y la fisiografía a nivel de zona enmarcada en su distribución (imagen 2).



1) Distribución potencial por Departamentos; 2) Distribución potencial en zonas fisiográficas.

Figura 21: Distribución potencial de *Polylepis racemosa*.

Podemos observar en la Figura 21.1. que la distribución potencial para *Polylepis racemosa* es amplia, desde Cajamarca hasta Puno, Departamentos como Apurímac, Huancavelica, Lima y La Libertad son incluidas como zonas de alta idoneidad, a pesar de no haber contado con registros iniciales en el modelamiento. La distribución potencial para *Polylepis racemosa* se muestra más amplia que la mostrada para *Buddleja incana* en la Figura 17, esto puede deberse a que la amplitud de altitud para *Polylepis racemosa* es mayor a la de *Buddleja incana*. En la figura 21(imagen 2) podemos observar como la Sierra - Zona altoandina es la que alcanza el 87 por ciento del área de distribución potencial, seguido de la Sierra - Zona mesoandina el 9 por ciento y la Sierra -Zona bajoandina el 4 por ciento.

Reynel y León (1990), manifiestan que *Polylepis racemosa* es la especie del género con más rápido crecimiento en Sierra Central y Sur, lo que va acorde con la Figura 21, donde se observa una amplia distribución potencial para estas zonas de Sierra, no cabe duda que la rusticidad de la especie le otorga amplias áreas para su reforestación.

3.3. *Cedrela Kuelapensis* T.D.Penn. & Daza

No hubo registros duplicados, ni registros agrupados en una sola celda por lo que se utilizó la misma cantidad de registros colectados para las especies. Estos 7 registros estuvieron distribuidos en los Departamentos de Amazonas, Cajamarca y La Libertad. La Figura 22, muestra los registros para *Cedrela kuelapensis*.

La configuración de complejidad elegida fueron *features linear-quadratic* y multiplicador de regularización 1, dado que fueron los que mostraron mejor desempeño en el modelado de las especies descritas con anterioridad. El algoritmo *maxent* le otorgo la siguiente abreviación LQ_1. Se obtuvo 7 modelos con la metodología “*leave-one-out*”.

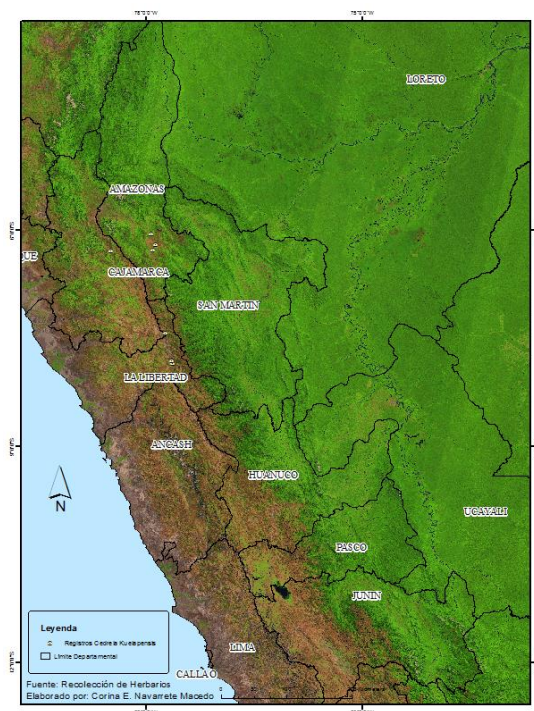


Figura 22: Distribución de Registros de *Cedrela kuelapensis*.

El valor de umbral que se aplicó a cada modelo fue el valor mínimo obtenido por los datos de entrenamiento, la Tabla 10 muestra si el registro excluido fue predicho o no.

Tabla 10: Resultado de predicción de registro excluido en *Cedrela kuelapensis*.

Modelos	Identificador del Registro No incluido	¿Fue predicho? Éxito(1), Fracaso(0)
Modelo 1	1	1
Modelo 2	2	0
Modelo 3	3	0
Modelo 4	4	1
Modelo 5	5	1
Modelo 6	6	1
Modelo 7	7	1

Como se observa 5 de los 7 modelos predijeron el registro excluido.

La Tabla 11 presenta los resultados de la prueba de significancia para la especie, a un nivel de confianza de 0,05:

Tabla 11: Resultado prueba de significancia Pearson para *Cedrela kuelapensis*.

Especie	Nro. registros	Proporción de éxito	p-valor
<i>Cedrela kuelapensis</i>	7	0,71	0,0001

En la Tabla 11 se observa un p-valor de 0,0001; menor al nivel de significancia 0,05; por lo cual se puede deducir que hay evidencia que se puede obtener un modelo estadísticamente significativo, por lo tanto, mejor que modelos al azar (Pearson 2007) con pocos registros. Sin embargo, el número de registros es muy bajo para realizar una evaluación de rendimiento. Esto demuestra que el trabajar con bajos registros implica que se puedan obtener modelos valiosos de gran ayuda, que puedan servir de base como primeras aproximaciones para identificar zonas potenciales, lo cual genere una mayor recolección de muestras y así ir obteniendo distribuciones potenciales más próximas a la realidad.

Se obtuvo un modelo final, con los 7 registros como entrenamiento y los *features linear-quadratic* con multiplicador de regularización 1. En la Figura 23, se muestra la importancia de las variables de importancia a través de su contribución en el modelo. La variable de mayor contribución fue nuevamente Precipitación en el período más seco, siendo esta una limitante en los Bosques Tropicales Estacionalmente Secos (BTES), de donde es endémica *Cedrela Kuelapensis* según Pennington y Muellner (2010).

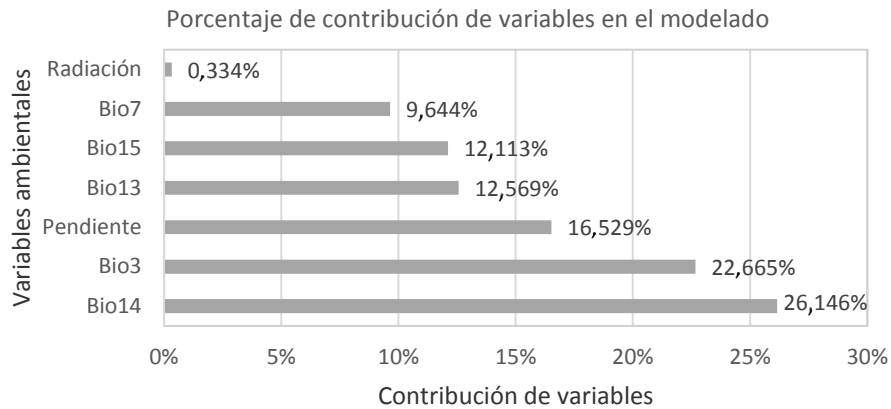
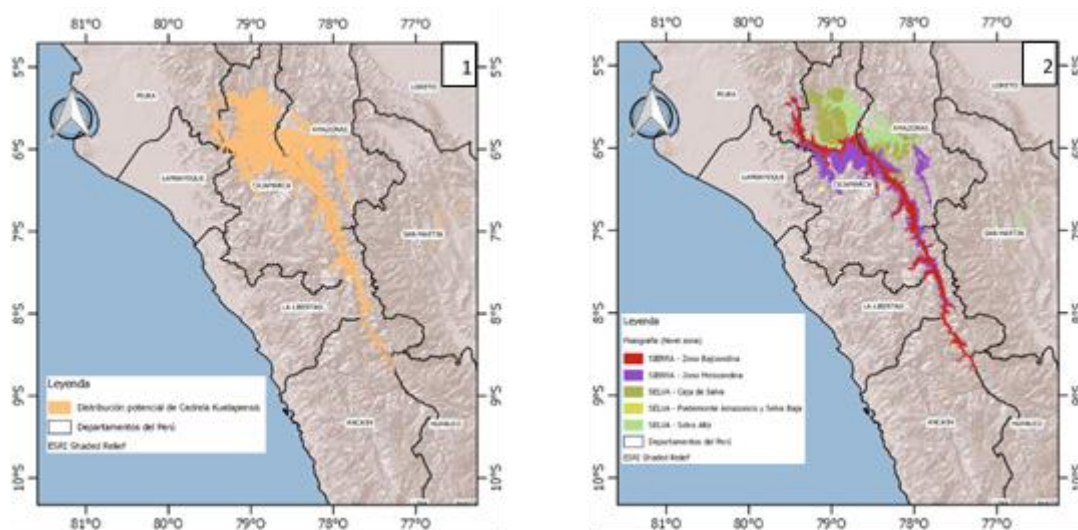


Figura 23: Contribución de variables en el modelo de *Cedrela kuelapensis*

Los valores de idoneidad del modelo para los registros de entrenamiento variaron de 0,68 a 0,82; por lo cual se estableció el umbral a partir del menor valor. Finalmente, a esta distribución potencial establecida por el umbral se interseco con la capa de Bosques Tropicales Estacionalmente Secos propuesta por Marcelo-Peña *et al.* (2015) y limitando a rangos de altitud entre 600 a 2500 según lo mencionado por Pennington y Muellner (2010).

La Figura 24 muestra la distribución potencial de *Cedrela kuelapensis* obtenida (imagen 1) y la fisiografía a nivel de zona enmarcada en su distribución (imagen 2).



1) Distribución potencial por Departamentos; 2) Distribución potencial en zonas fisiográficas.

Figura 24: Distribución potencial de *Cedrela kuelapensis*.

La Figura 24 (imagen 1) Muestra la Distribución potencial obtenida para *Cedrela kuelapensis*, la cual abarca Amazonas, Cajamarca, La Libertad y Ancash, Así como un sector de Piura y San Martín, todas estas zonas pertenecen a los Bosques Tropicales estacionalmente secos, a pesar que se contó con pocos registros, el modelo logro predecir zonas idóneas en otros sectores donde no se contó con registros como Ancash, Piura y San Martín. *Cedrela kuelapensis* es una especie endémica del bosque del Marañón, su idoneidad principalmente está focalizada en esta zona, lo que la hace una especie altamente vulnerable, ya que, se encuentra en zonas sujetas a fuerte deforestación, como lo sugiere Marcelo-Peña *et al.* (2010). En la Figura 24 (imagen 2) podemos observar que esta distribución potencial abarca Sierra – Zona bajoandina (35 por ciento), Sierra-Zona Mesoandina (26 por ciento), Ceja de Selva (21 por ciento), Selva alta (14 por ciento) y Selva-Piedemonte Amazónico y Selva baja (4 por ciento). La amplitud de estas zonas se debe al amplio rango altitudinal de la especie (600 – 2500 m), por lo cual se observa que San Martín presenta zonas idóneas para su restauración, así como Piura y Ancash.

3.4. *Tecoma rosaefolia* Kunth

No hubo registros duplicados, ni registros agrupados en una sola celda por lo que se utilizó los 17 registros colectados para la especie. En la Figura 25, se observa la distribución de estos registros en los Departamentos de Piura, Cajamarca y Amazonas.

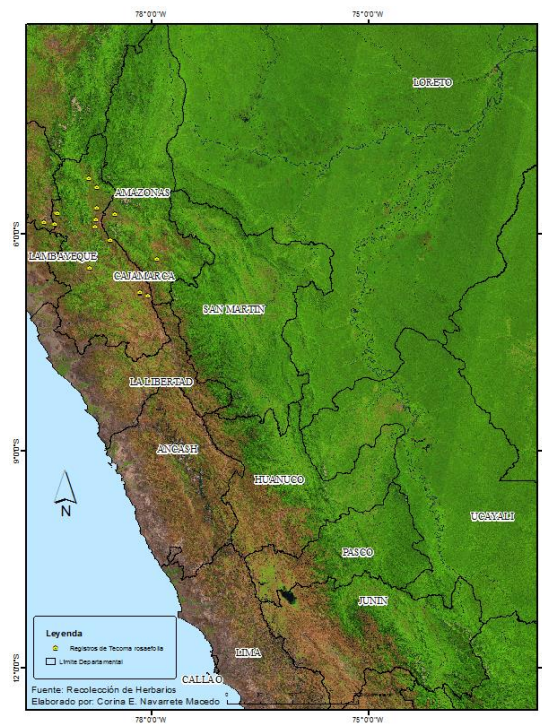


Figura 25: Distribución de Registros de *Tecoma rosaefolia*.

La configuración de complejidad elegida *features linear-quadratic* y multiplicador de regularización 1, el algoritmo *maxent* le otorgo la siguiente abreviación LQ_1. Se obtuvo 17 modelos siguiendo la metodología “*Leave-one-out*”. Seguidamente se aplicó el menor valor de entrenamiento como umbral y se obtuvo si el registro excluido fue predicho o no. La Tabla 12, muestra para cada modelo la condición de éxito o fracaso en la predicción del registro excluido. En ella se observa que 14 de los 17 modelos predijeron el registro excluido.

Tabla 12: Resultado de predicción de registro excluido en *Tecoma rosaefolia*.

<i>Modelos</i>	<i>Identificador del Registro No incluido</i>	<i>¿Fue predicho? Éxito(1), Fracaso(0)</i>
Modelo 1	1	0
Modelo 2	2	1
Modelo 3	3	0
Modelo 4	4	1
Modelo 5	5	1
Modelo 6	6	1
Modelo 7	7	1
Modelo 8	8	1
Modelo 9	9	1
Modelo 10	10	1
Modelo 11	11	1
Modelo 12	12	1
Modelo 13	13	1
Modelo 14	14	1
Modelo 15	15	1
Modelo 16	14	0
Modelo 17	15	0

En la Tabla 13, se muestra los resultados de la prueba de significancia para la especie, a un nivel de confianza de 0,05:

Tabla 13: Resultado prueba de significancia Pearson para *Tecoma rosaefolia*.

<i>Especie</i>	<i>Nro. registros</i>	<i>Proporción de éxito</i>	<i>p-valor</i>
<i>Tecoma rosaefolia</i>	17	0,82	0,0015

En la Tabla 13 se observa un p-valor de 0,0015; menor al nivel de significancia 0,05; por lo cual hay evidencia que se puede obtener un modelo estadísticamente significativo, por lo tanto, mejor que modelos al azar (Pearson 2007) con registros menores a 20. Sin embargo, el número de registros es muy bajo para realizar una evaluación de rendimiento.

Se obtuvo un modelo final, con los 17 registros como entrenamiento y los *features linear-quadratic* con multiplicador de regularización 1. En la Figura 26, se muestra la importancia de las variables de importancia a través de su contribución en el modelo. La variable de mayor contribución fue Isotermalidad (Bio3).

