



**DEPARTAMENTO DE CIENCIAS DE LA TIERRA Y LA
CONSTRUCCIÓN**

**CARRERA DE INGENIERÍA GEOGRÁFICA Y DEL MEDIO
AMBIENTE**

**TRABAJO DE TITULACIÓN, PREVIO A LA OBTENCIÓN DEL
TÍTULO DE INGENIERO GEÓGRAFO Y DEL MEDIO AMBIENTE**

**TEMA: MODELAMIENTO DEL NICHOS ECOLÓGICO DE
ANOPHELES SPP. EN EL ECUADOR MEDIANTE
HERRAMIENTAS GEO-INFORMÁTICAS.**

AUTOR: ROSAS MONTALVO, PABLO FRANCISCO

DIRECTOR: ING. PADILLA ALMEIDA, OSWALDO VINICIO

SANGOLQUÍ

2016



DEPARTAMENTO DE CIENCIAS DE LA TIERRA Y LA CONSTRUCCIÓN
CARRERA DE INGENIERÍA GEOGRÁFICA Y DEL MEDIO AMBIENTE

CERTIFICACIÓN

Certifico que el trabajo de titulación, "Modelamiento de nicho ecológico de *Anopheles spp* en el Ecuador mediante herramientas geo-informáticas" realizado por el señor Pablo Francisco Rosas Montalvo, ha sido revisado en su totalidad, el mismo cumple con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de Fuerzas Armadas ESPE, por lo tanto me permito acreditarlo y autorizar al señor Pablo Francisco Rosas Montalvo para que continúe con los tramites respectivos para su graduación.

Sangolquí, 29 de enero del 2016

Atentamente,

Ing. Oswaldo Padilla Almeida

Director del proyecto de investigación

Recibido
29-01-2016



DEPARTAMENTO DE CIENCIAS DE LA TIERRA Y LA CONSTRUCCIÓN
CARRERA DE INGENIERÍA GEOGRÁFICA Y DEL MEDIO AMBIENTE

AUTORIA DE RESPONSABILIDAD

Yo, Pablo Francisco Rosas Montalvo, con cédula de identidad N° 171882279-2, declaro que este trabajo de titulación “Modelamiento del nicho ecológico de *Anopheles spp.* en el Ecuador mediante herramientas geo-informáticas” ha sido desarrollado considerando los métodos de investigación existentes, así como también se ha respetado los derechos intelectuales de terceros considerándose en las citas bibliográficas.

Consecuentemente declaro que este trabajo es de mi autoría, en virtud de ello me declaro responsable del contenido, veracidad y alcance de la investigación mencionada

Sangolquí, 10 de febrero del 2016

Pablo Francisco Rosas Montalvo

C.C: 171882279-2



DEPARTAMENTO DE CIENCIAS DE LA TIERRA Y LA CONSTRUCCIÓN
CARRERA DE INGENIERÍA GEOGRÁFICA Y DEL MEDIO AMBIENTE

AUTORIZACIÓN

Yo, Pablo Francisco Rosas Montalvo, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar en la biblioteca Virtual de la institución la presente trabajo de titulación “Modelamiento del nicho ecológico de *Anopheles spp.* en el Ecuador mediante herramientas geo-informáticas” cuyo contenido, ideas y criterios son de mi autoría y responsabilidad

Sangolquí, 10 de febrero del 2016

A handwritten signature in blue ink, consisting of stylized initials 'PRM' enclosed within a large, circular flourish.

Pablo Francisco Rosas Montalvo

C.C: 171882279-2

DEDICATORIA

A mi familia fuente de apoyo constante e incondicional durante toda mi vida y en especial a mi madre por todo el sacrificio y entrega para sacar adelante a la familia siempre.

Pablo R.

AGREDECIMIENTOS

En primer lugar, agradezco al Instituto Espacial Ecuatoriano por brindarme todo el apoyo necesario para la ejecución del presente proyecto.

A mis profesores de la Universidad de las Fuerzas Armadas-ESPE , especialmente a mi director Ing. Oswaldo Padilla por toda su aporte técnico- científico y su calidez humana a lo largo de mi carrera universitaria y durante el desarrollo del trabajo de investigación.

A mi amigo y compañero de labores Ing. Mauricio Moreno, que con todo su conocimiento impartido aportó de sobremanera para el éxito de la investigación.

A Lina, mi compañera de vida, por ser mi soporte en momentos de flaqueza y mi luz en medio de la oscuridad.

A mis amigos y amigas por todas las vivencias compartidas a lo largo de este camino; me llevo muchas amistades valiosas que perduraran en el tiempo y una maleta cargada de gratos recuerdos.

Finalmente, todo esto no hubiera sido posible sin el apoyo de mis padres y de toda mi familia, fuente de inspiración, motivación y amor. Que con su ejemplo de esfuerzo, sacrificio y perseverancia me han enseñado grandes valores y han enriquecido mi alma con una palabra de apoyo o un consejo y porque no con un regaño.

A todos ustedes mi mayor reconocimiento y gratitud.

ÍNDICE

1	Introducción	1
1.1	Antecedentes	1
1.2	Identificación del problema.....	2
1.3	Justificación.....	2
1.4	Objetivo General.	5
1.5	Objetivos Específicos.....	5
1.6	Metas.	5
1.7	Área de Estudio.	5
2	Marco Teórico	7
2.1	Definición de nicho ecológico.....	7
2.1.1	Definición de Joseph Grinnell.....	7
2.1.2	Definición de Charles Elton	9
2.1.3	Definición de G. Evelyn Hutchinson	9
2.2	Conceptos relevantes sobre el nicho ecológico.....	11
2.2.1	Tipos de nicho ecológico.....	11
2.2.2	Diagrama BAM.....	11
2.2.3	Relación entre el espacio geográfico y espacio ecológico	13
2.3	Modelos de Nicho Ecológico (MNE).....	15
2.3.1	Proceso para el modelamiento de nichos ecológico.....	17
2.3.1.1	Datos biológicos de la especie	21
2.3.1.2	Datos medio ambientales	23
2.3.1.3	Métodos para el modelamiento de nichos ecológicos.....	27
2.3.1.3.1	Maxent.....	29
2.3.1.3.2	Modelo de regresión logística.	39
2.3.1.3.3	Evaluación multi-criterio (EMC) y lógica difusa (<i>Fuzzy</i>).....	42
2.3.1.4	Validación de los modelos de nicho ecológico.	47
2.3.1.4.1	Curva ROC (Receiver operating characteristic analysis).....	50
2.4	Ecología de <i>Anopheles spp.</i>	52
3	Metodología y Desarrollo.....	56

3.1	Esquema metodológico	56
3.2	Estructuración de la información	57
3.2.1	Puntos de presencia de <i>Anopheles spp.</i>	57
3.2.2	Variables Predictoras.....	59
3.3	Aplicación de los modelos de nicho ecológico	64
3.3.1	Maxent.....	64
3.3.1.1	Ajuste y calibración.....	64
3.3.1.2	Resultados	66
3.3.2	Regresión logística.	69
3.3.2.1	Ajuste y calibración.....	70
3.3.2.2	Resultados	74
3.3.3	Evaluación multicriterio con lógica difusa.....	76
3.3.3.1	Ajuste y calibración.....	76
3.3.3.2	Resultados	80
4	Análisis y discusión.....	81
4.1	Análisis estadístico de los modelos.	81
4.1.1	Análisis por desviación estándar	81
4.1.2	Análisis por rangos de probabilidad.....	81
4.1.3	Análisis por la Curva ROC.	87
4.2	Selección del mejor modelo y delimitación de la zonas idóneas	89
4.3	Caracterización paisajística	90
4.3.1	Uso de suelo de la provincia de Esmeraldas	91
4.3.2	Zonas de Vida de la provincia de Esmeraldas.....	92
4.3.3	Factores Climáticos de la provincia de Esmeraldas.....	94
5	Conclusiones y Recomendaciones	96
5.1	Conclusiones.	96
5.2	Recomendaciones.....	97
6	Anexos.....	99
6.1	Código en lenguaje R para regresión logística.....	99
6.2	Mapa de distribución de <i>Anopheles spp.</i> – Evaluación Multicriterio	99

6.3	Mapa de distribución de <i>Anopheles spp.</i> – Regresión logística	99
6.4	Mapa de distribución de <i>Anopheles spp.</i> – Maxent.....	99
6.5	Mapa de Uso de Suelo de la provincia de Esmeraldas.....	99
6.6	Mapa de Zonas de Vida de la provincia de Esmeraldas.....	99
6.7	Mapa de Temperatura Media Anual de la provincia de Esmeraldas.....	99
6.8	Mapa de Precipitación Media Anual de la provincia de Esmeraldas.....	99
7	Bibliografía.....	100