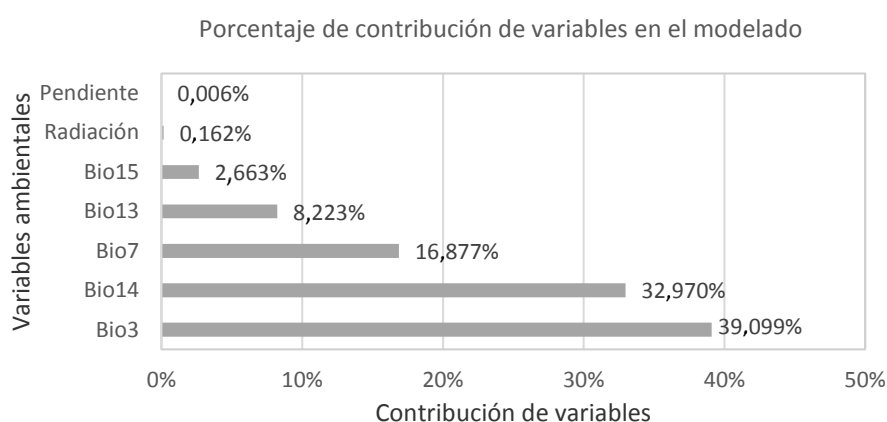


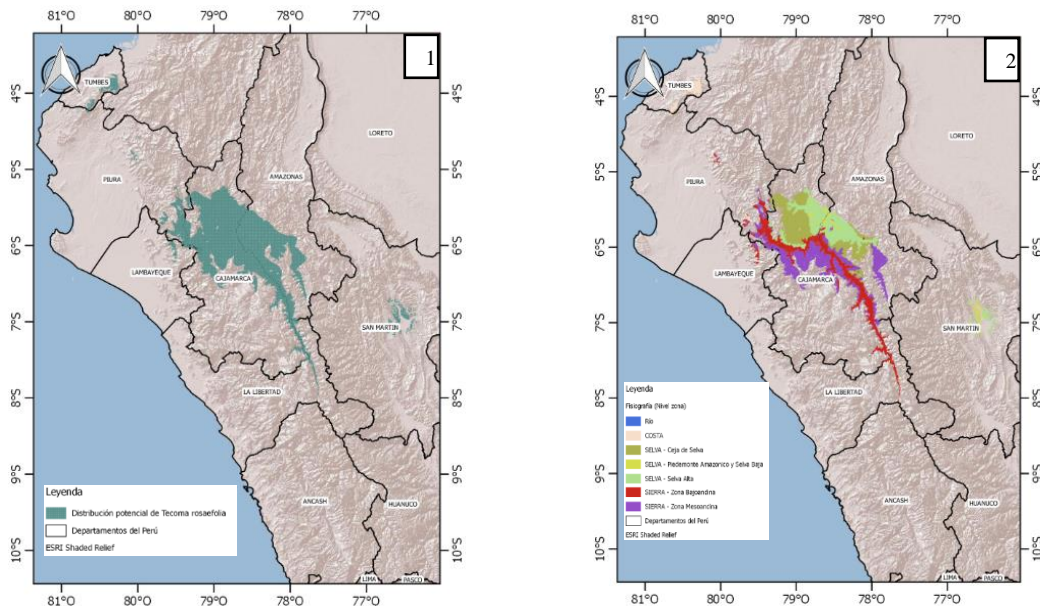
Figura 26: Contribución de variables en el modelo de *Tecoma rosaefolia*.

El mayor porcentaje de contribución de esta variable se debe a que la distribución de los registros, como se observó en la Figura 26, se encuentran dentro de una misma zona sujeta a un mismo patrón de estacionalidad climática. Por lo cual, la isotermalidad representa el comportamiento promedio mensual del rango de temperatura frente al rango anual de temperatura que considera como extremo la temperatura del mes más frío y del mes más cálido, acorde a Hijmans *et al.* (2006). Estos registros al parecer mantienen una tendencia común en la variable isotermalidad. Adicionalmente, se observa que la precipitación en el período más seco mantiene un alto porcentaje de contribución y como se mencionó esta variable es una limitante de los Bosques Tropicales Estacionalmente Secos (BTES).

Los valores de idoneidad del modelo para los registros de entrenamiento variaron de 0,51 a 0,91; por lo cual se estableció el umbral a partir del menor valor. Finalmente, a esta distribución potencial establecida por el umbral se intersecto con la capa de Bosques

Tropicales Estacionalmente Secos propuesta por Marcelo-Peña *et al.* (2015) y limitando a rangos de altitud entre 300 a 2800 según lo mencionado por Marcelo-Peña *et al.* (2010).

La Figura 27 muestra la distribución potencial de *Tecoma rosaefolia* obtenida (imagen 1) y la isiografía a nivel de zona enmarcada en su distribución (imagen 2).



1) Distribución potencial por Departamentos; 2) Distribución potencial en zonas fisiográficas.

Figura 27: Distribución potencial de *Tecoma rosaefolia*

La Figura 27 (imagen 1) muestra la distribución potencial de *Tecoma rosaefolia* en los Departamentos de Amazonas, Cajamarca, La Libertad, Lambayeque, Piura, San Martín y Tumbes. A pesar que los registros en su mayoría estuvieron focalizados en Cajamarca y alrededores, el modelo logro detectar condiciones de similitud ambiental en los Bosques Estacionalmente Secos de San Martín y Tumbes. La Figura 27 (imagen 2) muestra que la distribución potencial abarca Costa (3 por ciento), Ceja de Selva (23 por ciento), Selva-Piedemonte amazónico y selva baja (6 por ciento), Selva alta (12 por ciento), Sierra – Zona Bajoandina (27 por ciento) y Sierra - Zona Mesoandina (29 por ciento). Según Marcelo-Peña *et al.* (2010), esta especie es poco frecuente, endémica del norte y debe ser considerada como vulnerable, por lo cual es de gran importancia aquellas zonas donde el modelo considera existe un potencial para su establecimiento en pro de su conservación.

3.5. *Esenbeckia cornuta* Engl.

No hubo registros duplicados, ni registros agrupados en una sola celda por lo que se utilizó la misma cantidad de registros colectados para las especies. Estos 11 registros estuvieron distribuidos en los Departamentos de Amazonas y Cajamarca. La Figura 28, muestra los registros para *Esenbeckia cornuta*.

La configuración de complejidad elegida fue *features linear-quadratic* y multiplicador de regularización 1, el algoritmo *maxent* le otorgó la siguiente abreviación LQ_1. Obteniendo como resultado 11 modelos con la metodología “*Leave-one-out*”.

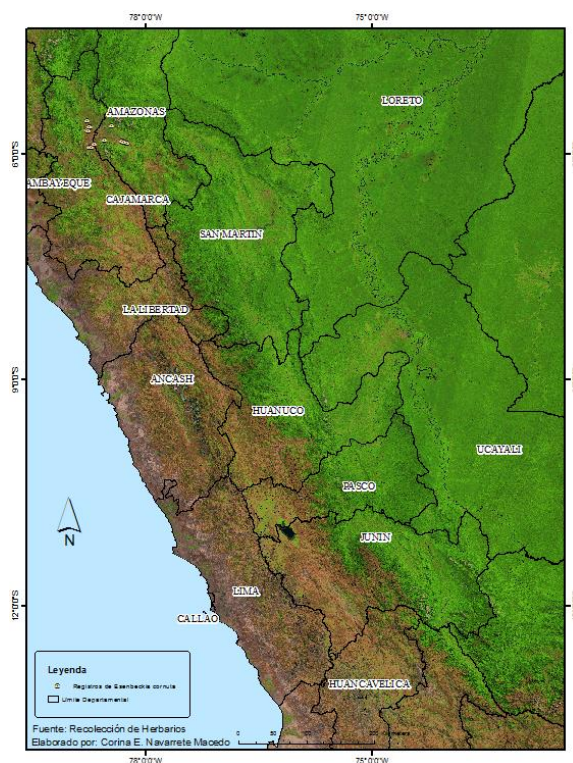


Figura 28: Distribución de Registros de *Esembeckia cornuta*

Como resultado de aplicar el umbral a partir del valor del menor registro de entrenamiento, se obtuvo si el registro excluido fue predicho o no. La Tabla 14, muestra para cada modelo la condición de éxito o fracaso en la predicción del registro excluido. En ella se observa que 11 de los 11 modelos predijeron el registro excluido.

Tabla 14: Resultado de predicción de registro excluido en *Esenbeckia cornuta*

Modelos	Identificador del Registro No incluido	¿Fue predicho? Éxito(1), Fracaso(0)
Modelo 1	1	1
Modelo 2	2	1
Modelo 3	3	1
Modelo 4	4	1
Modelo 5	5	1
Modelo 6	6	1
Modelo 7	7	1
Modelo 8	8	1
Modelo 9	9	1
Modelo 10	10	1
Modelo 11	11	1

Este resultado muestra que, a pesar de excluirse un registro en cada modelo, los demás contenían la información del registro excluido, lo que podría deberse a que los registros se encuentran focalizados en una zona. En comparación con las demás especies estudiadas *Esenbeckia cornuta* presenta sus registros como un agrupamiento, su distribución restringida va acorde a lo manifestado por Marcelo-Peña *et al.* (2010).

La Tabla 15, muestra los resultados de la prueba de significancia para la especie, a un nivel de confianza de 0,05:

Tabla 15: Resultado de prueba de significancia Pearson para *Esenbeckia cornuta*

Especie	Nro. registros	Proporción de éxito	p-valor
Esenbeckia cornuta	11	1	0,0003

Se observa un p-valor de 0,0003; menor al nivel de significancia 0,05; por lo cual hay evidencia que se puede obtener un modelo estadísticamente significativo, con registros menores a 20, en concordancia a lo manifestado por Pearson *et al.* (2007).

Se obtuvo un modelo final, con los 11 registros como entrenamiento y los *features linear-quadratic* con multiplicador de regularización 1. En la Figura 29, se muestra las variables de importancia a través de su contribución en el modelo. La variable de mayor contribución

Precipitación en el periodo más seco en mm (Bio14), seguido del Rango anual de Temperatura en °C (Bio7) y la Isotermalidad (Bio3).

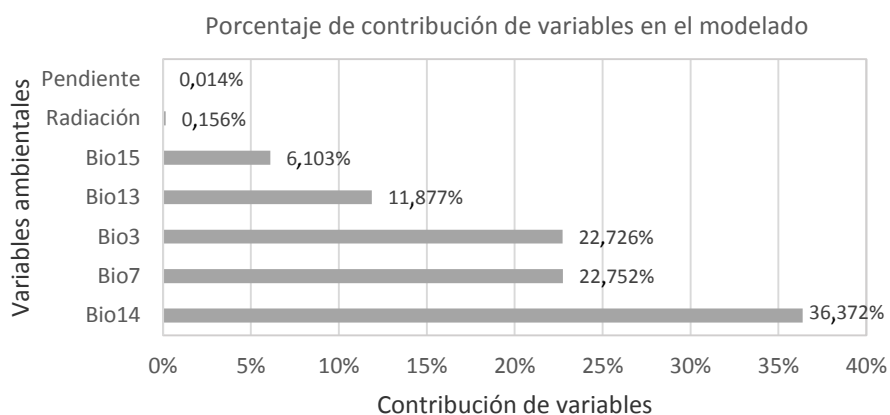
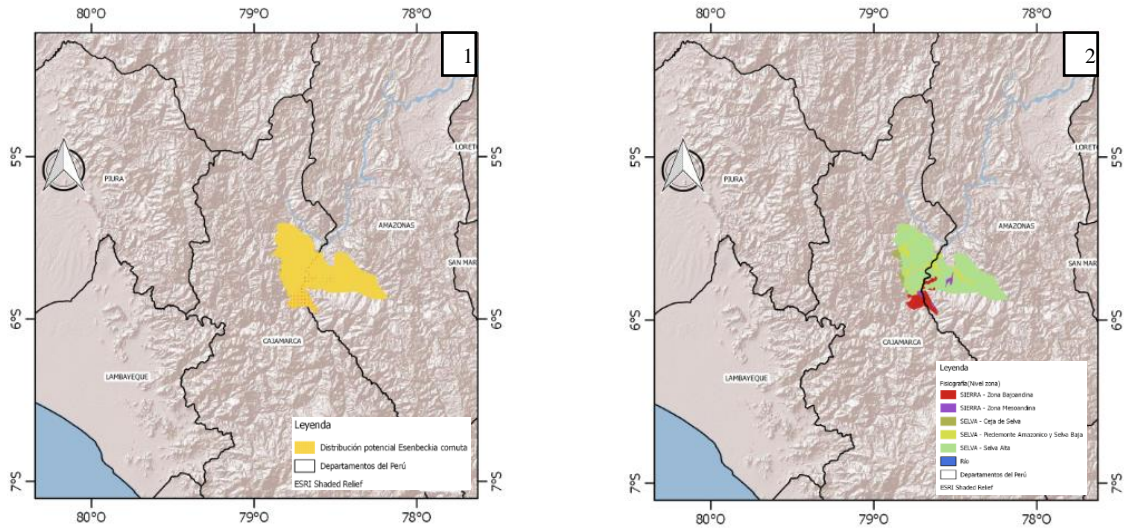


Figura 29: Contribución de variables en el modelo de *Esenbeckia cornuta*

La Figura 29, muestra una vez más que la variable Precipitación en el periodo más seco en mm, es la de mayor contribución en el modelo al igual que lo fue en los modelos de *Buddleja incana*, *Polylepis racemosa* y *Cedrela kuelapensis*. Al parecer esta variable que representa la resistencia a la sequía es de gran importancia para las especies altoandinas y aquellas de los BTES. Las otras variables de importancia en el aporte al modelo son Rango anual de Temperatura en °C y Isotermalidad que representan los extremos de temperatura a las que podría estar sujeta la especie.

Los valores de idoneidad del modelo para los registros de entrenamiento variaron de 0,77 a 0,89; por lo cual se estableció el umbral a partir del menor valor. Finalmente, a esta distribución potencial establecida por el umbral se interseco con la capa de Bosques Tropicales Estacionalmente Secos propuesta por Marcelo-Peña *et al.* (2015) y limitando a rangos de altitud entre 500 a 700 según lo mencionado por Marcelo-Peña *et al.* (2010).



1) Distribución potencial por Departamentos; 2) Distribución potencial en zonas fisiográficas.

Figura 30: Distribución potencial de *Esenbeckia cornuta*

Los resultados en la Figura 30 muestran que *Esenbeckia cornuta* presenta una primera aproximación a su distribución potencial de manera focalizada y continua, a diferencia de las distribuciones potenciales observadas para *Tecoma rosaefolia* en la Figura 27 y *Cedrela kuelapensis* en la Figura 24. Al parecer las condiciones presentadas por los registros no son similares en otras zonas, además de ser la especie considerada de rango geográfico restringido según Marcelo-Peña (2008). Otra limitante para su distribución es la altitud, que posee menor rango en comparación a las otras especies representativas de los BTES. En la Figura 30 (imagen 2) se observa que la distribución potencial se encuentra en las zonas de Selva alta (81 por ciento), Sierra- Zona bajoandina (10 por ciento), Selva-Piedemonte amazónico y selva baja (5 por ciento), Selva-Ceja de selva (2,5 por ciento), Sierra-Zona Mesoandina (1,5 por ciento). *Esenbeckia cornuta* es una especie sujeta a la desaparición de sus poblaciones por la demanda de su madera y la distribución potencial obtenida nos da indicios de lo restringida que podría ser su restauración, por lo cual es una especie que debería ser considerada en peligro crítico, en concordancia con Marcelo-Peña (2008).

V. CONCLUSIONES

- 1) La Distribución potencial obtenida para *Buddleja incana* Ruiz & Pav. y *Polylepis racemosa* Ruiz & Pav fue considerada como de “alta predictibilidad”.
- 2) La Distribución potencial obtenida para *Cedrela kuelapensis* T.D. Penn. & Daza, *Tecoma rosaefolia* Kunth y *Esenbeckia cornuta* Engl, debe considerarse como primeras aproximaciones geográficas a su idoneidad de hábitat.
- 3) Las variables de mayor importancia para especies andinas y de bosque seco fueron la precipitación en el periodo más seco y la isotermalidad.
- 4) La configuración de parámetros para obtener modelos optimos en el algoritmo *maxent*, no es la que presenta por defecto.

VI. RECOMENDACIONES

- Se recomienda para futuros modelamientos en tamaños pequeños de muestra seguir la metodología observada en la tesis, para cerciorarse que el mapa a producir es realmente significativo y refleja condiciones reales.
- Es recomendable que se realicen mayor número de investigaciones en modelamiento predictivo enfocados a nichos ecológicos y riqueza de especies, que comparen las distintas bases climáticas (Wordclim, Chelsa climate, etc) disponibles en la actualidad, con el fin de generar modelos cada vez más precisos.
- Se recomienda realizar modelamientos basados adicionalmente en el espectro, es decir teniendo como variable de entrada compuesto por pixeles multianuales como el que dispone la Universidad de Turku, Finlandia. Sería interesante poder integrar a los modelos basados en características ambientales, estos últimos basados en características espectrales, con el fin de obtener la Distribución de la especie.

VII. REFERENCIAS BIBLIOGRÁFICAS

- Anderson, R.P; Gonzalez, I. 2011. Species-specific tuning increases robustness to sampling bias in models of species distributions: An implementation with Maxent. *Ecological Modelling* 222:2796-2811.
- Araújo, M.B; Guisan, A. 2006. Five (or so) challenges for species distribution modelling. *Journal of Biogeography* 33:1677-1688.
- Arica, D. 2003. Beneficios del sistema agroforestal de la comunidad campesina de Antacusi. (En línea). Consultado 21 mar. 2017. Disponible en: <https://es.scribd.com/doc/293283614/BENEFICIOS-DEL-SISTEMA-AGROFORESTAL>.
- Austin, M. 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecological modelling* 200:1-19.
- Baselga, A; Araújo M.B. 2009. Individualistic vs. community modelling of species distributions under climate change. *Ecography* 32:55-65.
- Benito, B. 2015. XI Taller GBIF de modelización de nichos ecológicos. (En línea). Consultado 25 ene. 2017. Disponible en: <http://www.gbif.es/formaciondetalles.php?IDForm=135>.
- Brako, L; Zarucchi, J. 1993. Catálogo de las Angiospermas y Gimnospermas del Perú. *Monogr. Syst. Bot. Missouri Bot. Garden.* 45:763-867.
- Cain, S.A. 1944. *Foundations of Plant Geography*. New York, US, Harper & Brothers. 556 p.
- Cantor, S.B; Sun, C.C; Tortolero-Luna, G; Richards-Kortum, R; Follen, M. 1999. A comparison of C/B ratios from studies using receiver operating characteristic curve analysis. *Journal of Clinical Epidemiology* 52:885–892.
- CONABIO (Comisión Nacional de Biodiversidad). 2016. Nichos y áreas de distribución. (En línea). Consultado 6 feb. 2017. Disponible en: <http://nicho.co-nabio.gob.mx/>

- Costales, D. 2005. Guía para el establecimiento de plantaciones forestales en la zona andina. FAO.
- Cramer, J.S. 2003. Logit models: from economics and other fields. Cambridge, UK, Cambridge University Press. 184 p
- Elith, J; Graham, C; Anderson, R; Dudik, M; Ferrier, S; Guisan, A; Hijmans, R; Huettmann, F; Leathwick, J; Lehmann, A; Li, J; Lohmann, L; Loiselle B.A; Manion, G; Moritz, C; Nakamura, M; Nakazawa, Y; Overton, J.M; Peterson, A.T; Phillips, S; Richardson, K; Scachetti-Pereira, R; Schapire, R; Soberón, J; Williams, S; Wisz, M; Zimmermann, N. 2006. Novel methods improve prediction of species distributions from occurrence data. *Ecography* 29:129-151.
- Elith, J; Leathwick J. 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution and Systematics* 40:677-697.
- Elith, J; Phillips, S; Hastie, T; Dudik, M; Yung E.C; Yates, C. J. 2011. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* 17:43-57.
- Ferrer, F.J. 2012. Analisis de datos espaciales en ecología. Lección 5: Distribucion potencial de especies. (En línea). Consultado 4 feb. 2017. Disponible en: https://fjferreer.webs.ull.es/Apuntes05/Tema_5.pdf
- Fielding, A.H; Bell J.F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24:38-49.
- Fielding, A.H; Haworth, P.F. 1995. Testing the generality of bird-habitat models. *Conservation Biology* 9:1466-1481
- Franklin, J. 2009. Mapping Species Distributions: Spatial Inference and Prediction. Cambridge, UK, University Press. 340 p.
- Freeman, E.A; Moisen, G.G. 2008. A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and Kappa. *Ecological Modelling* 217:48-58.
- Good, R. 1931. A theory of plant geography. *The New Phytologist* 30:149-171.

- Graham, C; Ferrier, S; Huettman, F; Moritz, C; Peterson, AT. 2004. New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology and Evolution* 19:497–503.
- Grinnell, J. 1917. Field tests of theories concerning distributional control. *American Naturalist* 51:115-128.
- Guisan, A; Edwards, T.C; Hastie, T. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling* 157:89-100.
- Guisan, A; Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. *Ecological Modeling* 135:147-186.
- Guisan, A; Zimmermann, N.E., Elith, J., Graham, C.H., Phillips, S. & Peterson, A.T. 2007. What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? *Ecological Monographs* 77: 615–630.
- Hanley, J.A; McNeil, B. J. 1982. The meaning and use of the area under a receiver operating characteristic (ROC curve). *Radiology*, 143(1), 29–36.
- Hernández, P.A; Franke, I; Herzog, S.K; Pacheco, V; Paniagua, L; Quintana, H.L; Soto, A; Swenson J.J; Tovar, C; Valqui, T.H; Vargas, J; Young, B.E. 2008. Predicting species distributions in poorly-studied landscapes. *Biodiversity and Conservation* 17(6):1353-1366.
- Hijmans, R.J; Cameron, S; Parra, J; Jones, P; Jarvis, A; Richardson, K. 2006. World-Clim version 1.4. (En línea). Consultado 11 mar. 2017. Disponible en: <http://www.worldclim.org/>.
- Hijmans, R.J. 2016. *Spatial Data Analysis and Modeling with R*. (En línea). Consultado 8 feb. 2017. Disponible en: <http://rspatial.org>
- Hirzel, A.H; Hausser, J; Chessel, D; Perrin, N. 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data?. *Ecology* 83:2027-2036.
- Huntley, B; Berry, P.M; Cramer, W; McDonald, A.P. 1995. Modelling present and potential future ranges of some European higher plants using climate response surfaces. *Journal of Biogeography* 22:967-1001.

- Huntley, B; Green, R. E; Collingham, Y.C; Hill, J.K; Willis S.G; Bartlein, P.J; Cramer, W; Hagermeijer, W.J.M; Thomas, C. J. 2004. The performance of models relating species geographic distributions to climate is independent of trophic level. *Ecology Letters* 7:417-426.
- Hutchinson, G.E. 1957. Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology* (22):415-427.
- IUCN (Union Internacional para la Conservación de la Naturaleza). 2016. Lista Roja de especies amenazadas. (En línea). Consultado 8 feb. 2017. Disponible en: <http://www.iucnredlist.org/search>
- Jiménez- Valverde, A; Lobo, J.M. 2007. Threshold criteria for conversion of probability of species presence to either- or presence-absence. *Acta oecologica* 31:361-369.
- Jiménez- Valverde, A; Lobo, J.M; Hortal, J. 2008. Not as good as they seem: the importance of concept in species distribution modelling. *Diversity and Distributions* 14:885-890.
- Johnson, R.H. 1910. Determinate evolution in the color-pattern in the lady beetles. *Carnegie Institution of Washington* 122:87-89.
- Jung, M. 2014. QSDM – Species Distribution Modelling for the QGIS Processing Toolbox. (En línea). Consultado 22 feb. 2017. Disponible en: <https://conservationecology.wordpress.com/qgis-plugins-and-scripts/qsdm>
- Lao, R; Zevallos, P; De La Cruz, H. 1990. Información preliminar de la ecología, dendrología y distribución geográfica de las especies del género *Polylepis* en el Perú. *Revista Espacio y Desarrollo* 2: 47-62.
- Laurente, M. 2015. Efectos del cambio climático en la distribución del Cedro (*Cedrela odorata* L.) en la Amazonia Peruana. *The Biologist* 13(2):.213-224.
- León, J. 1988. Identificación y caracterización dendrológica de 18 especies leñosas de la zona de Chaipi Ayacucho. Tesis Ing. Lima, PE, Universidad Nacional Agraria La Molina. 158 p.
- Liras, E. 2008. IV Taller de modelización de nichos ecológicos. (En línea). Consultado 24 ene. 2017. Disponible en: <http://www.gbif.es/formaciondetalles.php?IDForm=41#ad-image-0>

- Liu, C.; Berry, P.M; Dawson, T.P; Pearson, R.G. 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* (28):385-393.
- Lobo, J.M; Jiménez-Valverde, A; Real, R. 2007. AUC: a misleading measure of the performance of predictive distribution models. *Global ecology and biogeography* 17:145-151.
- Londoño, MC. 2012. Nociones de modelización de nicho ecológico. (En línea). Consultado 5 mar. 2017. Disponible en: <http://www.recibio.net/wpcontent/uploads/2012/02/ModelosDistribucionEspecies-MCL.pdf>
- Luck, G.W. 2002. The habitat requirements of the rufous treecreeper (*Climacteris rufa*): Validating predictive habitat models. *Biological Conservation* 105:395-403.
- Mackey, B. G.; Lindenmayer, D. B. 2001. Towards a hierarchical framework for modelling the spatial distribution of animals. *Journal of Biogeography* 28:1147-1166.
- Manel, S; Williams H.C; Ormerod, S.J. 2001. Evaluating presences-absence models in ecology: The need to account for prevalence. *Journal of Applied Ecology* 38:921–931.
- Marcelo--Peña, J. L. 2008. Vegetación leñosa, endemismos y estado de conservación en los bosques estacionalmente secos de Jaén, Perú. *Rev. Per. Biol.* 15(1): 43-52.
- Marcelo--Peña, J. L; Pennington, R; Reynel, C; Zevallos, P. 2010. Guía ilustrada de la flora leñosa de los Bosques Estacionalmente Secos de Jaén, Perú. UNALM/Royal Botanic Garden Edinburgh. Lima, PE. 288 p.
- Marcelo--Peña, J. L; Huamantupa, I; Särkinen, T; Tomazello, M. 2016. Identifying Conservation Priority Areas in the Marañón Valley (Peru) Based On Floristic Inventories. *Edinburgh Journal of Botany*,73(1):95-123.
- Marmion, M; Parviainen, M; Luoto, M; Heikkinen, R.K; Thuiller, W. 2009. Evaluation of consensus methods in predictive species distribution modelling. *Diversity and distributions*, 15:59-69.
- Martínez-Meyer, E. 2014. Introducción al modelado de nicho ecológico. (En línea). Consultado 5 mar. 2017. Disponible en: http://www.recibio.net/wp-content/uploads/2014/01/Intro-Modelado-de-nichos_EMM.pdf

- Mateo, R; Felicísimo, A; Muñoz, J. 2011. Modelos de distribución de especies: Una revisión sintética. *Revista chilena de historia natural* 84(2):217-240. (En línea). Consultado 7 mar. 2017. Disponible en: <https://dx.doi.org/10.4067/S0716-078X2011000200008>
- Mejía, A.A. 2014. Metodología para la cartografía de bosques del generó Polylepis, aplicando geomática. Tesis Ing. Lima, PE, Universidad Nacional Agraria La Molina. 106 p.
- Merow, C; Smith, M.J; Silander, J.A. 2013. A practical 269 guide to MaxEnt for modeling species distributions: what it does, and why inputs and settings matter. *Ecography* 36:1058-1069.
- Morales, N. 2012. Modelos de distribución de especies: Software Maxent y sus aplicaciones en conservación. *Revista Conservación Ambiental* 2(1): 1-3.
- Morales, N; Fernandez, I; Baca, V. 2016. MaxEnt's parameter configuration and small samples: Are we paying attention to recommendations? Publimed. (En línea). Consultado 7 mar. 2017. Disponible en: <https://www.ncbi.nlm.nih.gov/pubmed/28316894>
- Osborne, P.E; Alonso, J.C; Bryant, R.G. 2001. Modelling landscape-scale habitat use using GIS and remote sensing: a case study with great bustards. *Journal of Applied Ecology* 38:458-471.
- Ortega-Huerta, M.A., Peterson, A.T., 2008. Modeling ecological niches and predicting geographic distributions: a test of six presence-only methods. *Rev. Mex. Biodivers.* 79 (1), 205—216.
- OSINFOR (Organismo de Supervisión de los Recursos Naturales y Fauna Silvestre). 2013. *Distribucion de las Especies Forestales del Perú*. Lima, PE. 294 p.
- OSINFOR (Organismo de Supervisión de los Recursos Naturales y Fauna Silvestre). 2016. *Modelamiento de la distribución potencial de 18 especies forestales en el departamento de Loreto*. Lima, PE. 90 p.
- Pearson, R.G; Dawson T. P. 2003. Predicting the impacts of climate change on the distribution of species: Are bioclimate envelope models useful? *Global Ecology and Biogeography* 12:361-371.