ÍNDICE DE TABLAS

Tabla 1. Fuentes de información de datos biológicos	22
Tabla 2. Fuentes de datos medioambientales usados para la modelación de nichos.....	26
Tabla 3. Algoritmos para el modelamiento de nichos ecológicos.....	28
Tabla 4. Parámetros estadísticos derivados de la matriz de confusión.	50
Tabla 5. Especies del genero Anophles para el modelamiento de nicho ecológico.	58
Tabla 6. Variables descriptivas para el modelamiento de Anopheles spp.	61
Tabla 7. Matriz de correlación de variables	62
Tabla 8. Variables empleadas para el modelamiento de Anopheles spp.....	63
Tabla 9. Ponderación de la cobertura Uso de suelo	73
Tabla 10. Variables y sus respectivas funciones fuzzy.	78
Tabla 11. Análisis de Jerarquías Analíticas.....	79
Tabla 12. Asignación de pesos para las variables.	80
Tabla 13. Análisis por desviación estándar	81
Tabla 14. Análisis por rangos de probabilidad - EMC.....	83
Tabla 15. Análisis por rangos de probabilidad – Regresión Logística.....	84
Tabla 16. Análisis por rangos de probabilidad – Maxent.....	86

ÍNDICE DE FIGURAS

Figura 1. Pobreza por NBI nacional, urbano y rural.	3
Figura 2. Casos confirmados de malaria por cada 1000 habitantes en el Ecuador.	4
Figura 3. Mapa de ubicación del Ecuador.	6
Figura 4. Representación del hiperespacio de Hutchinson con tres variables.	10
Figura 5. Diagrama BAM.	13
Figura 6. Representación del espacio geográfico y ecológico.	14
Figura 7. Proceso para el modelamiento de nichos ecológicos.	20
Figura 8. Curva logística.	40
Figura 9. Función $\text{Sen}^2 \alpha$	45
Figura 10. Funciones de pertenencia borrosa.	47
Figura 11. Matriz de confusión.	49
Figura 12. Matriz de confusión representada en el espacio geográfico	49
Figura 13. Curva ROC	51
Figura 14. <i>Anopheles albimanus</i>	54
Figura 15. Esquema metodológico para el modelamiento de <i>Anopheles spp.</i>	56
Figura 16. Puntos de presencia de <i>Anopheles spp.</i> en el Ecuador.	59
Figura 17. Corte de las variables descriptivas en base al área de estudio.	60
Figura 18. Puntos de presencia en formato .csv.	64
Figura 19. Entorno del software Maxent.	65
Figura 20. Opciones del software Maxent.	66
Figura 21. Curva ROC generada por el modelo Maxent.	67
Figura 22. Curvas de respuesta de Maxent.	67
Figura 23. Test Jackknife de Maxent.	68
Figura 24. Resultado de Maxent para <i>Anopheles spp.</i>	69
Figura 25. Buffer sobre los puntos de presencia de <i>Anopheles spp.</i>	70
Figura 26. Zona para la creación de pseudoausencias.	71
Figura 27. Datos de presencia/ausencia.	72
Figura 28. Mapas de Distancias	72

Figura 29. Mapa de uso de suelo ponderado.....	73
Figura 30. Extracción de los valores de cada variable.	74
Figura 31. Resumen del modelo logístico en R.	75
Figura 32. Resultado de la regresión logística en R para <i>Anopheles spp.</i>	75
Figura 33. Curva ROC del modelo regresión logística.	76
Figura 34. Esquema para la evaluación multicriterio.....	77
Figura 35. Resultado de la EMC para <i>Anopheles spp.</i>	80
Figura 36. Evaluación Multicriterio - Rangos de probabilidad.....	82
Figura 37. Histograma de frecuencias - EMC.....	83
Figura 38. Regresión Logística - Rangos de probabilidad.....	84
Figura 39. Histograma de frecuencias – Regresión Logística.....	85
Figura 40. Maxent - Rangos de probabilidad.....	85
Figura 41. Histograma de frecuencias – Maxent	86
Figura 42. Análisis de Curva ROC – EMC.....	87
Figura 43. Análisis de Curva ROC – Regresión Logística	88
Figura 44. Análisis de Curva ROC – Maxent	88
Figura 45. Análisis de Curva ROC general.....	89
Figura 46. Zonas idóneas para la presencia de la especie <i>Anopheles spp.</i>	90
Figura 47. Uso de suelo de la provincia de Esmeraldas.....	92
Figura 48. Zonas de vida de la provincia de Esmeraldas	93
Figura 49. Temperatura Media Anual de la provincia de Esmeraldas.....	94
Figura 50. Precipitación Media Anual de la provincia de Esmeraldas.	95

RESUMEN

Se debe considerar a las enfermedades de transmisión vectorial como patologías que históricamente han presentado un comportamiento endémico – epidémico y que siguen constituyendo un problema de salud pública para el país. La malaria en el Ecuador tiene un comportamiento cíclico en su dinámica y está estrechamente relacionada con fenómenos climáticos, ecológicos y socioeconómicos. El principal objetivo de la presente investigación es determinar el nicho ecológico de las especies anofelinas (*Anopheles spp.*) que transmiten la malaria dentro del Ecuador mediante la aplicación de tres modelos de predicción: Maxent, Regresión Logística y Evaluación Multicriterio con lógica difusa. En primer lugar, se llevó a cabo una revisión bibliográfica desde los conceptos ecológicos fundamentales, pasando por los procesos dentro del modelamiento de nichos ecológicos, hasta la descripción biogeográfica de las especies. Posteriormente, se realizó la recopilación y tratamiento de la información para el entrenamiento dentro de los modelos. Por último, los análisis estadísticos dejaron en evidencia que el modelo que mejor describe la distribución de *Anopheles spp.* dentro del territorio fue el modelo Maxent; con el cual, se realizó la delimitación de las zonas donde la probabilidad de ocurrencia era la más alta, encontrándose varias localidades de suma importancia para el desarrollo nacional como: Esmeraldas, Guayaquil, Santo Domingo de los Tsáchilas, y otras. Esmeraldas fue la zona seleccionada para realizar una caracterización más detallada que permita brindar información referente a la presencia del vector. La existencia de manglares, actividad pecuaria y características climáticas propias de la provincia, hacen a Esmeraldas un sitio propicio para el desarrollo de *Anopheles spp.*

PALABRAS CLAVE

- MODELO DE NICHOS ECOLÓGICO
- MAXENT
- REGRESIÓN LOGÍSTICA
- EVALUACIÓN MULTICRITERIO
- ANOPHELES.

ABSTRACT

There is an important consideration should be given to vector-borne diseases such as pathologies which have historically presented an endemic – epidemic behavior and a public health problem for the country. Malaria in Ecuador has a cyclical behavior in its dynamics and is closely related to climatic, ecological and socio-economic phenomena. The main objective of this research is to determine the ecological niche of anopheline species (*Anopheles spp.*) which transmit malaria within Ecuador by applying three models of prediction: Maxent, Logistic Regression Multi-Criteria Evaluation with fuzzy logic. First, it was carried out a literature review, from the fundamental ecological concepts, including the processes within the ecological niche modeling, to biogeographical description of the species. Subsequently, the collection and processing of information for training in the models was performed. Finally, the statistical analysis made it clear that the model which best describes the distribution of *Anopheles spp.* in the territory was the Maxent model; in which, the delimitation of the areas where the probability of occurrence was the highest was performed, finding several important locations for national development such as: Esmeraldas, Guayaquil, Santo Domingo de los Tsáchilas, and others. Esmeraldas was the target area to perform a detailed characterization, which will provide more information about the presence of the vector. The existence of mangroves, livestock farming and own climatic characteristics of the province makes Esmeraldas a site conducive for the development of *Anopheles spp.*

KEY WORDS

- ECOLOGICAL NICHE MODEL
- MAXENT
- LOGISTIC REGRESSION
- MULTI-CRITERIA EVALUATION
- ANOPHELES.

CAPÍTULO I

1 Introducción

1.1 Antecedentes

En la actualidad problemas como las enfermedades transmitidas por vectores y roedores, enfermedades asociadas con la transmisión hídrica y alimentaria, infecciones respiratorias, enfermedades humanas relacionadas con cambios bruscos de temperatura, cada vez se asocian más con alteraciones del medio ambiente que propician su aparición, reaparición, incremento o incluso disminución o desaparición temporal o definitiva (Rotela *et al*, 2014).

Existen diversos factores representativos como las condiciones climáticas, fenómenos naturales, decisiones políticas, y otros; que condicionan el medio ambiente y en muchos casos significan un riesgo para la salud del ser humano, donde puede ser él mismo un agente modificador sin tener conocimiento de ello. En este sentido la geografía es útil para analizar las características de localización y distribución de las manifestaciones de los grupos humanos, sus relaciones con los elementos naturales, sus diferencias locales, su dinámica, sus inter-relaciones e inter-actuaciones en la superficie terrestre o lugar (Rotela *et al*, 2014).