- Pearson, R.G; Nakamura, M; Peterson A.T; Raxworthy, C. 2007. Predicting species distributions from small numbers of occurrence records: A test case using cryptic geckos in Madagascar. *Journal of Biogeography* 34:102-117.
- Pennington, T.D; Muellner, A, N. 2010. A monograph of *Cedrela*(Meliaceae). California, US, Masalai Press. 112 p.
- Peterson, A.T. 2006. Uses and requirements of ecological niche models and related distributional models. *Biodiversity Informatics* 3:59-72.
- Peterson; Soberón, J; Pearson, R; Anderson, R; Martínez-Meyer, E; Nakamura, M; Araújo, M. 2011. *Ecological Niches and Geographic Distributions*. Nueva Jersey, US, Princeton University Press. 328 p.
- Peterson; Papes, M; Carroll, D.S; Leirs, H; Johnson K. M. 2007a. Mammaltaxa constituting potential coevolved reservoirs of filoviruses. *Journal of Mammalogy* 88:1544-1554.
- Peterson; Papes, M; Eaton, M. 2007b. Transferability and model evaluation in ecological niche modeling: A comparison of GARP and Maxent. *Ecography* 30:550-560.
- Peterson; Papes, M; Soberon, J. 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling* 213:63-72.
- Phillips, S.J; Anderson, R.P; Schapire, R.E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190:231-259.
- Phillips, S.J; Dudík, M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31:161-175.
- Pliscoff, P; Fuentes-Castillo, T. 2011. Modelación de la distribución de especies y ecosistemas en el tiempo y en el espacio: una revisión de las nuevas herramientas y enfoques disponibles. *Revista de geografía Norte Grande* (48):61-79. (En línea). Consultado 7 mar. 2017. Disponible en: <https://dx.doi.org/10.4067/S0718-34022011000100005>
- Reynel, C. 1988. Plantas para leña en el Sur-occidente de Puno. Proyecto Árbol andino-Puno. Puno-Peru. 165 p.
- Reynel, C; León, J. 1990. Árboles y arbustos andinos para agroforestería y conservación de suelos. FAO/HOLANDA/INFOR. Lima, Perú. 500 p.

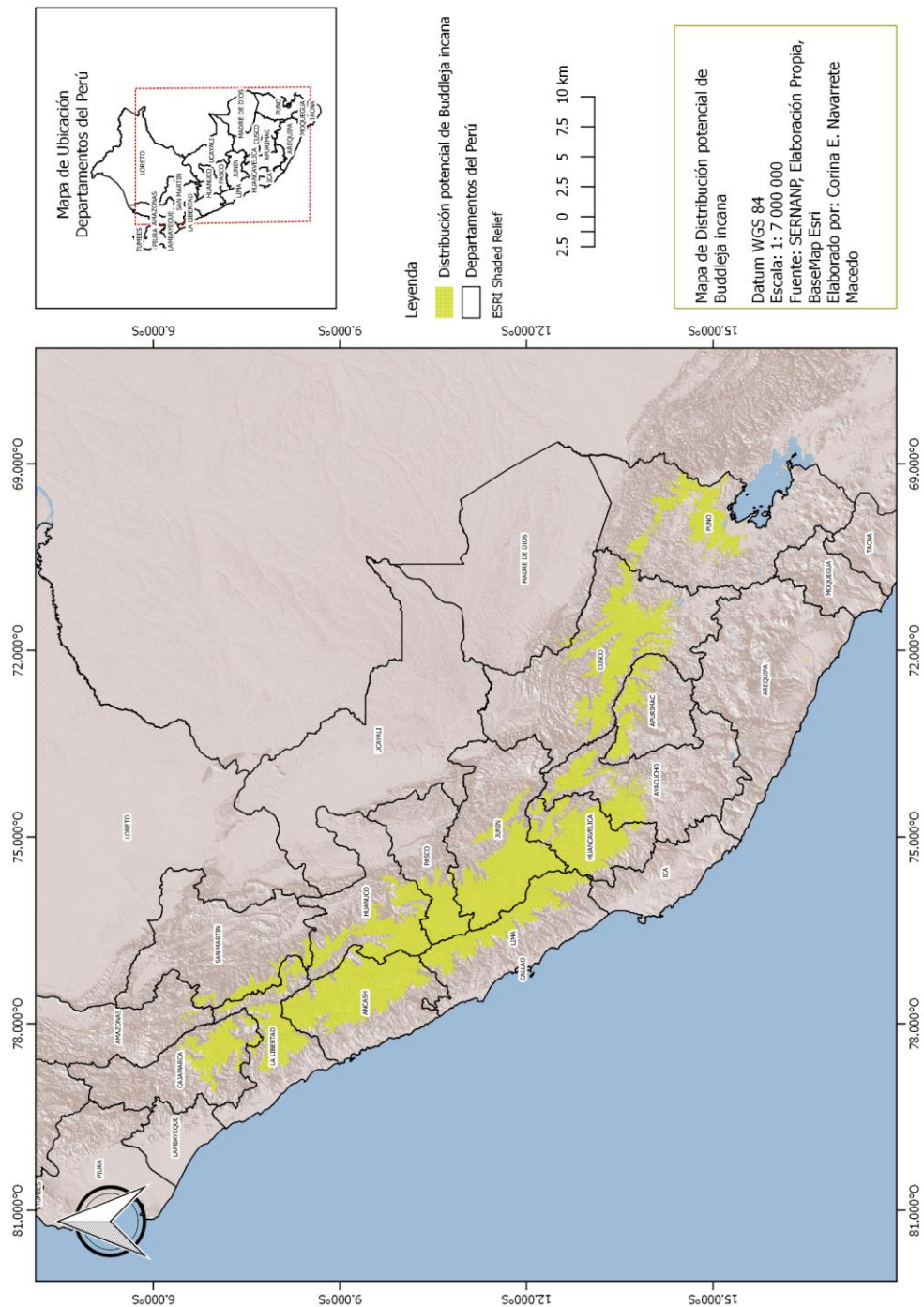
- Reynel, C; y Marcelo, J. 2010. Arboles de los Ecosistemas Forestales Andinos. (En línea). Consultado 23 mar. 2017. Disponible en: <http://www.asocam.org/biblioteca/files/original/02190ca87f921d9feb250ae672ae6653.pdf>
- Rykiel, E.J., 1996. Testing ecological models: the meaning of validation. *Ecol. Model.* 90, 229–244
- Särkinen, T; Gonzáles, P; Knapp, S. 2013. Distribution models and species discovery: the story of a new *Solanum* species from the Peruvian Andes. *PhytoKeys* (31):1–20.
- Shcheglovitova, M; Anderson, R.P. 2013. Estimating optimal complexity for ecological niche models: A jackknife approach for species with small sample sizes. *Ecological Modelling* 269:9-17.
- Silver, N. 2012. *La señal y el ruido*. España, Grupo planeta. 656 p.
- Simpson, B. B. 1979. A revisión of the Genus *Polylepis* (Rosaceae: Sanguisorbeae). (En línea). Consultado 3 feb. 2017. Disponible en: https://www.researchgate.net/publication/44338887_A_Revision_of_the_Genus_Polylepis_Rosaceae_Sanguisorbeae
- Soberón, J; Peterson A.T. 2005. Interpretation of models of fundamental ecological niches and species distributional areas. *Biodiversity Informatics* 2:1-10.
- Soberón, J; Peterson, A.T. 2008. Monitoring biodiversity loss with primary species-occurrence data: Toward national-level indicators for the 2010 Target of the Convention on Biological Diversity. *AMBIO* 38:29-34.
- Soberón, J; Nakamura, M. 2009. Niches and distributional areas: concepts, methods and assumptions. (En línea). Consultado 7 mar. 2017. Disponible en: <https://www.ncbi.nlm.nih.gov/pubmed/19805041>
- Soberón, J; Osorio-Olvera, L; Peterson, T. 2017. Diferencias conceptuales entre modelación de nichos y modelación de áreas de distribución. *Revista Mexicana de Biodiversidad* 88(2):437-441. (En línea). Consultado 2 jun. 2017. Disponible en: <http://www.sciencedirect.com/science/article/pii/S1870345317300611>
- Stockwell, D.R; Peters, D. 1999. The GARP modelling system: Problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science* 13:143-158

- Stockwell, D.R; Peterson, A.T. 2002. Controlling Bias in Biodiversity Data. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A. and Samson, F.B., Eds., *Predicting Species Occurrences: Issues of Accuracy and Scale*. Washington DC, US, Island Press. 537-546p.
- Stockwell, D.R; Peterson, A.T. 2002. Effects of sample size on accuracy of species distribution models. *Ecological Modelling* 148:1-13.
- Thuiller, W; Lafourcade, B; Engler, R; Araujo, M. 2009. Biomod a platform for ensemble forecasting of species distributions. *Ecography* 32:369-373.
- Udvardy, M. 1969. *Dynamic Zoogeography. With Special Reference to Land Animals*. (En línea). Consultado 7 mar. 2017. Disponible en: <http://trove.nla.gov.au/work/21157155?selectedversion=NBD44545>
- Warren, D.L; Seifert, S.N. 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications* 21:335-342.
- Wisz, M.S; Hijmans, R.J; Li, J; Peterson, A.T; Graham, C.H; Guisan, A. 2008. Effects of sample size on the performance of species distribution models. *Diversity and Distributions* 14: 763-773.
- Yee, T. W; Mitchell, N.D. 1991. Generalized additive models in plant ecology. *Journal of Vegetation Science* 2:587-602.
- Zutta, B.R; Rundel P.W; Saatchi, S; Casana, J.D; Gauthier, P; Soto, A; Velazco, Y; Buermann, W. 2012. Prediciendo la distribución de *Polylepis*: bosques Andinos vulnerables y cada vez más importantes. *Revista Peruana de Biología* 19(2): 205-212. (En línea). Consultado 6 feb. 2017. Disponible en: http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S1727-99332012000200013&lng=es&nrm=iso. ISSN 1727-9933.

VIII. ANEXOS

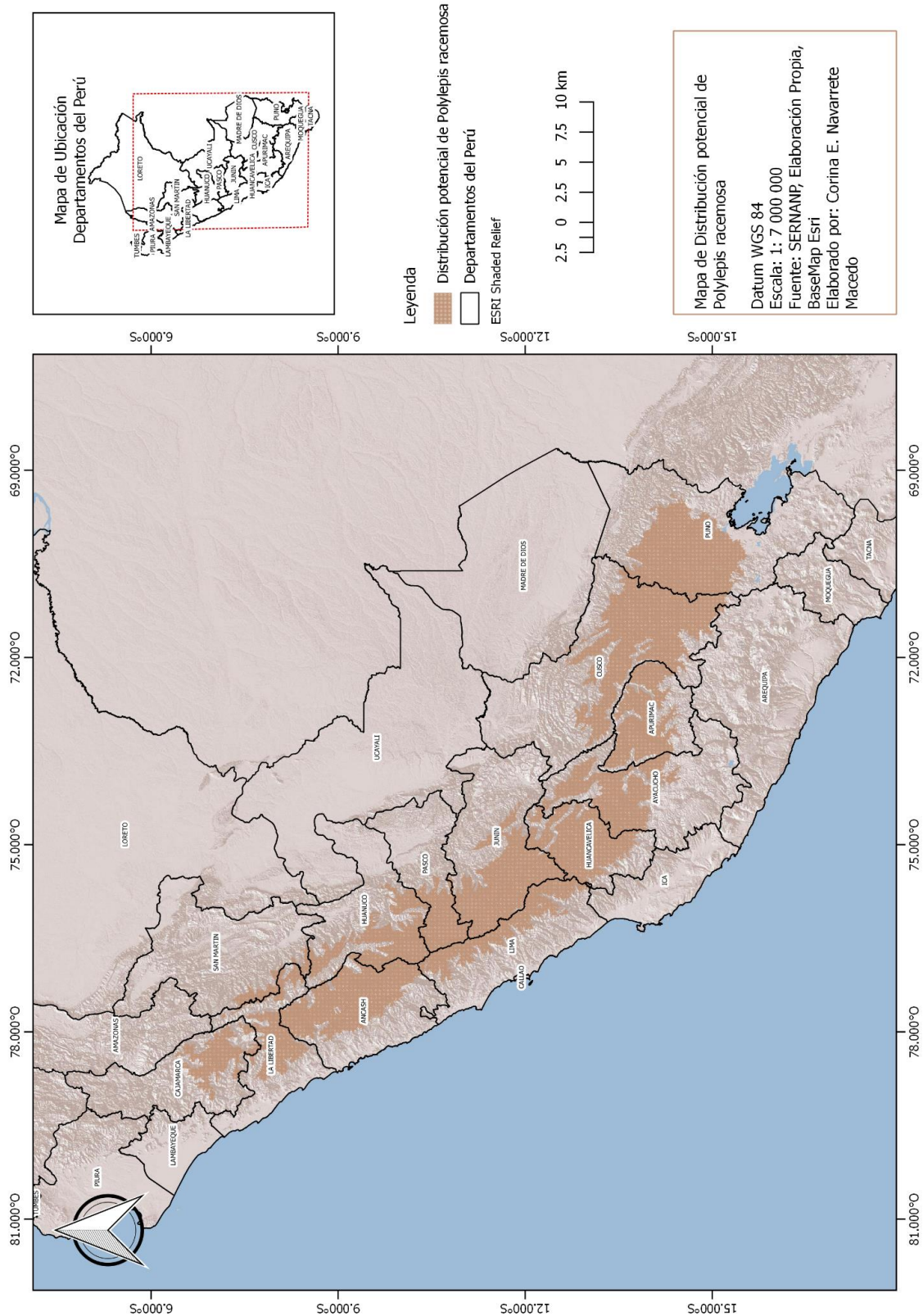
ANEXO 1

MAPA DE DISTRIBUCIÓN POTENCIAL DE *BUDDLEJA INCANA* RUIZ & PAV.



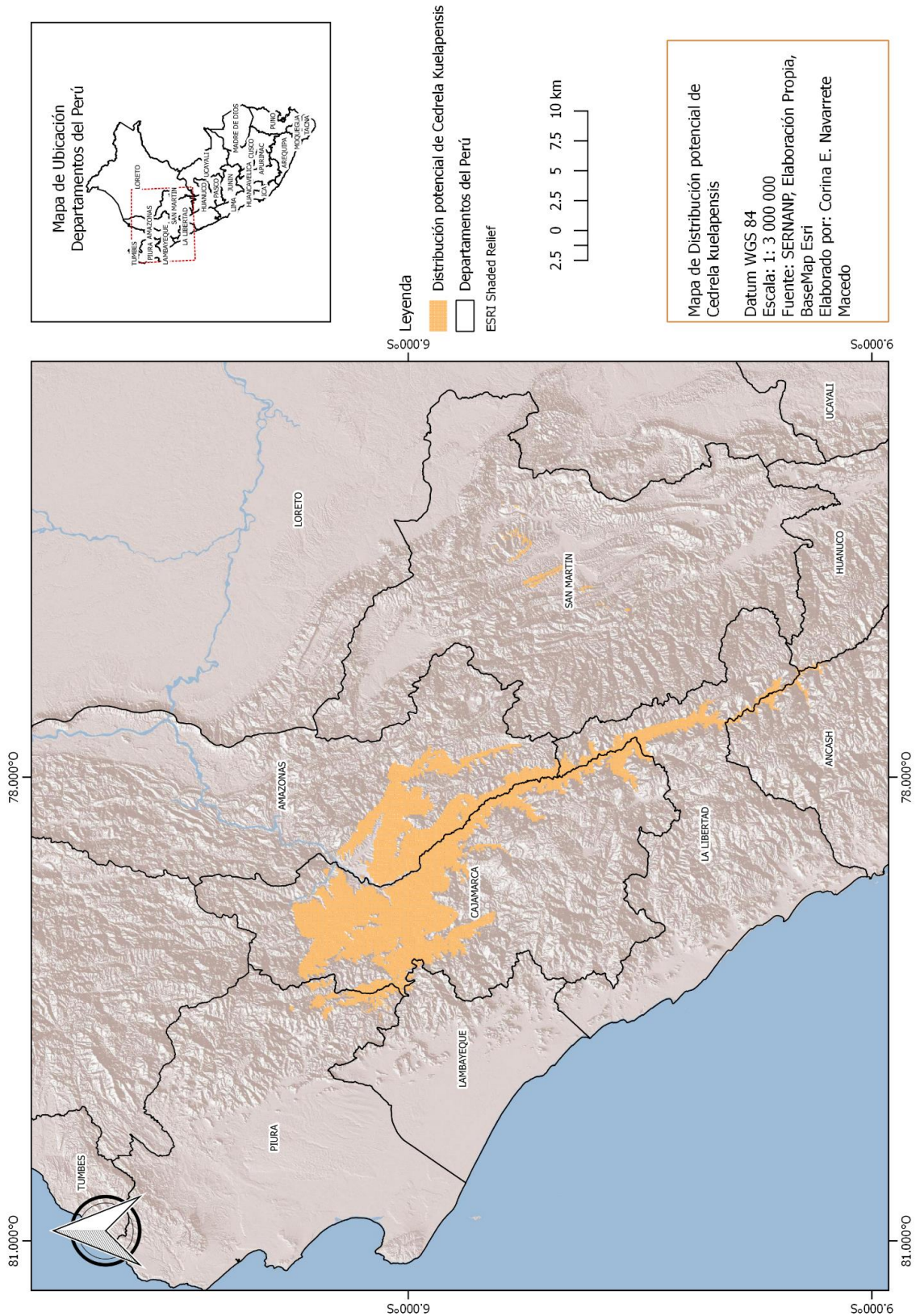
ANEXO 2

MAPA DE DISTRIBUCIÓN POTENCIAL DE *POLYLEPIS RACEMOSA* RUIZ & PAV.



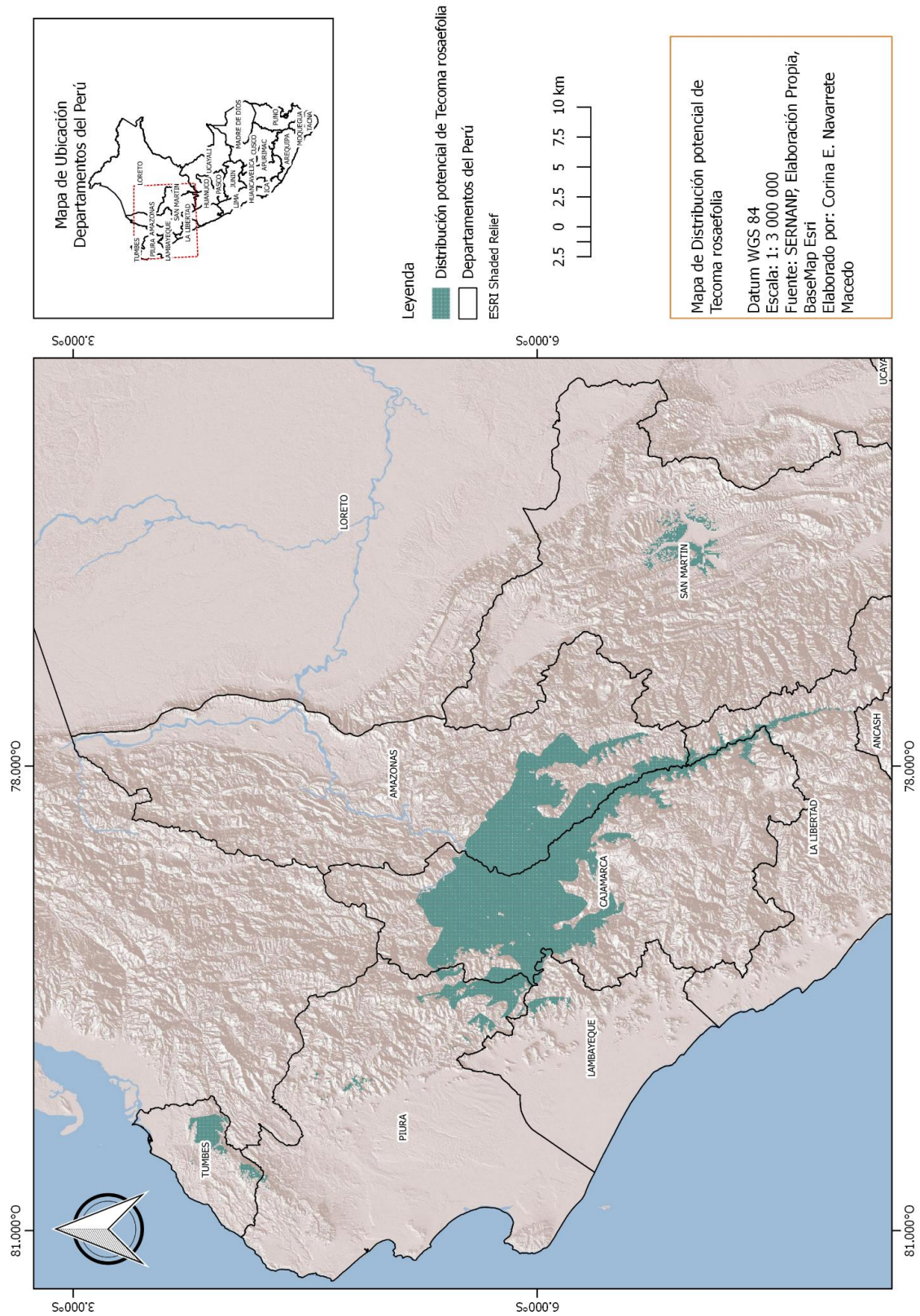
ANEXO 3

MAPA DE DISTRIBUCIÓN POTENCIAL DE *CEDRELA KUELAPENSIS* T.D. PENN. & DAZA



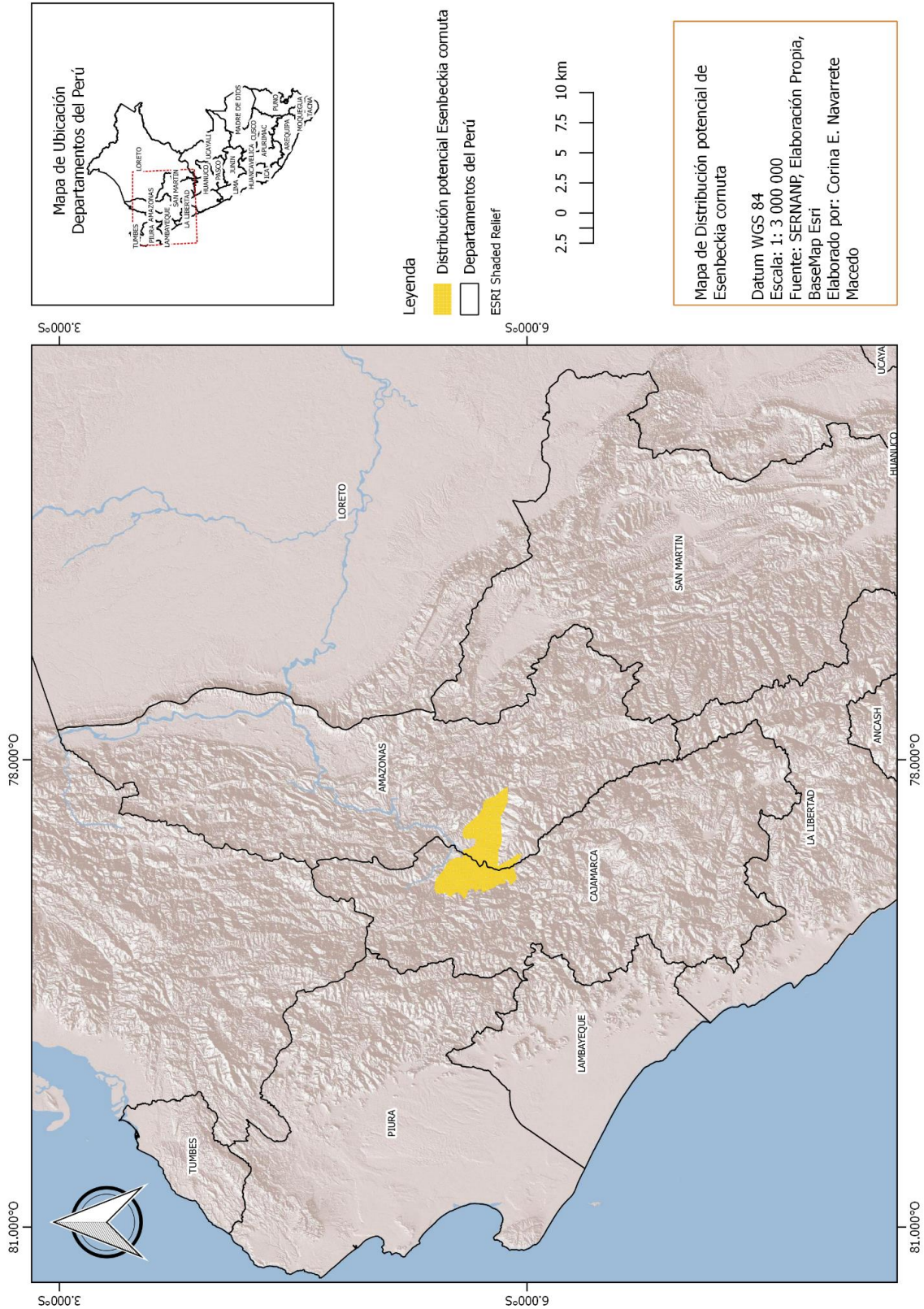
ANEXO 4

MAPA DE DISTRIBUCIÓN POTENCIAL DE *TECOMA ROSAEFOLIA* KUNTH



ANEXO 5

MAPA DE DISTRIBUCIÓN POTENCIAL DE *ESENBECKIA CORNUTA* ENGL.



ANEXO 6

COMBINACIONES DE CONFIGURACIONES DE PARÁMETROS PARA BUDDLEJA INCANA Y POLYLEPIS RACEMOSA

Para *Buddleja incana* y *Polylepis racemosa*: Se obtuvo 35 configuraciones de parámetros de complejidad, las cuales se aplicó a las dos especies. Se enlista dichas configuraciones y sus abreviaturas correspondientes otorgadas por el algoritmo *maxent*:

Configuracion	Features	Multiplicador de regularización
L_0.5	<i>Linear</i>	0,5
L_1	<i>Linear</i>	1
L_1.5	<i>Linear</i>	1,5
L_2	<i>Linear</i>	2
L_2.5	<i>Linear</i>	2,5
LQ_0.5	<i>Linear-quadratic</i>	0,5
LQ_1	<i>Linear-quadratic</i>	1
LQ_1.5	<i>Linear-quadratic</i>	1,5
LQ_2	<i>Linear-quadratic</i>	2
LQ_2.5	<i>Linear-quadratic</i>	2,5
LH_0.5	<i>Linear-Hinge</i>	0,5
LH_1	<i>Linear-Hinge</i>	1
LH_1.5	<i>Linear-Hinge</i>	1,5
LH_2	<i>Linear-Hinge</i>	2
LH_2.5	<i>Linear-Hinge</i>	2,5
LQH_0.5	<i>Linear-quadratic-Hinge</i>	0,5
LQH_1	<i>Linear-quadratic-Hinge</i>	1
LQH_1.5	<i>Linear-quadratic-Hinge</i>	1,5
LQH_2	<i>Linear-quadratic-Hinge</i>	2
LQH_2.5	<i>Linear-quadratic-Hinge</i>	2,5
Q_0.5	<i>Quadratic</i>	0,5
Q_1	<i>Quadratic</i>	1
Q_1.5	<i>Quadratic</i>	1,5
Q_2	<i>Quadratic</i>	2
Q_2.5	<i>Quadratic</i>	2,5
QH_0.5	<i>Quadratic-Hinge</i>	0,5
QH_1	<i>Quadratic-Hinge</i>	1
QH_1.5	<i>Quadratic-Hinge</i>	1,5
QH_2	<i>Quadratic-Hinge</i>	2
QH_2.5	<i>Quadratic-Hinge</i>	2,5
H_0.5	<i>Hinge</i>	0,5
H_1	<i>Hinge</i>	1
H_1.5	<i>Hinge</i>	1,5
H_2	<i>Hinge</i>	2
H_2.5	<i>Hinge</i>	2,5

ANEXO 7

EVALUACIÓN DE RENDIMIENTO DE MODELOS PARA BUDDLEJA INCANA

Configuración de parametros	AUCtest	AUCeval	delta.AICc
LQ_2.5	0,912	0,909	0
LQH_2.5	0,921	0,909	2,861
L_2	0,914	0,876	3,146
LQ_1	0,928	0,918	3,240
L_2.5	0,922	0,912	4,017
LQH_2	0,940	0,922	4,141
LQ_2	0,926	0,917	4,516
L_0,5	0,922	0,915	4,797
LQ_0,5	0,929	0,919	4,847
L_1	0,922	0,915	4,937
L_1,5	0,922	0,914	5,515
LQ_1,5	0,927	0,918	5,522
LH_2	0,939	0,920	6,755
LH_2,5	0,935	0,915	11,081
Q_2,5	0,904	0,900	19,274
Q_1	0,904	0,899	21,922
Q_0,5	0,904	0,899	21,922
Q_1,5	0,904	0,899	22,027
Q_2	0,904	0,899	22,208
QH_2	0,946	0,925	31,428
QH_2,5	0,944	0,925	32,816
LH_1	0,949	0,923	36,412
LH_1,5	0,942	0,921	37,751
LQH_1,5	0,944	0,922	46,033
H_2	0,937	0,917	54,733
H_2,5	0,935	0,913	54,930
QH_1,5	0,947	0,924	58,155
LQH_1	0,949	0,925	82,983
QH_1	0,950	0,924	163,347
H_1,5	0,941	0,919	182,456
LH_0,5	0,957	0,926	1043,969
LQH_0,5	0,957	0,926	2234,518
H_0,5	0,956	0,924	NA
QH_0,5	0,956	0,924	NA
H_1	0,946	0,919	NA