Cada año, más de 200 millones de personas son infectadas con malaria en todo el mundo. Si bien el 80 % de los casos de malaria se producen en África, la región latinoamericana no queda exenta. Son 2,7 millones de personas cada año que resultan infectadas en América Latina. La enfermedad es endémica de la zona intertropical, básicamente en la cuenca amazónica (que representa más del 90 % de las infecciones), en América central y al sur de México. Si bien el riesgo es relativamente bajo en las grandes ciudades, no lo es tanto, en las zonas rurales de Bolivia, Colombia, Ecuador, Perú y Venezuela donde la enfermedad todavía persiste (Prugnotte & Yalcindag, 2012).

Tener la visión del entorno o el paisaje que rodea al individuo que puede enfermarse es lo que genera el concepto de eco epidemiología o epidemiología panorámica, en el cual se están desarrollando nuevos avances para abordar complejos problemas de salud pública. El estudio del entorno puede implicar distintos puntos de vista, componentes, factores y disciplinas para evaluarlos dado que estos problemas son multifactoriales; por ejemplo, la vegetación, la fauna, el suelo, la hidrografía, los elementos climáticos o meteorológicos, entre otros, estudiados por diferentes ciencias, entre ellas la ecología como ciencia integradora (Rotela *et al*, 2014).

1.2 Identificación del problema.

El modelamiento de nicho ecológico de especies es un tópico de investigación ampliamente explorado en todo el mundo y que ha tenido relevancia significativa durante los últimos años debido al gran potencial de aplicaciones que se le puede dar, de las cuales resalta, la investigación de vectores de enfermedades parasitarias. Dentro de este contexto, en el Ecuador las investigaciones han sido muy escasas, ya sea por falta de información, por la inexperiencia en el tema o por falta de recursos. El conocimiento de la distribución geográfica, y por ende de su nicho ecológico, del vector de la malaria en el territorio serviría como un insumo para la planificación en el control y erradicación de enfermedades causadas por vectores.

1.3 Justificación.

Se debe considerar a las enfermedades de transmisión vectorial como patologías emergentes y reemergentes que históricamente han presentado un comportamiento endémico – epidémico y siguen constituyendo un problema de salud pública en el Ecuador por ser la mayoría potencialmente letales (SNEM, 2013).

La malaria en el Ecuador tiene un comportamiento cíclico en su dinámica y está estrechamente relacionada con fenómenos climáticos y socioeconómicos. En el Ecuador causa pérdidas económicas directas e indirectas que afectan principalmente a los estratos

bajos de la sociedad. La malaria afecta principalmente a la población infantil y económicamente activa entre las edades de 0 a 39 años de edad. No existen cálculos actualizados sobre el impacto económico de la malaria, pero en 1994 se determinó que las pérdidas correspondían a un 16 - 20,8% del salario mensual vital por cada persona enferma con malaria en el Ecuador (IEE, 2014).

La mejora en la calidad de vida de los ecuatorianos ha sido evidente, en el periodo 2006-2014 el índice de Necesidades Básicas Insatisfechas (NBI) se redujo en 16,2 puntos entre 2006 y 2014 (de 52% a 35,8%) equivalente a una reducción relativa de 31,1% (ver Figura 1); lo que significó un total de 1,6 millones de personas que salieron de la pobreza estructural medida por NBI (INEC I. N., 2014). Sin embargo, a pesar del mejoramiento de las condiciones de vida y vivienda, en los últimos años en el Ecuador, aún se mantienen factores que influyen en la ocurrencia de las enfermedades transmitidas por vectores en áreas urbano–marginales y rurales del país ya que persisten las condiciones bio-ecológicas para la vida y desarrollo de artrópodos de importancia médica en aproximadamente el 70% de la extensión territorial del país (SNEM, 2013).

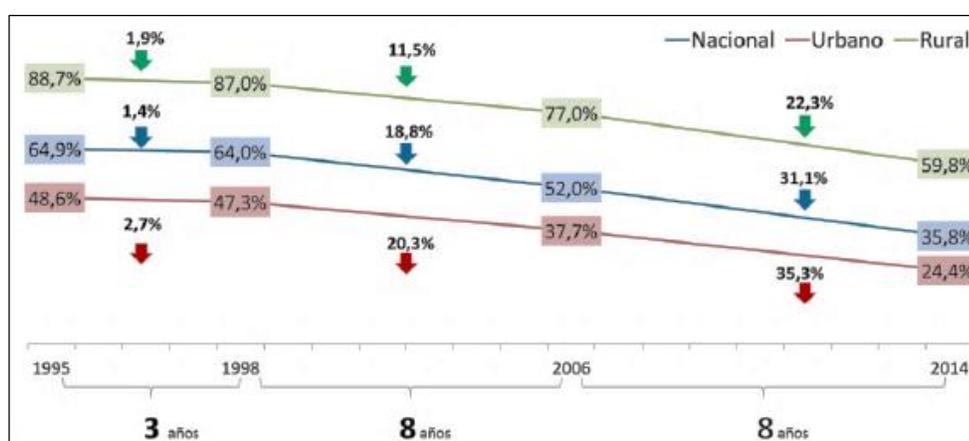


Figura 1. Pobreza por NBI nacional, urbano y rural.
Fuente: INEC, 2014.

De acuerdo al Reporte Mundial de la Malaria 2014 realizado por la Organización Mundial de la Salud se ha registrado grandes avances en la erradicación de esta enfermedad. En el Ecuador, en el año 2003 se reportaron 51.345 casos de paludismo

comparados con el año 2012 en el cual se reportaron 558, por lo que se evidencia un decrecimiento muy significativo de la incidencia de malaria, sin embargo existen zonas en las cuales la enfermedad todavía persiste (ver Figura 2).



Figura 2. Casos confirmados de malaria por cada 1000 habitantes en el Ecuador.
Fuente: Organización Mundial de la Salud, 2014.

Tal persistencia se debe a aspectos que son de carácter no modificable como: La regiones naturales del territorio continental como Costa, Amazonía, Sierra con estribaciones y valles subtropicales andinos que presentan condiciones propicias para la existencia y reproducción del anofeles, además de la presencia de fenómenos climatológicos ocasionales (El Fenómeno de El Niño) y el calentamiento global (SNEM, 2013). Además cabe señalar que el agente infeccioso (plasmodium) que permanece en el humano, y en el portador asintomático, puede permanecer desde los pocos meses hasta los 5 años (MSPC, 2010).

Considerando que la malaria es una enfermedad compleja es necesario utilizar un enfoque interdisciplinario para entender su epidemiología, estudiando la ecología de los vectores, el parásito y datos que integren varias ciencias como la Matemática, Estadística, Climatología, Entomología, y Geografía. Bajo esta visión el presente estudio pretende determinar la distribución potencial de *Anopheles spp*, transmisor del parásito que causa la malaria, para delimitar las zonas de mayor riesgo biológico para la población y posteriormente una caracterización paisajística de las mismas; de tal forma que el resultado de la investigación pueda ser una herramienta para la toma de decisiones en gestión de la salud pública y del territorio.

1.4 Objetivo General.

Modelar el nicho ecológico de especies vectores de la malaria, *Anopheles spp*, dentro Ecuador Continental mediante el empleo de modelos probabilísticos y herramientas geoinformáticas.

1.5 Objetivos Específicos.

- Determinar la distribución biogeográfica de las especies vectores para establecer la probabilidad de localización de la especie.
- Evaluar y determinar entre los modelos de distribución potencial de especies: maxent, regresión logística y evaluación multicriterio con lógica difusa; el modelo idóneo para el vector de la malaria *Anopheles spp*.
- Caracterizar paisajísticamente las zonas delimitadas de mayor riesgo biológico.

1.6 Metas.

- Tres mapas de distribución y de probabilidad de ocurrencia de las especies vectores, escala de trabajo 1:3'000 000.
- Un informe estadístico para la validación del modelo idóneo de nicho ecológico, que contenga indicadores estadísticos como AUC y curvas ROC.
- Creación de una base de datos geográfica y tres mapas temáticos de las zonas caracterizadas, escala de trabajo 1: 250 000.

1.7 Área de Estudio.

La República del Ecuador se encuentra ubicada en el sur oeste de América, bordeada por el Océano Pacífico, limitando al norte con Colombia; por el sur y el oriente con el Perú (ver Figura 3). Su superficie territorial es de 283 561 kilómetros cuadrados. Su población se estima en 14'448.499 habitantes, correspondiendo el 50,4% a mujeres y 49.6% a hombres según el Censo de población y vivienda 2010.