ANEXO 8

EVALUACIÓN DE RENDIMIENTO DE MODELOS PARA *POLYLEPIS RACEMOSA*

Configuración de parametros	AUCtest	AUCeval	delta,AICc
LQ_1	0,907	0,898	0
LH_2	0,906	0,897	0,002
QH_2	0,951	0,925	0,869
LQ_1,5	0,907	0,899	1,245
LQ_2	0,908	0,898	2,642
LQH_2	0,949	0,912	3,067
LQ_2,5	0,908	0,895	4,092
LQ_0,5	0,916	0,900	4,377
QH_2,5	0,949	0,897	8,087
LH_2,5	0,940	0,892	8,684
L_1	0,899	0,888	8,979
L_1,5	0,899	0,889	9,458
L_2	0,901	0,890	10,154
L_2,5	0,901	0,891	10,990
Q_0,5	0,888	0,883	11,213
LH_1,5	0,953	0,923	11,363
Q_1	0,889	0,884	11,458
L_0,5	0,899	0,887	11,812
Q_1,5	0,889	0,885	11,845
Q_2	0,890	0,886	12,314
Q_2,5	0,891	0,887	12,8546
LQH_2,5	0,942	0,894	15,742
LQH_1,5	0,955	0,926	16,347
H_2,5	0,932	0,888	57,177
QH_1,5	0,958	0,929	87,634
H_2	0,945	0,896	101,177
H_1,5	0,954	0,911	173,517
LQH_1	0,960	0,936	232,994
LH_1	0,960	0,934	519,415
QH_0,5	0,967	0,938	719,193
QH_1	0,962	0,937	1182,615
H_0,5	0,967	0,938	NA
LH_0,5	0,967	0,937	NA
LQH_0,5	0,968	0,938	NA
H_1	0,962	0,932	NA

ANEXO 9

SCRIPT DESARROLLADO EN R STUDIO PARA EL MODELADO DE *BUDDLEJA INCANA*

```
options(java.parameters = "-Xmx1g" )
library(raster)
library(dismo)
library(ENMeval)

#Directorio de trabajo
setwd("C:/Users/Corina/Desktop/Distribucionespecies/")
source("funcion_reescalar_raster,R")

#Carga de registros
registros,buddleja<- read.csv("./coordenadas_especies/buddlejaincana.csv")
registros,buddleja <- registros,buddleja[!duplicated(registros,buddleja),]
registros,buddleja <- registros,buddleja[,-1]

#Variables ambientales
variables,ambientales <-stack(list,files("variablesfinalesmaxent/" ,"*tif$", full, names = T))
bg <- randomPoints(variables,ambientales[[1]], n=10000)
bg <- as.data.frame(bg)

#Configuración de modelos
eval1 <- ENMevaluate(occ=registros,buddleja, env=variables,ambientales, bg,coords=bg,
method='randomkfold', kfolds = 4, RMvalues=c(0,5,1,1,5,2,2,5),
fc=c('L','Q','H','LQ','LH','QH','LQH'))

# Evaluadores de rendimiento y gráficos
str(eval1, max,level=3)
eval1@results
tablaestadisticos <- data.frame(eval1@results)
write.csv(tablaestadisticos,file = "modelos/Especiebuddleja/estadisticosrendimiento.csv",
row, names = T)
eval1@results[which(eval1@results$delta,AICc==0),]
eval1@predictions
plot(eval1@predictions[[which(eval1@results$delta,AICc==0)]], main="Relative
occurrence rate")
aic,opt <- eval1@models[[which(eval1@results$delta,AICc==0)]]
aic,opt@results
var,importance(aic,opt)
aic,opt@lambdas
eval,plot(eval1@results)
eval,plot(eval1@results, 'Mean,AUC,DIFF', legend,position = F)
df <- var,importance(aic,opt)
barplot(df$percent,contribution, names,arg=df$variable, las=2, beside=T)
write.csv(df, file = "modelos/Especiebuddleja/importanciavariables.csv", row, names = T)
```

```

##Ploteando las predicciones
par(mfrow=c(1,2), mar=c(2,2,1,0))
plot(eval1@predictions[[25]], ylim=c(-30,20), xlim=c(-90,-40), legend=F, main='L_2
prediction')
plot(eval1@predictions[['LQ_2']], ylim=c(-30,20), xlim=c(-90,-40), legend=F, main='LQ_1
prediction')

```

```

#Modelado con parametros óptimos elegidos
presencia,val,ambientales <- extract(variables,ambientales, registros,buddleja[,2:3],
cellnumbers=T)
presencia,val,ambientales <- na.omit(presencia,val,ambientales)
presencia,val,ambientales <- unique(presencia,val,ambientales)
presencia,val,ambientales <- presencia,val,ambientales[,-1]
bkg,val,ambientales <- sampleRandom(variables,ambientales, 10000, cells=T)
bkg,val,ambientales <- unique(bkg,val,ambientales)
bkg,val,ambientales <- bkg,val,ambientales[, -1]
mxnt,args <- c("autofeature=FALSE", "linear=TRUE", "quadratic=TRUE",
"product=FALSE", "hinge=FALSE", "threshold=FALSE", "betamultiplier=2,5")
env,values <- data.frame(rbind(presencia,val,ambientales, bkg,val,ambientales))
y <- c(rep(1,nrow(presencia,val,ambientales)), rep(0,nrow(bkg,val,ambientales)))
objetomaxentfi<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especiebuddleja/resultadofi")
predichofi<-predict(objetomaxentfi, variables,ambientales,args=c("outputformat=raw"))
predichofi <- rasterRescale(predichofi)
writeRaster(predichofi, filename="modelos/Especiebuddleja/prediccionfinal,tif",
format="GTiff" )

```

```

#Aplicación de umbral
extrayendo,valores<-extract(predichofi,registros,buddleja[,2:3])
umbralafi<- predichofi>=min(extrayendo,valores)
writeRaster(umbralafi, filename = "modelos/Especiebuddleja/umbralafi,tiff", format="GTiff")
plot(umbralafi)
points(registros,buddleja)
plot(points(registros,buddleja))

```

ANEXO 10

SCRIPT DESARROLLADO EN RSTUDIO PARA EL MODELADO DE *POLYLEPIS RACEMOSA*

```
options(java.parameters = "-Xmx1g" )
library(raster)
library(dismo)
library(ENMeval)

#Directorio de trabajo
setwd("C:/Users/Corina/Desktop/Distribucionespecies/")
source("funcion_reescalar_raster,R")

#Carga de registros
registros.polylepis<- read.csv("./coordenadas_especies/polylepis.csv")
registros.polylepis <- registros.polylepis[!duplicated(registros.polylepis),]
registros.polylepis <- registros.polylepis[,-1]

#Variables ambientales
variables.ambientales <-stack(list,files("variablesfinalesmaxent/" ,"*.tif$", full.names =
T))
plot(variables.ambientales)
bg <- randomPoints(variables.ambientales[[1]], n=10000)
bg <- as.data.frame(bg)

#Configuración de modelos
eval1 <- ENMevaluate(occ=registros.polylepis, env=variables.ambientales,
bg,coords=bg, method='randomkfold', kfolds = 4, RMvalues=c(0,5,1,1,5,2),
fc=c('L','Q','H','LQ','LH','QH','LQH'))

# Evaluadores de rendimiento y gráficos
str(eval1, max,level=3)
eval1@results
eval1@results[which(eval1@results$delta,AICc==0),]
eval1@predictions
plot(eval1@predictions[[which(eval1@results$delta,AICc==0)]], main="Relative
occurrence rate")
aic,opt <- eval1@models[[which(eval1@results$delta,AICc==0)]]
aic,opt@results
var,importance(aic,opt)
aic,opt@lambdas
eval,plot(eval1@results)
eval,plot(eval1@results, 'Mean,AUC')
df <- var,importance(aic,opt)
barplot(df$permutation,importance, names,arg=df$variable, las=2, ylab="Permutation
Importance")
```

```

#Ploteando las predicciones
par(mfrow=c(1,2), mar=c(2,2,1,0))
plot(eval1@predictions[[25]], ylim=c(-30,20), xlim=c(-90,-40), legend=F, main='L_2
prediction')
plot(eval1@predictions[['LQ_2']], ylim=c(-30,20), xlim=c(-90,-40), legend=F,
main='LQ_1 prediction')

```

```

#Modelado con parametros óptimos elegidos
presencia,val,ambientales <- extract(variables,ambientales, registros,polylepis[,2:3],
cellnumbers=T)
presencia,val,ambientales <- na.omit(presencia,val,ambientales)
presencia,val,ambientales <- unique(presencia,val,ambientales)
presencia,val,ambientales <- presencia,val,ambientales[,-1]
bkg,val,ambientales <- sampleRandom(variables,ambientales, 10000, cells=T)
bkg,val,ambientales <- unique(bkg,val,ambientales)
bkg,val,ambientales <- bkg,val,ambientales[, -1]
mxnt,args <- c("autofeature=FALSE", "linear=TRUE", "quadratic=TRUE",
"product=FALSE","hinge=FALSE", "threshold=FALSE", "betamultiplier=1")
env,values <- data.frame(rbind(presencia,val,ambientales, bkg,val,ambientales))
y <- c(rep(1,nrow(presencia,val,ambientales)), rep(0,nrow(bkg,val,ambientales)))
objetomaxentfi<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especiepolylepis/resultadofi")
predichofi<-predict(objetomaxentfi, variables,ambientales,args=c("outputformat=raw"))
predichofi <- rasterRescale(predichofi)
writeRaster(predichofi, filename="modelos/Especiepolylepis/prediccionfinal,tif",
format="GTiff" )

```

```

#Aplicación de umbral
extrayendo,valores<-extract(predichofi,registros,polylepis[,2:3])
umbralfi<- predichofi>=min(extrayendo,valores)
writeRaster(umbralfi, filename = "modelos/Especiepolylepis/umbralfi,tiff",
format="GTiff")

```

ANEXO 11

SCRIPT DESARROLLADO PARA EL MODELADO DE *CEDRELA ODORATA*

```
options(java,parameters="-Xmx1g")
library(raster)
library(dismo)
#Directorio de trabajo
setwd("C:/Users/Corina/Desktop/Distribucionespecies/")
source("funcion_reescalar_raster,R")

#Carga de registros
presencias,cedrela <- read.csv("coordenadas_especies/cedrelasiete.csv")

#variables ambientales
variables,ambientales <-stack(list,files("variablesfinalesmaxent/" ,"*.tif$", full, names =
T))
presencia,val,ambientales <- extract(variables,ambientales, presencias,cedrela[,2:3],
cellnumbers=T)
presencia,val,ambientales <- na.omit(presencia,val,ambientales)
presencia,val,ambientales <- unique(presencia,val,ambientales)
presencia,val,ambientales <- presencia,val,ambientales[,-1]
bkg,val,ambientales <- sampleRandom(variables,ambientales, 10000, cells=T)
bkg,val,ambientales <- unique(bkg,val,ambientales)
bkg,val,ambientales <- bkg,val,ambientales[, -1]

#registros con índices
fold<- kfold(presencia,val,ambientales, k=7)

#configuración de parámetros
mxnt,args <- c("autofeature=FALSE", "linear=TRUE", "quadratic=TRUE",
"product=FALSE","hinge=FALSE", "threshold=FALSE", "betamultiplier=1")

#modelado excluyendo registro 1
pres,test <- presencia,val,ambientales[fold==1,]
pres,train <- presencia,val,ambientales[fold!=1,]
env,values <- data.frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent1<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especiecedrelakuela7/resultado1"
)
predicho1<-predict(objetomaxent1, variables,ambientales,args=c("outputformat=raw"))
predicho1 <- rasterRescale(predicho1)
writeRaster(predicho1,
filename="modelos/Especiecedrelakuela7/predicciones/prediccion1.tif", format="GTiff"
)
extrayendo,valores<-extract(predicho1,presencias,cedrela[,2:3])
umbral1<- predicho1>= 0,66
```



```
writeRaster(umbral1, filename =
"modelos/Especiecedrelakuela7/predicciones/umbral1.tif", format="GTiff")
```

```
## modelado excluyendo registro 2
pres,test <- presencia,val,ambientales[fold==2,]
pres,train <- presencia,val,ambientales[fold!=2,]
env,values <- data,frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent2<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especiecedrelakuela7/resultado2"
)
predicho2<-predict(objetomaxent2, variables,ambientales,args=c("outputformat=raw"))
predicho2 <- rasterRescale(predicho2)
writeRaster(predicho2,
filename="modelos/Especiecedrelakuela7/predicciones/prediccion2.tif", format="GTiff"
)
extrayendo,valores<-extract(predicho2,presencias,cedrela[,2:3])
umbral2<- predicho2>=0,71
writeRaster(umbral2, filename =
"modelos/Especiecedrelakuela7/predicciones/umbral2.tif", format="GTiff")
```

```
## modelado excluyendo registro 3
pres,test <- presencia,val,ambientales[fold==3,]
pres,train <- presencia,val,ambientales[fold!=3,]
env,values <- data,frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent3<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especiecedrelakuela7/resultado3"
)
predicho3<-predict(objetomaxent3, variables,ambientales,args=c("outputformat=raw"))
predicho3 <- rasterRescale(predicho1)
writeRaster(predicho3,
filename="modelos/Especiecedrelakuela7/predicciones/prediccion3.tif", format="GTiff"
)
extrayendo,valores<-extract(predicho3,presencias,cedrela[,2:3])
umbral3<- predicho3>=0,55
writeRaster(umbral3, filename =
"modelos/Especiecedrelakuela7/predicciones/umbral3.tif", format="GTiff")
```

```
## modelado excluyendo registro 4
pres,test <- presencia,val,ambientales[fold==4,]
pres,train <- presencia,val,ambientales[fold!=4,]
env,values <- data,frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent4<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especiecedrelakuela7/resultado4"
)
```

```

predicho4<-predict(objetomaxent4, variables,ambientales,args=c("outputformat=raw"))
predicho4 <- rasterRescale(predicho4)
writeRaster(predicho4,
filename="modelos/Especiecedrelakuela7/predicciones/prediccion4.tif", format="GTiff"
)
extrayendo,valores<-extract(predicho4,presencias,cedrela[,2:3])
umbral4<- predicho4>=0,69
writeRaster(umbral4, filename =
"modelos/Especiecedrelakuela7/predicciones/umbral4.tif", format="GTiff")

## modelado excluyendo registro 5
pres,test <- presencia,val,ambientales[fold==5,]
pres,train <- presencia,val,ambientales[fold!=5,]
env,values <- data,frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent5<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especiecedrelakuela7/resultado5"
)
predicho5<-predict(objetomaxent5, variables,ambientales,args=c("outputformat=raw"))
predicho5 <- rasterRescale(predicho5)
writeRaster(predicho5,
filename="modelos/Especiecedrelakuela7/predicciones/prediccion5.tif", format="GTiff"
)
extrayendo,valores<-extract(predicho5,presencias,cedrela[,2:3])
umbral5<- predicho5>=0,72
writeRaster(umbral5, filename =
"modelos/Especiecedrelakuela7/predicciones/umbral5.tif", format="GTiff")

## modelado excluyendo registro 6
pres,test <- presencia,val,ambientales[fold==6,]
pres,train <- presencia,val,ambientales[fold!=6,]
env,values <- data,frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent6<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especiecedrelakuela7/resultado6"
)
predicho6<-predict(objetomaxent6, variables,ambientales,args=c("outputformat=raw"))
predicho6 <- rasterRescale(predicho6)
writeRaster(predicho6,
filename="modelos/Especiecedrelakuela7/predicciones/prediccion6.tif", format="GTiff"
)
extrayendo,valores<-extract(predicho6,presencias,cedrela[,2:3])
umbral6<- predicho6>=0,66
writeRaster(umbral6, filename =
"modelos/Especiecedrelakuela7/predicciones/umbral6.tif", format="GTiff")
plot(predicho6)

## modelado excluyendo registro 7
pres,test <- presencia,val,ambientales[fold==7,]

```

```

pres,train <- presencia,val,ambientales[fold!=7,]
env,values <- data.frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent7<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especiecedrelakuela7/resultado7"
)
predicho7<-predict(objetomaxent7, variables,ambientales,args=c("outputformat=raw"))
predicho7 <- rasterRescale(predicho7)
writeRaster(predicho7,
filename="modelos/Especiecedrelakuela7/predicciones/prediccion7,tif", format="GTiff"
)
extrayendo,valores<-extract(predicho7,presencias,cedrela[,2:3])
umbral7<- predicho7>=0,51
writeRaster(umbral7,
filename
"modelos/Especiecedrelakuela7/predicciones/umbral7,tif", format="GTiff")
=

##modelo final 7 registros
env,values <- data.frame(rbind(presencia,val,ambientales, bkg,val,ambientales))
y <- c(rep(1,nrow(presencia,val,ambientales)), rep(0,nrow(bkg,val,ambientales)))
objetomaxentfi<- maxent(env,values,y,args=mxnt,args)
predichofi<-predict(objetomaxentfi, variables,ambientales,args=c("outputformat=raw"))
predichofi <- rasterRescale(predichofi)
writeRaster(predichofi,
filename="modelos/Especiecedrelakuela7/predicciones/prediccionfi,tif", format="GTiff"
)
extrayendo,valores<-extract(predichofi,presencias,cedrela[,2:3])
umbral7fi<- predichofi>=0,68
writeRaster(umbral7fi,
filename
"modelos/Especiecedrelakuela7/prediccionfinal/umbral7fi,tif", format="GTiff")
=

```

ANEXO 12

SCRIPT DESARROLLADO EN RSTUDIO PARA EL MODELADO DE *TECOMA ROSAEFOLIA*

```
options(java,parameters="-Xmx1g")
library(raster)
library(dismo)

#Directorio de trabajo
setwd("C:/Users/Corina/Desktop/Distribucionespecies/")
source("funcion_reescalar_raster,R")

#Carga de registros
presencias,tecoma <- read.csv("coordenadas_especies/tecomarosaeolia.csv")

#variables ambientales
variables,ambientales <- stack(list,files("variablesfinalesmaxent/", "*tif$", full.names =
T))
presencia,val,ambientales <- extract(variables,ambientales, presencias,tecoma[,2:3],
cellnumbers=T)
presencia,val,ambientales <- na.omit(presencia,val,ambientales)
presencia,val,ambientales <- unique(presencia,val,ambientales)
presencia,val,ambientales <- presencia,val,ambientales[,-1]
bkg,val,ambientales <- sampleRandom(variables,ambientales, 10000, cells=T)
bkg,val,ambientales <- unique(bkg,val,ambientales)
bkg,val,ambientales <- bkg,val,ambientales[, -1]

#registros con índices
fold<- kfold(presencia,val,ambientales, k=17)

#configuración de parámetros
mxnt,args <- c("autofeature=FALSE", "linear=TRUE", "quadratic=TRUE",
"product=FALSE", "hinge=FALSE", "threshold=FALSE", "betamultiplier=1")

#modelado excluyendo registro 1
pres,test <- presencia,val,ambientales[fold==1,]
pres,train <- presencia,val,ambientales[fold!=1,]
env,values <- data.frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent1<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especietecoma/resultado1")
predicho1<-predict(objetomaxent1, variables,ambientales,args=c("outputformat=raw"))
predicho1 <- rasterRescale(predicho1)
writeRaster(predicho1,
filename="modelos/Especietecoma/predicciones/prediccion1.tif", format="GTiff" )
extrayendo,valores<-extract(predicho1,presencias,tecoma[,2:3])
umbral1<- predicho1>= 0,52
```

```
writeRaster(umbral1, filename = "modelos/Especietecoma/predicciones/umbral1.tif",
format="GTiff")
```

```
#modelado excluyendo registro 2
```

```
pres,test <- presencia,val,ambientales[fold==2,]
pres,train <- presencia,val,ambientales[fold!=2,]
env,values <- data,frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent2<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especietecoma/resultado2")
predicho2<-predict(objetomaxent2, variables,ambientales,args=c("outputformat=raw"))
predicho2 <- rasterRescale(predicho2)
writeRaster(predicho2,
filename="modelos/Especietecoma/predicciones/prediccion2.tif", format="GTiff" )
extrayendo,valores<-extract(predicho2,presencias,tecoma[,2:3])
umbral2<- predicho2>= 0,55
writeRaster(umbral2, filename = "modelos/Especietecoma/predicciones/umbral2.tif",
format="GTiff")
```

```
#modelado excluyendo registro 3
```

```
pres,test <- presencia,val,ambientales[fold==3,]
pres,train <- presencia,val,ambientales[fold!=3,]
env,values <- data,frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent3<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especietecoma/resultado3")
predicho3<-predict(objetomaxent3, variables,ambientales,args=c("outputformat=raw"))
predicho3<- rasterRescale(predicho3)
writeRaster(predicho3,
filename="modelos/Especietecoma/predicciones/prediccion3.tif", format="GTiff" )
extrayendo,valores<-extract(predicho3,presencias,tecoma[,2:3])
umbral3<- predicho3>= 0,61
writeRaster(umbral3, filename = "modelos/Especietecoma/predicciones/umbral3.tif",
format="GTiff")
```

```
#modelado excluyendo registro 4
```

```
pres,test <- presencia,val,ambientales[fold==4,]
pres,train <- presencia,val,ambientales[fold!=4,]
env,values <- data,frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent4<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especietecoma/resultado4")
predicho4<-predict(objetomaxent4, variables,ambientales,args=c("outputformat=raw"))
predicho4 <- rasterRescale(predicho4)
writeRaster(predicho4,
filename="modelos/Especietecoma/predicciones/prediccion4.tif", format="GTiff" )
extrayendo,valores<-extract(predicho4,presencias,tecoma[,2:3])
umbral4<- predicho4>= 0,53
```

```
writeRaster(umbral4, filename = "modelos/Especietecoma/predicciones/umbral4.tif",
format="GTiff")
```

```
#modelado excluyendo registro 5
pres,test <- presencia,val,ambientales[fold==5,]
pres,train <- presencia,val,ambientales[fold!=5,]
env,values <- data,frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent5<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especietecoma/resultado5")
predicho5 <-predict(objetomaxent5, variables,ambientales,args=c("outputformat=raw"))
predicho5 <- rasterRescale(predicho5)
writeRaster(predicho5,
filename="modelos/Especietecoma/predicciones/prediccion5.tif", format="GTiff" )
extrayendo,valores<-extract(predicho5,presencias,tecoma[,2:3])
umbral5<- predicho5>= 0,71
writeRaster(umbral5, filename = "modelos/Especietecoma/predicciones/umbral5.tif",
format="GTiff")
```

```
#modelado excluyendo registro 6
pres,test <- presencia,val,ambientales[fold==6,]
pres,train <- presencia,val,ambientales[fold!=6,]
env,values <- data,frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent6<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especietecoma/resultado6")
predicho6<-predict(objetomaxent6, variables,ambientales,args=c("outputformat=raw"))
predicho6 <- rasterRescale(predicho6)
writeRaster(predicho6,
filename="modelos/Especietecoma/predicciones/prediccion6.tif", format="GTiff" )
extrayendo,valores<-extract(predicho6,presencias,tecoma[,2:3])
umbral6<- predicho6>= 0,61
writeRaster(umbral6, filename = "modelos/Especietecoma/predicciones/umbral6.tif",
format="GTiff")
```

```
#modelado excluyendo registro 7
pres,test <- presencia,val,ambientales[fold==7,]
pres,train <- presencia,val,ambientales[fold!=7,]
env,values <- data,frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent7<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especietecoma/resultado7")
predicho7<-predict(objetomaxent7, variables,ambientales,args=c("outputformat=raw"))
predicho7 <- rasterRescale(predicho7)
writeRaster(predicho7,
filename="modelos/Especietecoma/predicciones/prediccion7.tif", format="GTiff" )
extrayendo,valores<-extract(predicho7,presencias,tecoma[,2:3])
umbral7<- predicho7>= 0,52
writeRaster(umbral1, filename = "modelos/Especietecoma/predicciones/umbral7.tif",
format="GTiff")
```

```

#modelado excluyendo registro 8
pres,test <- presencia,val,ambientales[fold==8,]
pres,train <- presencia,val,ambientales[fold!=8,]
env,values <- data,frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent8<-
maxent(env,values,y,args=mxnt,args,path="./modelos/Especietecoma/resultado8")
predicho8<-predict(objetomaxent8, variables,ambientales,args=c("outputformat=raw"))
predicho8<- rasterRescale(predicho8)
writeRaster(predicho8,
filename="modelos/Especietecoma/predicciones/prediccion8.tif", format="GTiff" )
extrayendo,valores<-extract(predicho8,presencias,tecoma[,2:3])
umbral8<- predicho8>= 0,55
writeRaster(umbral1, filename = "modelos/Especietecoma/predicciones/umbral8.tif",
format="GTiff")

```

```

#modelado excluyendo registro 9
pres,test <- presencia,val,ambientales[fold==9,]
pres,train <- presencia,val,ambientales[fold!=9,]
env,values <- data,frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent9<-
maxent(env,values,y,args=mxnt,args,path="./modelos/Especietecoma/resultado9")
predicho9<-predict(objetomaxent9, variables,ambientales,args=c("outputformat=raw"))
predicho9 <- rasterRescale(predicho9)
writeRaster(predicho9,
filename="modelos/Especietecoma/predicciones/prediccion9.tif", format="GTiff" )
extrayendo,valores<-extract(predicho9,presencias,tecoma[,2:3])
umbral9<- predicho9>= 0,54
writeRaster(umbral9, filename = "modelos/Especietecoma/predicciones/umbral9.tif",
format="GTiff")
plot(umbral9)

```

```

#modelado excluyendo registro 10
pres,test <- presencia,val,ambientales[fold==10,]
pres,train <- presencia,val,ambientales[fold!=10,]
env,values <- data,frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent10<-
maxent(env,values,y,args=mxnt,args,path="./modelos/Especietecoma/resultado10")
predicho10<-predict(objetomaxent10,
variables,ambientales,args=c("outputformat=raw"))
predicho10 <- rasterRescale(predicho10)
writeRaster(predicho10,
filename="modelos/Especietecoma/predicciones/prediccion10.tif", format="GTiff" )
extrayendo,valores<-extract(predicho10,presencias,tecoma[,2:3])
umbral10<- predicho10>= 0,57

```

```
writeRaster(umbral10, filename = "modelos/Especietecoma/predicciones/umbral10.tif",
format="GTiff")
```

```
#modelado excluyendo registro 11
pres,test <- presencia,val,ambientales[fold==11,]
pres,train <- presencia,val,ambientales[fold!=11,]
env,values <- data,frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent11<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especietecoma/resultado11")
predicho11<-predict(objetomaxent11,
variables,ambientales,args=c("outputformat=raw"))
predicho11 <- rasterRescale(predicho11)
writeRaster(predicho11,
filename="modelos/Especietecoma/predicciones/prediccion11.tif", format="GTiff" )
extrayendo,valores<-extract(predicho11,presencias,tecoma[,2:3])
umbral11<- predicho11>= 0,56
writeRaster(umbral11, filename = "modelos/Especietecoma/predicciones/umbral11.tif",
format="GTiff")
```

```
#modelado excluyendo registro 12
pres,test <- presencia,val,ambientales[fold==12,]
pres,train <- presencia,val,ambientales[fold!=12,]
env,values <- data,frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent12<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especietecoma/resultado12")
predicho12<-predict(objetomaxent12,
variables,ambientales,args=c("outputformat=raw"))
predicho12 <- rasterRescale(predicho12)
writeRaster(predicho12,
filename="modelos/Especietecoma/predicciones/prediccion12.tif", format="GTiff" )
extrayendo,valores<-extract(predicho12,presencias,tecoma[,2:3])
umbral12<- predicho12>= 0,59
writeRaster(umbral12, filename = "modelos/Especietecoma/predicciones/umbral12.tif",
format="GTiff")
```

```
#modelado excluyendo registro 13
pres,test <- presencia,val,ambientales[fold==13,]
pres,train <- presencia,val,ambientales[fold!=13,]
env,values <- data,frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent13<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especietecoma/resultado13")
predicho13<-predict(objetomaxent13,
variables,ambientales,args=c("outputformat=raw"))
predicho13 <- rasterRescale(predicho13)
writeRaster(predicho13,
filename="modelos/Especietecoma/predicciones/prediccion13.tif", format="GTiff" )
extrayendo,valores<-extract(predicho13,presencias,tecoma[,2:3])
```



```
umbral13<- predicho13>= 0,59
writeRaster(umbral13, filename = "modelos/Especietecoma/predicciones/umbral13.tif",
format="GTiff")
```

```
#modelado excluyendo registro 14
pres,test <- presencia,val,ambientales[fold==14,]
pres,train <- presencia,val,ambientales[fold!=14,]
env,values <- data,frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent14<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especietecoma/resultado14")
predicho14<-predict(objetomaxent14,
variables,ambientales,args=c("outputformat=raw"))
predicho14 <- rasterRescale(predicho14)
writeRaster(predicho14,
filename="modelos/Especietecoma/predicciones/prediccion14.tif", format="GTiff" )
extrayendo,valores<-extract(predicho14,presencias,tecoma[,2:3])
umbral14<- predicho14>= 0,61
writeRaster(umbral14, filename = "modelos/Especietecoma/predicciones/umbral14.tif",
format="GTiff")
```

```
#modelado excluyendo registro 15
pres,test <- presencia,val,ambientales[fold==15,]
pres,train <- presencia,val,ambientales[fold!=15,]
env,values <- data,frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent15<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especietecoma/resultado15")
predicho15<-predict(objetomaxent15,
variables,ambientales,args=c("outputformat=raw"))
predicho15 <- rasterRescale(predicho15)
writeRaster(predicho15,
filename="modelos/Especietecoma/predicciones/prediccion15.tif", format="GTiff" )
extrayendo,valores<-extract(predicho15,presencias,tecoma[,2:3])
umbral15<- predicho15>= 0,54
writeRaster(umbral15, filename = "modelos/Especietecoma/predicciones/umbral15.tif",
format="GTiff")
```

```
#modelado excluyendo registro 16
pres,test <- presencia,val,ambientales[fold==16,]
pres,train <- presencia,val,ambientales[fold!=16,]
env,values <- data,frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent16<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especietecoma/resultado16")
predicho16<-predict(objetomaxent16,
variables,ambientales,args=c("outputformat=raw"))
predicho16 <- rasterRescale(predicho16)
writeRaster(predicho16,
filename="modelos/Especietecoma/predicciones/prediccion16.tif", format="GTiff" )
```

```

extrayendo,valores<-extract(predicho16,presencias,tecoma[,2:3])
umbral16<- predicho16>= 0,51
writeRaster(umbral16, filename = "modelos/Especietecoma/predicciones/umbral16,tif",
format="GTiff")