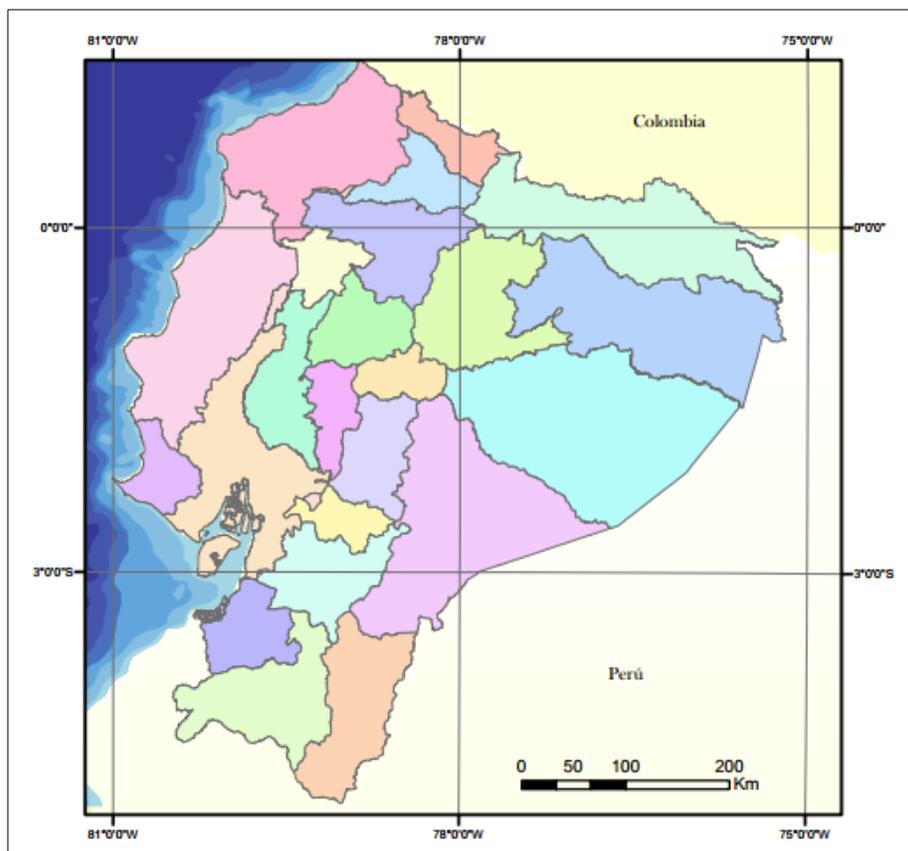


Figura 3. Mapa de ubicación del Ecuador.

CAPÍTULO II

2 Marco Teórico

2.1 Definición de nicho ecológico

La definición de nicho ecológico tiene un sitio preponderante dentro de la teoría ecológica ya que muchos preceptos se fundamentan implícita o explícitamente en el mismo. El concepto nació a partir del siglo XX como una generalización empírica, basado en analogías de distribución geográfica, formas de vida y relaciones tróficas de las especies, sin embargo durante las décadas de 1960 – 1970 se profundizó en el tema para darle profundidad y solidez a otros tópicos como el uso de recursos y hábitat, la competencia y la selección natural. (Milesi & de Casenave, 2005).

Una de las razones por la cual el concepto de nicho ecológico ha tenido varias interpretaciones es porque los científicos lo han usado para responder a una pregunta muy compleja: ¿Que combinaciones de factores ambientales permiten a las especies existir en una región geográfica dada o en una comunidad biótica dada, y cuales efectos tienen sobre las especies dichos factores ambientales? La pregunta hace referencia a un complejo conjunto de problemas, los cuales pueden ser abordados y medidos de diferentes formas (Peterson, et al., 2011).

2.1.1 Definición de Joseph Grinnell

“El nicho ecológico es la unidad de distribución final, en la que cada especie está condicionada por sus limitaciones instintivas y estructurales. La presencia o persistencia de las especies esta vitalmente ligada con su ambiente” (Grinnell, 1924).

Cada especie tiene sus características fisiológicas, morfológicas y de comportamiento, lo que hace posible que ocupen determinados espacios ofrecidos por la naturaleza. Bajo este concepto existen nichos vacíos y vacantes (Grinnell, 1924).

Al existir nichos similares en distintos lugares, éstos nichos son ocupados por organismos también similares en características morfológicas, fisiológicas y conductuales; dichos organismos son llamados equivalentes ecológicos, donde la exclusión competitiva es la interacción principal entre ellos, en la cual, competidores totales no pueden coexistir a largo plazo (Martínez, 2010).

El nicho ecológico según Grinnell ha sido interpretado más rigurosamente por Soberón (2007) como el conjunto de todos los valores de los vectores de las variables medioambientales descriptivas en los cuales la tasa de crecimiento intrínseco de la especie es positivo, como se observa en la Ecuación 1 .

$$N = \{\vec{e}_j \mid r(\vec{e}_j) > 0\} \quad (1)$$

Dónde:

$r(\vec{e}_j)$ Representa la tasa de crecimiento intrínseca de la población como función de las variables medioambientales j .

En términos más simples, es posible resumir la definición de nicho como la menor unidad de distribución de especies, la cual consiste en la suma de los requerimientos de hábitat y cuya principal característica es que el nicho es propiedad del medio y no de los organismos.

2.1.2 Definición de Charles Elton

“El nicho describe el estatus de un animal en su comunidad, indicando que hace y no solamente como se ve. El nicho de un animal es el lugar que ocupa en el ambiente biótico y su relación con la comida y sus enemigos naturales” (Elton, 1927).

Esta definición enfatiza la función de una especie dentro de una cadena alimenticia (carnívoros, herbívoros, y otros.). El nicho en teoría no está restringido a una especie, por tanto los organismos relacionados como equivalentes ecológicos serían un indicio de nichos similares, aunque las comunidades estuvieran en lugares muy alejados (Martínez, 2010).

Elton analizó al nicho como el rol funcional de un organismo en una comunidad (sus efectos locales). La existencia del nicho fue dada por echo, pero el énfasis fue colocado sobre los impactos

Cabe señalar que este concepto de nicho es aplicado en su mayoría en la ecología de comunidades, además, está definido dentro de un espacio geográfico pequeño en el cual las manipulaciones experimentales son factibles, enfatizando el rol de las especies e incluyendo modelos de consumo de recursos y su impacto en las comunidades (Peterson, et al., 2011).

2.1.3 Definición de G. Evelyn Hutchinson

“El termino nicho es definido como la suma de todos los factores medioambientales que actúan sobre el organismo; el nicho así definido es una región de un hiperespacio n-dimensional si este proceso puede ser llevado a cabo con todas las variables, físicas y biológicas, el nicho fundamental de las especies será completamente definido por sus propiedades ecológicas. El nicho fundamental definido de esta forma es meramente una

formalización abstracta de lo que usualmente se dice nicho ecológico” (Hutchinson G. , 1957).

Se pueden definir las condiciones y recursos que las especies necesitan a partir de estudiar la interacción de sus tolerancias y requerimientos. Para ello, cada factor (Ej. temperatura) es una dimensión del nicho en la que la especie tendrá un rango que le resulta adecuado. De la misma manera procedió con todas las demás dimensiones incluyendo variables ambientales abióticas o bióticas (Ej. humedad relativa, pH, salinidad, complejidad de la vegetación, y otros.) Se pueden graficar dos dimensiones como dispersión de puntos, e incluso tres dimensiones (ver Figura 4). Si se pudiera graficar la intersección de todas las características se obtendría un hipervolumen n-dimensional que sería el verdadero nicho de la especie, sin embargo resulta imposible visualizarlo y medir todas las variables posibles no es realizable. Este nicho es dinámico pues la presencia de unas especies constriñe la presencia de otras, modificando la posición de las primeras en el espacio ecológico multidimensional (Martínez Ainsworth, 2013).

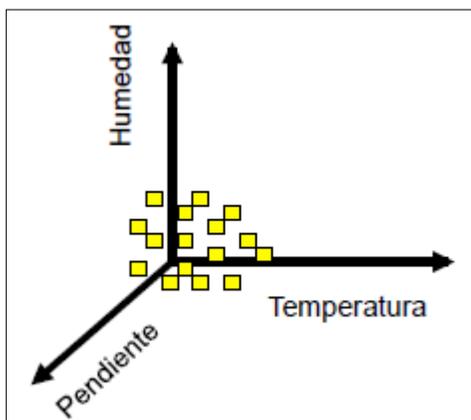


Figura 4. Representación del hiperespacio de Hutchinson con tres variables.
Fuente: Martínez (2010).

2.2 Conceptos relevantes sobre el nicho ecológico

2.2.1 Tipos de nicho ecológico

Del trabajo realizado por Hutchinson se derivan algunos términos relevantes para el modelamiento de nichos ecológicos que se relacionan o aplican directamente con otras definiciones como:

- **Nicho fundamental:** Todos los aspectos (variables) medioambientales del espacio o hipervolumen, en la ausencia de otras especies. En pocas palabras es donde la especie puede vivir (Martínez, 2010).
- **Nicho realizado o efectivo:** Es un subconjunto del nicho fundamental en el cual las especies están restringidas debido a sus interacciones interespecíficas. También se puede definir como la región donde la probabilidad de encontrar a una especie sería alta. Es también, por hipótesis, la región donde coinciden los factores abióticos, bióticos y migratorios. De manera menos compleja, es el espacio ecológico y geográfico donde le especie vive (Martínez 2010, Martínez-Meyer 2011).

2.2.2 Diagrama BAM

El área de distribución de una especie representa la compleja expresión de su ecología y de su evolución histórica, la cual, está determinada por varios factores como (Soberón & Peterson, 2005) :

- Factores abióticos: que incluye los aspectos climáticos, físicos, condiciones edáficas, y otros.
- Factores bióticos: que consiste en el conjunto de interacciones con otras especies que modifican la habilidad de la especie para mantener su población. Estas

interacciones pueden ser positivas (mutualistas como las dispersión de semillas, polinizadores, y otros.) o negativas (competidores, predadores, parásitos). Ya sea por limitar o mejorar los procesos dentro de la población estas interacciones afectan la distribución de la especie.

- Las regiones que son accesibles para la dispersión de las especies.
- La capacidad evolutiva de las especies para adaptarse a las nuevas condiciones: usualmente este factor queda excluido en el análisis de la distribución de una especie, ya que los cambios evolutivos se dan después periodos largos de tiempo (cientos o miles de años).

Todos estos factores interactúan dinámicamente a diferentes intensidades y a diferentes escalas para determinar la distribución de una especie. La forma más fácil de visualizar e interpretar estas interacciones es el diagrama BAM (Biótico, Abiótico y Movimiento) (ver Figura 5).

El diagrama BAM resume los tres principales factores que limitan la distribución de una especie A (bióticos), B (abióticos) M (movimiento o dispersión). El hábitat es el ideal para la especie dentro de la zona común a estos tres factores, la cual se conoce como nicho ocupado o realizado (NR); la región que intersecan A y M corresponde a al nicho definido por Grinnell (NG); el espacio común a A y B es el nicho eltoniano (NE) y toda el área de A es el nicho fundamental de Hutchinson (NF). La especie consigue sobrevivir en aquellas condiciones climáticas favorables, donde se logra dispersar y en las que no hay interacciones bióticas que la excluyan. Las regiones que no cumplan con alguna de estas condiciones no son idóneas para la presencia de la especie.

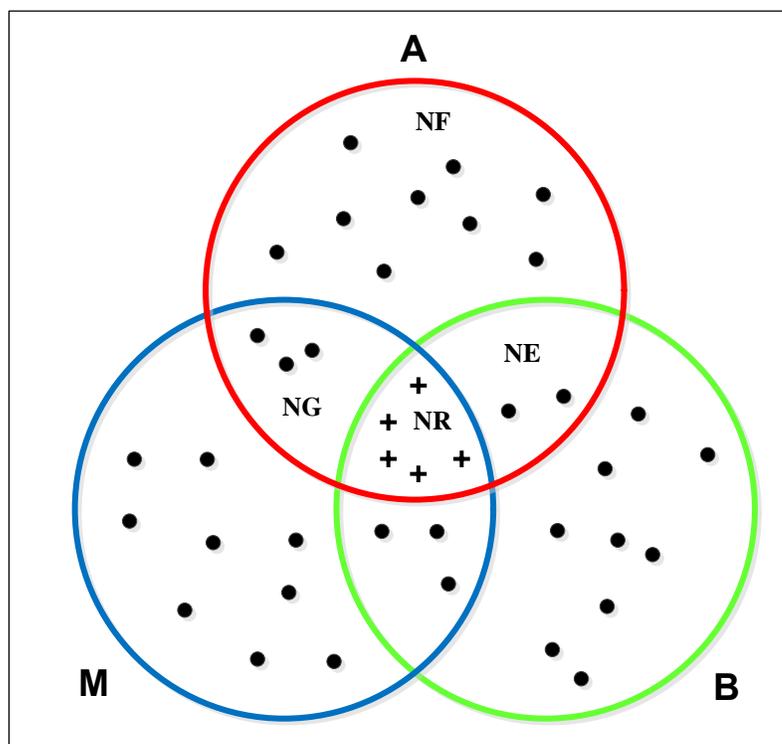


Figura 5. Diagrama BAM.
Fuente: Siliero et al (2010)

2.2.3 Relación entre el espacio geográfico y espacio ecológico

Generalmente cuando se habla de ocurrencia de una especie se le relaciona inmediatamente al contexto geográfico, donde la distribución de la especie puede ser representada dentro del espacio geográfico a través de un mapa. Para entender los modelos de distribución de especies es importante también entender que la ocurrencia de una especie puede verse desde el espacio ecológico, el cual es un espacio conceptual definido por las variables medioambientales o escenopoéticas, llamadas así por Hutchinson, a las cuales la especie responde (Villar Hernández, 2014).

Los puntos de ocurrencia constituye todo lo que se sabe acerca de la distribución actual de la especie; es probable que la especie ocurra en otras áreas en las cuales aún no ha sido detectada (Figura 6, área A). Si la distribución actual se representa en el espacio ecológico, entonces podemos identificar que parte del espacio ecológico está ocupado por la especie y se define a este espacio como nicho ocupado; el cual, refleja todas las

restricciones impuestas a la distribución actual, incluyendo restricciones espaciales relacionadas con la capacidad de dispersión, y las múltiples interacciones con otros organismos (Pearson, 2008).

Si las condiciones medioambientales encapsuladas dentro del nicho fundamental se representan en el espacio geográfico, entonces lo que se tiene es la distribución potencial de la especie. Note que algunas regiones de la distribución potencial no pueden ser habitadas por la especie (Figura 6, áreas *B* y *C*), ya sea porque la especie ha sido excluida debido a interacciones bióticas (por ejemplo la presencia de un competidor o la falta de fuente de alimentación); o bien porque la especie aún no se ha dispersado por el área (por ejemplo una barrera geográfica tal como una montaña); o bien porque la especie se ha extinguido (por ejemplo por la modificación humana del paisaje) (Pearson, 2008).

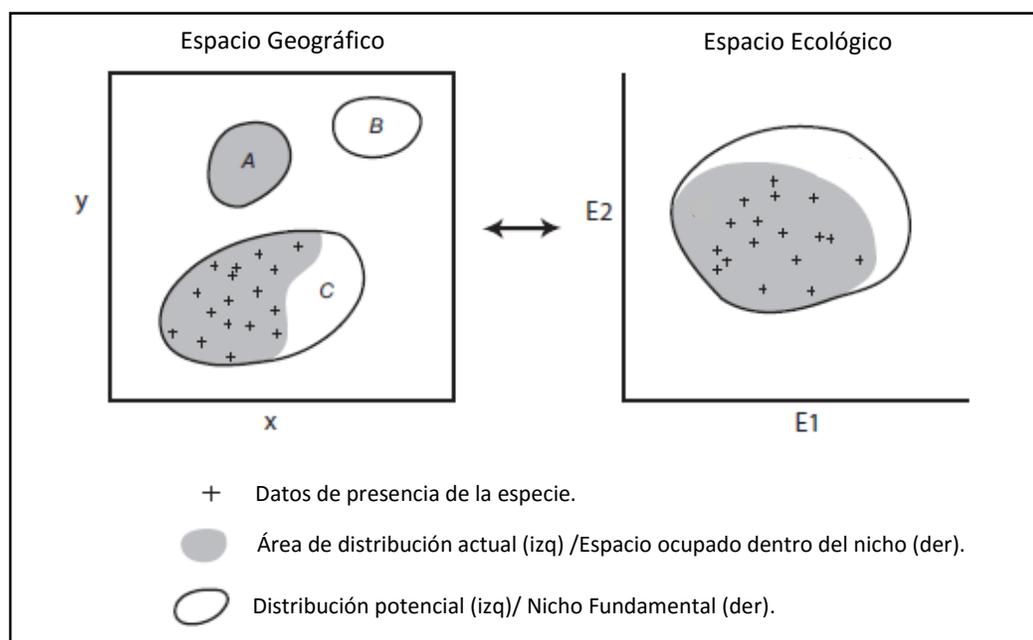


Figura 6. Representación del espacio geográfico y ecológico.
Fuente: Pearson (2008)

Para el modelamiento del nicho ecológico de una especie, se debe tener en mente algunas consideraciones adicionales de suma importancia. Cabe destacar que es improbable que las variables medioambientales contengan todas las posibles dimensiones del espacio ecológico para la construcción de modelos, por tanto, las

variables disponibles para modelar solo representan un subconjunto de los posibles factores medioambientales que influyen en la distribución de las especies. Otro factor de suma importancia a considerar son las dinámicas llamadas “Fuente-sumidero” son aquellos casos donde se observan especies en áreas que no proporcionan todos los recursos para la sobrevivencia de la especie (“sumidero”), sin embargo existen zonas aledañas que si prestan todas las condiciones necesarias para la presencia de la especie (“fuente”); lógicamente estos eventos ocurren con mayor frecuencia en especies con alta capacidad de dispersión, tales como las aves (Pearson, 2008).

En base a los conceptos establecidos por Hutchinson (1957) y el análisis hecho por Pearson, (2008); Martínez (2010) concluye que para cada punto en el espacio geográfico existe uno y sólo un punto en el espacio ecológico; en cambio, para cada punto en el espacio ecológico puede haber más de un punto en el espacio geográfico.