```

```

#modelado excluyendo registro 17
pres,test <- presencia,val,ambientales[fold==17,]
pres,train <- presencia,val,ambientales[fold!=17,]
env,values <- data,frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent17<-
maxent(env,values,y,args=mxnt,args,path="./modelos/Especietecoma/resultado17")
predicho17<-predict(objetomaxent17,
variables,ambientales,args=c("outputformat=raw"))
predicho17 <- rasterRescale(predicho17)
writeRaster(predicho17,
filename="modelos/Especietecoma/predicciones/prediccion17,tif", format="GTiff")
extrayendo,valores<-extract(predicho17,presencias,tecoma[,2:3])
umbral17<- predicho17>= 0,63
writeRaster(umbral17, filename = "modelos/Especietecoma/predicciones/umbral17,tif",
format="GTiff")

```

```

#modelo final con 17 registros
env,values <- data,frame(rbind(presencia,val,ambientales, bkg,val,ambientales))
y <- c(rep(1,nrow(presencia,val,ambientales)), rep(0,nrow(bkg,val,ambientales)))
objetomaxentfi<-
maxent(env,values,y,args=mxnt,args,path="./modelos/Especietecoma/resultadofi")
predichofi<-predict(objetomaxentfi, variables,ambientales,args=c("outputformat=raw"))
predichofi <- rasterRescale(predichofi)
writeRaster(predichofi,
filename="modelos/Especietecoma/predicciones/prediccionfi,tif", format="GTiff" )
extrayendo,valores<-extract(predichofi,presencias,tecoma[,2:3])
umbralfi<- predichofi>=0,51
writeRaster(umbralfi, filename = "modelos/Especietecoma/prediccionfinal/umbralfi,tiff",
format="GTiff")

```

ANEXO 13

SCRIPT DESARROLLADO EN RSTUDIO PARA EL MODELADO DE *ESENBECKIA CORNUTA*

```
options(java,parameters="-Xmx1g")
library(raster)
library(dismo)
#Directorio de trabajo
setwd("C:/Users/Corina/Desktop/Distribucionespecies/")
source("funcion_reescalar_raster,R")

#Carga de registros
presencias,esenbeckia <- read.csv("coordenadas_especies/Esenbeckia.csv")

#variables ambientales
variables,ambientales <- stack(list,files("variablesfinalesmaxent/" ,"*",tif$", full, names =
T))
presencia,val,ambientales <- extract(variables,ambientales, presencias,esenbeckia[,2:3],
cellnumbers=T)
presencia,val,ambientales <- na.omit(presencia,val,ambientales)
presencia,val,ambientales <- unique(presencia,val,ambientales)
presencia,val,ambientales <- presencia,val,ambientales[,-1]
bkg,val,ambientales <- sampleRandom(variables,ambientales, 10000, cells=T)
bkg,val,ambientales <- unique(bkg,val,ambientales)
bkg,val,ambientales <- bkg,val,ambientales[, -1]

#registros con índices
fold<- kfold(presencia,val,ambientales, k=11)

#configuración de parámetros
mxnt,args <- c("autofeature=FALSE", "linear=TRUE", "quadratic=TRUE",
"product=FALSE","hinge=FALSE", "threshold=FALSE", "betamultiplier=1")

#modelado excluyendo registro 1
pres,test <- presencia,val,ambientales[fold==1,]
pres,train <- presencia,val,ambientales[fold!=1,]
env,values <- data.frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent1<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especieesenbeckia/resultado1")
predicho1<-predict(objetomaxent1, variables,ambientales,args=c("outputformat=raw"))
predicho1 <- rasterRescale(predicho1)
writeRaster(predicho1,
filename="modelos/Especieesenbeckia/predicciones/prediccion1.tif", format="GTiff" )
extrayendo,valores1<-extract(predicho1,presencias,esenbeckia[,2:3])
umbral1<- predicho1>= 0,77
writeRaster(umbral1, filename = "modelos/Especieesenbeckia/predicciones/umbral1.tif",
format="GTiff")
```

```

#modelado excluyendo registro 2
pres,test <- presencia,val,ambientales[fold==2,]
pres,train <- presencia,val,ambientales[fold!=2,]
env,values <- data,frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent2<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especieesenbeckia/resultado2")
predicho2<-predict(objetomaxent2, variables,ambientales,args=c("outputformat=raw"))
predicho2 <- rasterRescale(predicho2)
writeRaster(predicho2,
filename="modelos/Especieesenbeckia/predicciones/prediccion2.tif", format="GTiff" )
extrayendo,valores2<-extract(predicho2,presencias,esenbeckia[,2:3])
umbral2<- predicho2>= 0,77
writeRaster(umbral2, filename = "modelos/Especieesenbeckia/predicciones/umbral2.tif",
format="GTiff")

```

```

#modelado excluyendo registro 3
pres,test <- presencia,val,ambientales[fold==3,]
pres,train <- presencia,val,ambientales[fold!=3,]
env,values <- data,frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent3<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especieesenbeckia/resultado3")
predicho3<-predict(objetomaxent3, variables,ambientales,args=c("outputformat=raw"))
predicho3 <- rasterRescale(predicho3)
writeRaster(predicho3,
filename="modelos/Especieesenbeckia/predicciones/prediccion3.tif", format="GTiff" )
extrayendo,valores3<-extract(predicho3,presencias,esenbeckia[,2:3])
umbral3<- predicho3>= 0,78
writeRaster(umbral3, filename = "modelos/Especieesenbeckia/predicciones/umbral3.tif",
format="GTiff")

```

```

#modelado excluyendo registro 4
pres,test <- presencia,val,ambientales[fold==4,]
pres,train <- presencia,val,ambientales[fold!=4,]
env,values <- data,frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent4<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especieesenbeckia/resultado4")
predicho4<-predict(objetomaxent4, variables,ambientales,args=c("outputformat=raw"))
predicho4 <- rasterRescale(predicho4)
writeRaster(predicho4,
filename="modelos/Especieesenbeckia/predicciones/prediccion4.tif", format="GTiff" )
extrayendo,valores4<-extract(predicho4,presencias,esenbeckia[,2:3])
umbral4<- predicho4>= 0,77

```

```
writeRaster(umbral4, filename = "modelos/Especieesenbeckia/predicciones/umbral4.tif",
format="GTiff",overwrite=TRUE)
```

```
#modelado excluyendo registro 5
pres,test <- presencia,val,ambientales[fold==5,]
pres,train <- presencia,val,ambientales[fold!=5,]
env,values <- data,frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent5<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especieesenbeckia/resultado5")
predicho5<-predict(objetomaxent5, variables,ambientales, args=c("outputformat=raw"))
predicho5 <- rasterRescale(predicho5)
writeRaster(predicho5,
filename="modelos/Especieesenbeckia/predicciones/prediccion5.tif", format="GTiff" )
extrayendo,valores5<-extract(predicho5,presencias,esenbeckia[,2:3])
umbral5<- predicho5>= 0,79
writeRaster(umbral5, filename = "modelos/Especieesenbeckia/predicciones/umbral5.tif",
format="GTiff")
plot(umbral5)
```

```
#modelado excluyendo registro 6
pres,test <- presencia,val,ambientales[fold==6,]
pres,train <- presencia,val,ambientales[fold!=6,]
env,values <- data,frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent6<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especieesenbeckia/resultado6")
predicho6<-predict(objetomaxent6, variables,ambientales,args=c("outputformat=raw"))
predicho6 <- rasterRescale(predicho6)
writeRaster(predicho6,
filename="modelos/Especieesenbeckia/predicciones/prediccion6.tif", format="GTiff" )
extrayendo,valores<-extract(predicho6,presencias,esenbeckia[,2:3])
umbral6<- predicho6>= 0,79
writeRaster(umbral6, filename = "modelos/Especieesenbeckia/predicciones/umbral6.tif",
format="GTiff")
```

```
#modelado excluyendo registro 7
pres,test <- presencia,val,ambientales[fold==7,]
pres,train <- presencia,val,ambientales[fold!=7,]
env,values <- data,frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent7<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especieesenbeckia/resultado7")
predicho7<-predict(objetomaxent7, variables,ambientales,args=c("outputformat=raw"))
predicho7 <- rasterRescale(predicho7)
```

```

writeRaster(predicho7,
filename="modelos/Especieesenbeckia/predicciones/prediccion7.tif", format="GTiff" )
extrayendo, valores<-extract(predicho7,presencias,esenbeckia[,2:3])
umbral7<- predicho7>= 0,77
writeRaster(umbral7, filename = "modelos/Especieesenbeckia/predicciones/umbral7.tif",
format="GTiff")

```

```

#modelado excluyendo registro 8
pres,test <- presencia,val,ambientales[fold==8,]
pres,train <- presencia,val,ambientales[fold!=8,]
env,values <- data.frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent8<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especieesenbeckia/resultado8")
predicho8<-predict(objetomaxent8, variables,ambientales,args=c("outputformat=raw"))
predicho8 <- rasterRescale(predicho8)
writeRaster(predicho8,
filename="modelos/Especieesenbeckia/predicciones/prediccion8.tif", format="GTiff" )
extrayendo, valores<-extract(predicho8,presencias,esenbeckia[,2:3])
umbral8<- predicho8>= 0,78
writeRaster(umbral8, filename = "modelos/Especieesenbeckia/predicciones/umbral8.tif",
format="GTiff")

```

```

#modelado excluyendo registro 9
pres,test <- presencia,val,ambientales[fold==9,]
pres,train <- presencia,val,ambientales[fold!=9,]
env,values <- data.frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent9<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especieesenbeckia/resultado9")
predicho9<-predict(objetomaxent9, variables,ambientales,args=c("outputformat=raw"))
predicho9 <- rasterRescale(predicho9)
writeRaster(predicho9,
filename="modelos/Especieesenbeckia/predicciones/prediccion9.tif", format="GTiff" )
extrayendo, valores<-extract(predicho9,presencias,esenbeckia[,2:3])
umbral9<- predicho9>= 0,78
writeRaster(umbral9, filename = "modelos/Especieesenbeckia/predicciones/umbral9.tif",
format="GTiff")

```

```

#modelado excluyendo registro 10
pres,test <- presencia,val,ambientales[fold==10,]
pres,train <- presencia,val,ambientales[fold!=10,]
env,values <- data.frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent10<-
maxent(env,values,y,args=mxnt,args,path="/modelos/Especieesenbeckia/resultado10")

```

```

predicho10<-predict(objetomaxent10,
variables,ambientales,args=c("outputformat=raw"))
predicho10<- rasterRescale(predicho10)
writeRaster(predicho10,
filename="modelos/Especieesenbeckia/predicciones/prediccion10.tif", format="GTiff" )
extrayendo,valores<-extract(predicho10,presencias,esenbeckia[,2:3])
umbral10<- predicho10>= 0,79
writeRaster(umbral10, filename =
"modelos/Especieesenbeckia/predicciones/umbral10.tif", format="GTiff")

```

```

#modelado excluyendo registro 11
pres,test <- presencia,val,ambientales[fold==11,]
pres,train <- presencia,val,ambientales[fold!=11,]
env,values <- data,frame(rbind(pres,train, bkg,val,ambientales))
y <- c(rep(1,nrow(pres,train)), rep(0,nrow(bkg,val,ambientales)))
objetomaxent11<-
maxent(env,values,y,args=mxnt,args,path="./modelos/Especieesenbeckia/resultado11")
predicho11<-predict(objetomaxent11,
variables,ambientales,args=c("outputformat=raw"))
predicho11 <- rasterRescale(predicho11)
writeRaster(predicho11,
filename="modelos/Especieesenbeckia/predicciones/prediccion11.tif", format="GTiff" )
extrayendo,valores<-extract(predicho11,presencias,esenbeckia[,2:3])
umbral11<- predicho11>= 0,77
writeRaster(umbral11, filename =
"modelos/Especieesenbeckia/predicciones/umbral11.tif", format="GTiff")

```

```

#modelo final 11 registros
env,values <- data,frame(rbind(presencia,val,ambientales, bkg,val,ambientales))
y <- c(rep(1,nrow(presencia,val,ambientales)), rep(0,nrow(bkg,val,ambientales)))
objetomaxentfi<-
maxent(env,values,y,args=mxnt,args,path="./modelos/Especieesenbeckia/resultadofi")
predichofi<-predict(objetomaxentfi, variables,ambientales,args=c("outputformat=raw"))
predichofi<- rasterRescale(predichofi)
writeRaster(predichofi,
filename="modelos/Especieesenbeckia/predicciones/prediccionfi.tif", format="GTiff" )
extrayendo,valores<-extract(predichofi,presencias,esenbeckia[,2:3])
umbral1fi<- predichofi>=0,77
writeRaster(umbral1fi, filename =
"modelos/Especieesenbeckia/prediccionfinal/umbral1fi.tiff", format="GTiff")

```