2.3 Modelos de Nicho Ecológico (MNE)

En años recientes ha surgido un creciente interés en el estudio de nichos ecológicos, esto ha sido posible gracias a la confluencia de la creciente necesidad de información de la distribución geográfica de la biodiversidad y de las nuevas y mejoradas técnicas de recolección de datos idóneos para abordar esta información como: sensores remotos, tecnología de sistemas de posicionamiento global, sistemas de información geográfica y métodos de aprendizaje estadístico (Franklin, 2010).

Los modelos de nicho ecológico (MNE) son aproximaciones empíricas o matemáticas al nicho ecológico de una especie cuyo objetivo primordial es relacionar diferentes tipos de variables sean estas bio-físicas, topográficas, climatológicas y humanas con la distribución de las especies, esto es, identificar los factores que limitan y definen dicho nicho. El resultado final de un MNE puede ser una representación espacial de los hábitats favorables a la presencia de una especie (Silero, Barbosa, Martínez-Freira, & Real, 2010).

Los MNE pueden predecir nichos favorables para la existencia de la especie en áreas donde no se haya muestreado o ha existido un muestreo escaso, así mismo puede hacer una proyección temporal de la existencia de una especie debido a cambios previsibles en el entorno. Los MNE pueden también ser relacionados a estudios de biodiversidad para determinar la abundancia de una especie o en investigaciones sobre plagas con la probabilidad de persistencia en un determinado sector (Silero, Barbosa, Martínez-Freira, & Real, 2010).

Los modelos de distribución de especies son técnicamente la distribución de los factores ambientales adecuados que se modelan o mapean. Esta generalización hace que estos modelos sean apropiados para el estudio de la distribución de las variables biogeográficas relacionadas, la riqueza de las especies, especies potencialmente invasoras y vectores de enfermedades (Miller, 2010).

Los MNE pueden clasificarse en mecanicistas (explicativos) o estadísticos/correlativos (predictivos). Los MNE mecanicistas basan sus predicciones en relaciones de causa-efecto entre las variables y la distribución de la especie, por lo que tienen relevancia ecológica principalmente. Para esto, utilizan variables que demuestran (por ejemplo mediante experimentos de termo-tolerancia) que afectan directamente a la supervivencia de la especie, como la temperatura, la humedad, alimento, y otros. Por otro lado, los MNE estadísticos se basan en correlaciones significativas entre la distribución de la especie y las variables que no tienen necesariamente un efecto directo sobre ella, como la altitud o la precipitación, pero que resumen el efecto de diversos factores más directos; además son más sencillas de medir. Estos modelos suelen ser más precisos y pueden tener también un componente explicativo: más allá de proporcionar simplemente una predicción de las distribuciones geográficas, pueden reflejar aspectos relevantes de la biología e historia natural de las especies, sugiriendo factores ecológicos subyacentes (Silero, Barbosa, Martínez-Freira, & Real, 2010)

La existencia de una correlación entre una variable y la distribución de una especie no implica una relación de causa-efecto. La interpretación explicativa de los MNE debe hacerse, por tanto, con precaución, ya que el efecto causal de una variable sobre la especie puede ser enmascarado por los efectos de otras variables no causales correlacionadas con ella (Silero, Barbosa, Martínez-Freira, & Real, 2010).

Pearson (2014) resalta ciertas características de estos dos enfoques en cuanto al modelamiento de nichos ecológicos: El primero, el modelamiento correlativo donde se destacan las siguientes consideraciones:

- La mayoría de los modelos utilizados para determinar la distribución de especies utiliza este enfoque ya que se da por hecho que la distribución actual de la especie da un buen indicador de los requerimientos ecológicos de la misma.
- Delimita el espacio que ocupa actualmente una especie.

Por otro lado el modelamiento mecanístico resalta que:

- Los datos de ocurrencia registrados no bastan para modelar la distribución de la especie por ello se fundamenta en los datos fisiológicos y morfológicos de la misma (por ejemplo: termo-tolerancia, porcentaje de humedad necesaria, y otros.)
- Delimita el espacio que potencialmente puede ser ocupado por una especie.

2.3.1 Proceso para el modelamiento de nichos ecológico.

El modelamiento de nichos ecológicos es un proceso complejo, el cual requiere de una serie de decisiones que están relacionadas con el objetivo del estudio, la clase y la calidad de datos disponibles, y la habilidad del modelador (Graham *et al*, 2012). Debido al constante avance dentro de este campo de investigación, no existe un proceso ya establecido para determinar la distribución de una especie, sin embargo para Peterson *et*

al (2011) existen ciertas directrices o conceptos generales en los que se enmarca toda investigación referente a este tema:

- El área de estudio es conceptualizada como un raster de una extensión determinada, compuesto de celdas de una resolución específica.
- La variable dependiente es la distribución de la especie la cual es obtenida de los registros de solo presencia o presencia-ausencia, a estos puntos se los llamara **G_{data}**.
- El conjunto de variables medioambientales descriptivas es recopilado para caracterizar cada celda del área de estudio en términos ecológicos, conocido como **E**.
- Una función $\mu(\mathbf{G}_{data}, \mathbf{E})$ que caracteriza la variable dependiente en términos de las variables medio-ambientales generadas, dicha función determina el grado en el cual cada celda es idónea para la especie en cuestión.
- Por último es necesaria una evaluación de la precisión de las predicciones en términos cuantitativos.

En términos generales la noción de aproximar los fenómenos naturales para su estudio se puede expresar matemáticamente con una función dada por $Y = f(X)$. El término “modelo” es frecuentemente usado como un sinónimo de las palabras algoritmo, predicción o método, esta denominación es representada por la expresión $\mu(\mathbf{G}_{data}, \mathbf{E})$, descrita anteriormente. Los diferentes algoritmos o modelos tienen un objetivo en común que es producir funciones $f_n(X_n)$ que puedan ser usados para computar una predicción de Y para una X dada.

Por lo tanto el modelo puede ser expresado como:

$$Y = \mu (G_{\text{data}}, E) \quad (2)$$

$$Y = f_1 (X_1) + f_2 (X_2) + f_2 (X_2) + f_3 (X_3) + \dots + f_n (X_n) \quad (3)$$

Algunos métodos dan como resultado una salida de ceros o unos ($Y = 0$ ó $Y = 1$), en cambio otros métodos producen una salida continua donde mientras más alto sea el valor más alta es la probabilidad de presencia (Peterson, et al., 2011).

Es muy importante incorporar la teoría ecológica en cada paso del proceso, desde la conceptualización de los modelos basados con asunciones ecológicas de pseudo-equilibrio hasta la obtención de las curvas de respuesta. Los tipos de datos y algoritmos de modelamiento usados determinan el mapa resultante, por lo tanto el objetivo general deberá ser dar soporte a cada paso y decisión dentro del modelo (Miller, 2010). En la Figura 7, se muestra una visión en conjunto del proceso delineada por Miller (2010); el cual será detallado paso a paso en los siguientes apartados de este documento.

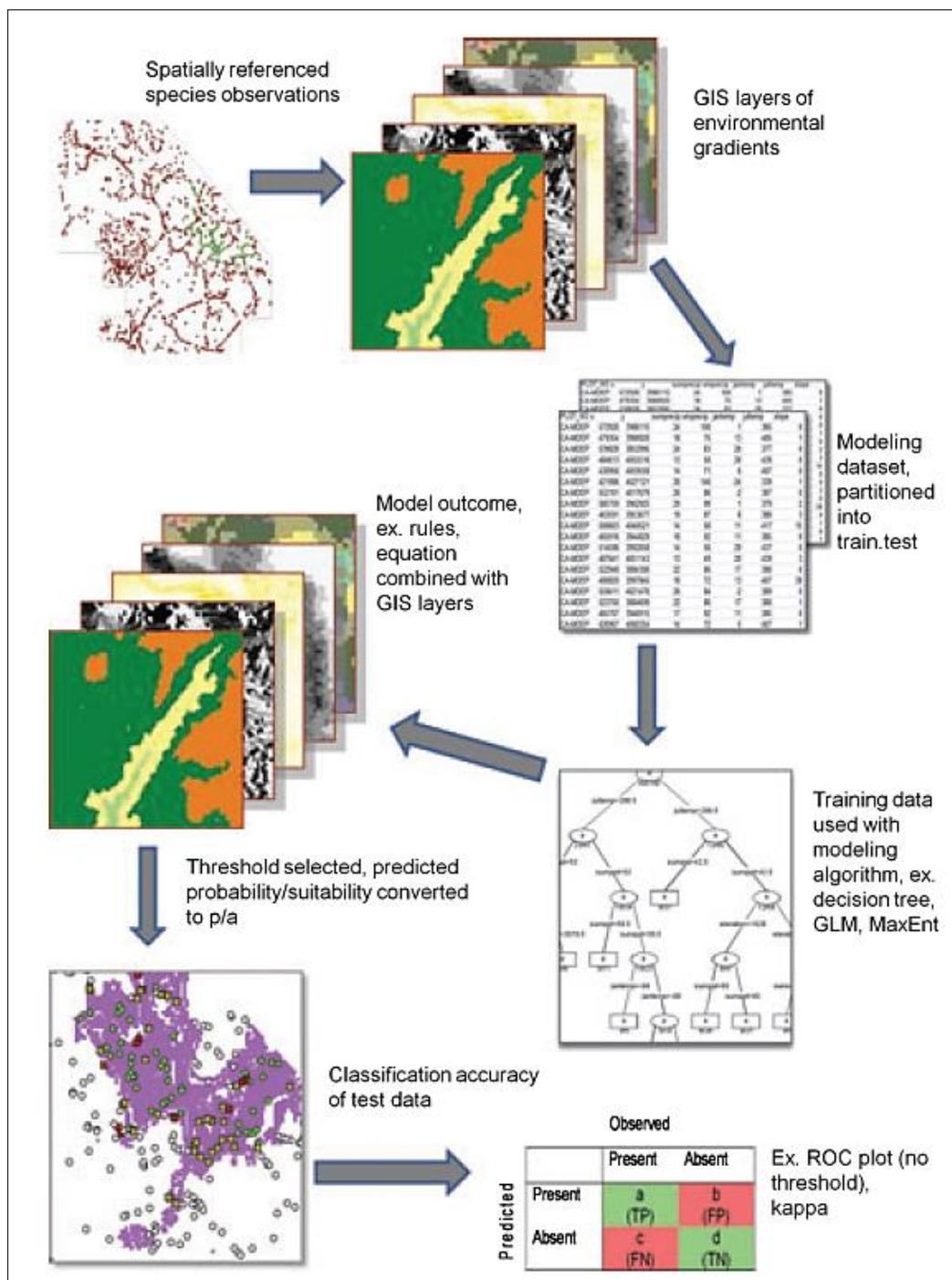


Figura 7. Proceso para el modelamiento de nichos ecológicos.
Fuente: Miller (2010).

2.3.1.1 Datos biológicos de la especie

Los datos biológicos que describen la distribución de una especie pueden ser medidos de forma nominal (Ej. Presencia/ ausencia), ordinal (Ej. Número de individuos), además por niveles de proporción o índices (Ej. Abundancia, riqueza), todos ellos tienen un efecto diferente y un uso en cada algoritmo de distribución de especies, por lo cual el producto generado será diferente para cada uno (Ej. Probabilidad de ocurrencia, media esperada, y otros.) (Miller, 2010). De acuerdo con Vaughan & Ormerod (2003) para un adecuado análisis una proporción de 10 observaciones por cada variable predictora es considerado como mínimo para la construcción de un modelo de datos, sin embargo esta proporción puede ser afectada por la autocorrelación espacial de los datos. Entendiéndose por autocorrelación espacial como la concentración o dispersión de valores de una variable en un mapa dicho de otro modo el grado en que objetos o actividades en una unidad geográfica son similares a otros objetos o actividades en unidades geográficas próximas (Vilalta y Perdomo, 2005). Franklin (2010) sugiere un tamaño de muestra mínima sea una proporción 10 a 1 y 20 a 1 cuando se trata de modelos de regresión.

El aspecto más relevante a considerar sobre los datos biológicos es que describan el hábitat idóneo para la especie en estudio, por ello es necesario que sean colectados dentro de un amplio rango medioambiental y geográfico. Las gradientes medioambientales han sido usadas para el diseño de las estrategias de muestreo, con ello se puede encontrar una amplia variedad de especies. Sin embargo, más recientemente se han enfocado los esfuerzos por maximizar la utilidad de la información ya compilada en museos, colecciones de historia natural e investigaciones previas (Miller, 2010). Se estima que existen registros de 2.5 billones de especímenes en bases de datos digitales en todo el mundo compilados desde los años 70 (Graham, Ferrier, Huetteman, Moritz, & Peterson, 2004). La limitación que presentan este tipo de datos, es que se encuentran únicamente registros de solo presencia, los cuales pueden estar sesgados por efectos propios del muestreo. La falta de puntos de ausencia hace que estos datos no sean muy

útiles para ciertos algoritmos de modelamiento de nicho ecológico, sin embargo; algunos algoritmos son más robustos que otros y tienen la capacidad de lidiar con estos problemas, por ejemplo, “pseudo-ausencias” pueden ser generadas dentro de las áreas que no han sido muestreadas para compensar esta falta de información (Miller, 2010). A continuación se presenta en la Tabla 1 algunas fuentes de información disponibles en la web.

Tabla 1.

Fuentes de información de datos biológicos

Organización	URL	Información disponible
Global Biodiversity Information Facility	http://data.gbif.org/welcome.htm	Especies de plantas, animales, hongos y micro-organismos.
National Invasive Species Information Center	http://www.invasivespeciesinfo.gov/resources/regdb.shtml	Bases de datos regionales de especies invasoras.
NatureServe	http://www.natureserve.org/	Animales, plantas y ecosistemas del Norte y Sur América
North American Breeding Bird Survey	http://www.pwrc.usgs.gov/BBS/	Datos colectados desde 1966 de más de 400 especies de aves de todo el mundo.
VertNet	http://vertnet.org/index.php	Base de datos sobre vertebrados colectados por los museos de historia natural de todo el mundo (MaNIS, HerpNet, ORNIS y FishNet)

Fuente: Miller, J. (2010). Species Distribution Modeling. *Geography Compass* No.46, 490-509.

2.3.1.2 Datos medio ambientales

Una adecuado modelamiento también recae en la disponibilidad de datos espaciales que contengan la información medioambiental necesaria para determinar la distribución de la especie (Franklin, 2010). En la literatura, se pueden reconocer tres tipos de variables ambientales: las variables directas, indirectas y de recursos (Guisan & Zimmermann, 2000). Las variables directas representan parámetros ambientales de importancia fisiológica pero que no son consumidas por los individuos (Ej. Temperatura); mientras que las variables indirectas corresponden a aquellas que no tienen un efecto fisiológico directo sobre el individuo (Ej. Pendiente) pero presentan una alta correlación para la presencia de la especie. Bajo este concepto, los animales presentan respuestas indirectas ante variables como la vegetación (Ej. Refugio), topografía (Ej. Locomoción), y otros. Por último, las variables de recursos se refieren a materia y energía consumida por animales (Ej. Agua, nutrientes o presas) (Peterson, et al., 2011; Riquelme, 2013).

A su vez, diversos autores han considerado la inclusión de variables que no están relacionadas directamente con ninguno de estos fenómenos, rasgos del hábitat (Ej. Caminos, áreas urbanas, uso de suelo), pero que en la modelación presentan buenas correlaciones con los patrones de distribución de las especies (Riquelme, 2013). Gracias al incremento de diversas técnicas de modelamiento espacial y sensores remotos ahora está disponible una abundante cantidad de datos medioambientales de alta calidad, dicha información está siendo generada a diversas escalas y con diferentes resoluciones, desde globales hasta locales, con pixeles que van en rangos de $<10^0$ hasta 10^5 metros (Peterson, et al., 2011).

Las variables climáticas y topográficas son las más ampliamente usadas en los modelos de nicho ecológico, ya que estas describen a gran escala factores de tolerancia fisiológica relacionadas con la temperatura y la precipitación y en escalas más finas la

variación espacial correspondiente a fuentes de energía y humedad disponible (Miller, 2010).

Un grupo de científicos de la Universidad de California definieron un conjunto de parámetros climáticos que a su consideración son los más relevantes para determinar la distribución de una especie en escalas espaciales amplias. El resultado de esta investigación fue llamado WORLDCLIM; que es un conjunto de datos espaciales resumidos en 19 variables bioclimáticas resultantes de la aplicación de un algoritmo llamado “Thin-plate smoothing spline”, siendo uno de los más eficientes comparado con otros métodos de interpolación múltiple (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005). Este conjunto de datos utiliza largos periodos (30 años) mensuales de precipitación y de temperatura máxima y mínima diaria; de estos dos parámetros se derivan 16 variables más, que resultan en los promedios, descripción de los parámetros por estaciones y las condiciones extremas. Actualmente, la mayor parte de estudios relacionados al modelamiento de nichos ecológicos que engloban escalas regionales o a nivel de país usan este tipo de variables bioclimáticas por su buen desempeño y fácil acceso (Franklin, 2010).

Además de las variables climáticas, generadas normalmente a partir de la interpolación de datos de estaciones climatológicas, se suele utilizar información de diversa índole que puede ayudar al modelo como: edafología, litología y geología, que representan la dependencia de la vegetación con el tipo de sustrato, la elevación y variables derivadas, tanto topográficas (Ej. Pendiente, rugosidad) y microclimáticas (Ej. Radiación solar), hidrológicas (Ej. Red de drenaje o flujo potencial); además de variables obtenidas mediante teledetección, como índices de vegetación, temperatura en superficie o clasificaciones de la cubierta del suelo. Finalmente, algunos modelos han tenido en cuenta variables de tipo demográfico y de ocupación del espacio, como índices de población, accesibilidad, densidad o proximidad a vías de comunicación, que son útiles como indicadores del grado de influencia antrópica (Mateo, Felicísimo, & Muñoz, 2011).

Es claro que el tipo y número de variables que describen las dimensiones del nicho ecológico varían de una especie a otra, por lo que en la práctica, se han desarrollado diferentes estrategias para la selección de variables que mejor describan a la especie. Algunos investigadores prefieren usar pocas variables preseleccionadas, que se encuentran poco correlacionadas y que corresponden directamente a la respuesta fisiológica de la especie (Ej. Temperatura, humedad relativa, y otras.) (Pearson, 2008). Por otro lado, cuando no existe conocimiento previo de los factores claves que influyen la distribución de la especie se utilizan conjuntos de datos grandes (20 variables o más); muchas de estas variables están altamente correlacionadas, lo que ha sido objeto de muchas críticas. Ambos enfoques tienen sus ventajas e inconvenientes. Cuando se usa muy pocas variables los modelos corren el riesgo de perder información valiosa lo que resulta en una pobre caracterización del nicho ecológico, donde las áreas de distribución de la especie son demasiado amplias y no corresponden a la realidad. Mientras que un excesivo número de variables tiende a sobre-ajustar el modelo, lo que causa una pérdida de los grados de libertad en los modelos de regresión y una escasa representación geográfica de la distribución de los organismos, particularmente cuando se utiliza muy pocos datos de presencia; además estas variables generalmente poseen un grado de correlación muy alto, lo que impide la aplicación de algunos modelos de predicción basados en parámetros estadísticos (Peterson, et al., 2011).

Para evitar todos estos problemas se han propuesto varias metodologías para explorar la relación que existe entre las variables. Algunos investigadores han usado análisis de correlaciones para identificar los conjuntos de datos que menos se correlacionan entre si y otros han transformado las variables originales en nuevas variables no correlacionadas mediante análisis de componentes principales (PCA) y otras técnicas (Ej. Transformación de la variable temperatura en anomalías de temperatura) (Peterson, et al., 2011). En resumen, existe una diversidad de fuentes de información espacial (ver Tabla 2) que describen varios fenómenos biofísicos, donde el principal reto para el modelador es identificar las variables medio-ambientales que

realmente representan gradientes u otros factores que determinan la distribución de una especie en una escala apropiada.

Tabla 2.

Fuentes de datos medioambientales usados para la modelación de nichos.

Organización	URL	Información disponible
WorldClim	http://www.worldclim.org/	Coberturas del clima para todo el mundo a 1 km de resolución.
Intergovernmental Panel on Climate Change (IPCC)	http://www.ipcc-data.org/	Escenarios futuros de cambio climático para todo el globo
NOAA	http://www.ncdc.noaa.gov/paleo/paleo.html	Paleoclimas
USGS	http://edc.usgs.gov/products/elevation/gtopo30/hydro/index.html	Elevación y otras variables relacionadas a 1 km de resolución.
Global Landcover Facility	http://edc.usgs.gov/products/elevation/gtopo30/hydro/index.html	Imágenes de satélite y datos provenientes de sensores remotos
NASA	http://modis.gsfc.nasa.gov/data/	Información climatológica y NDVI proveniente del sensor MODIS.
UNEP	http://www.grid.unep.ch/data/data.php?category=lithosphere	Tipo de suelo
NOAA	www.nodc.noaa.gov	Varios datos que describen los océanos de todo el mundo.
SNI	http://sni.gob.ec/inicio	Portal de información proveniente de diversas fuentes gubernamentales del Ecuador como: IGM, MAGAP, MAE, IEE, y otros

Fuente: Pearson, R. (2008). *Species' Distribution Modeling for Conservation Educators and Practitioners*.

American Museum of Natural History: Center for Biodiversity and Conservation & Department of Herpetology.

2.3.1.3 Métodos para el modelamiento de nichos ecológicos.

Comúnmente se usa el término “algoritmo de modelamiento” o “método de modelamiento” para referirse al proceso, regla o función matemática que es usada para estimar el nicho ecológico como una función de las variables medio ambientales.

Existe un gran número de algoritmos y paquetes de software que permiten describir la relación entre el medioambiente y las especies (Miller, 2010). Aunque muchos SIG ya están provistos de un conjunto de teoría y técnicas adecuadas para manejar los datos espaciales usados en el MNE, los métodos estadísticos básicos (Ej. Regresión lineal) disponibles como parte de muchos software SIG están basados en asunciones poco apropiadas para usarse con datos biogeográficos. Los SIG juegan un papel fundamental para llevar a cabo el modelamiento, desde la entrada de datos, el manejo de los mismos, hasta la etapa de salida donde se obtienen los productos cartográficos en sí. El análisis estadístico es ampliamente desarrollado por sí solo o a través de un software estadístico (Ej. R) y posteriormente integrado a un SIG para la producción de mapas o resultados similares (Miller, 2010). Numerosos enfoques han sido usados para el modelamiento de nichos ecológicos y nuevos métodos (ya sea variantes de los mismos) son publicados cada año. Los estudios publicados dan varios nombres a los algoritmos o en su defecto son acrónimos de dichos métodos, por ejemplo “Maxent” se refiere a la implementación del principio de máxima entropía (Phillips et al. 2006); otros suelen ser GLM, GAM, GARP, ANN, etc. En algunos casos, los algoritmos han sido implementados de forma amigable para el usuario a través de paquetes de software que generalmente están disponibles de forma gratuita (Peterson, et al., 2011).

Existen importantes diferencias entre los algoritmos de modelamiento: incluyendo (1) los tipos de datos biológicos que requieren (Ej. Solo presencia, presencia-ausencia, presencia y datos de fondo, o presencias y pseudo-ausencias); (2) el enfoque metodológico (Ej. Algoritmos basados en métodos de regresión, procesos de

clasificación, aprendizaje automático, o estadística Bayesiana); (4) la habilidad de incorporar variables categóricas.

Tabla 3.

Algoritmos para el modelamiento de nichos ecológicos.

Software	Algoritmo	URL, Referencias
ANUCLIM	EE	http://cres.anu.edu.au/outputs/anuclim.php
BIOCLIM	EE	http://cres.anu.edu.au/outputs/anuclim/doc/bioclim.html (Busby, 1991)
BIOMAP	EE	http://cres.anu.edu.au/outputs/anuclim.php
BIOMAPPER	Ecological Niche Factor analysis (ENFA)	www.unil.ch/biomapper , (Hirzelet al., 2002)
BIOMOD	GLM,GAM,CART,ANN	At the discretion of the author (Thuilleret al.,2003)
DIVA	EE	http://www.diva-gis.org (Hijmans et al.,2001)
DOMAIN	Distancias de Gower	www.cifor.cgiar.org/scripts/default.asp?ref=research tools/domain/index.htm http://uwadmnweb.uwyo.edu/wyndd/ (Carpenter et al.,1993)
ECOSPAT	Regresión logística	http://www.ecospat.unil.ch ; with permission of the developer
GARP	Algoritmo genético	http://lifemapper.org/desktopgarp (Stockwell & Peters, 1999) (Stockwell, 2006)
HyperNiche	Regresión no paramétrica multiplicativa	http://home.centurytel.net/~mjm/hyperniche.htm (McCune, 2006)
MARS	Splines de regresión adaptativa multivariada	www.salford-systems.com/mars.php
Maxent	Máxima entropía	http://www.cs.princeton.edu/~schapire/maxent/ (Phillipset al.,2006)
NNETW	Redes neuronales artificiales	
OpenModeller	GARP, Envolturas bioclimáticas	http://openmodeller.sourceforge.net
PRESENCE	Regresión logística	
SAM	OLS, Autoregresión	http://www.ecoevol.ufg.br/sam/ (Rangelet al.,2006)
STATMOD ZONE	CART	http://www.gis.usu.edu/~chrisg/avext

Fuente: Miller, J. (2010). *Species Distribution Modeling*. Geography Compass N.46, 490-509.

Nota: Muchos métodos, incluyendo GLM, GAM, CART, ANN, MARS y distancias de Mahalanobis, están también disponibles en software estadísticos y SIG.

CART = Árboles de regresión y clasificación

EE = Envoltura climática.

GAM = Modelos aditivos generalizados.

GLM = Modelos lineales generalizados.

OLS = Regresión ordinaria por mínimos cuadrados.

2.3.1.3.1 Maxent

Máxima entropía o Maxent es un modelo de propósito general que ha sido desarrollado mediante conceptos de mecánica estadística (Jaynes, 1957) y la teoría de la información de Shannon, y es aplicado en diversos campos del conocimiento como: física, química, finanzas, astronomía y ecología (Franklin, 2010). El principio de máxima entropía establece que una distribución de probabilidad con máxima entropía (la más dispersa o cercana a la uniforme) sujeta a restricciones conocidas es la mejor aproximación de una distribución desconocida, porque ésta incluye toda la información que se conoce pero evita cualquier asunción que no esté dentro de la información disponible (Phillips, Anderson, & Schapire, 2006). Cuando es aplicada al problema del modelamiento de nichos ecológicos, la distribución de la especie debe estar acorde con todo lo que es conocido (o inferido a partir de las condiciones medioambientales dónde la especie han sido observada) pero debe evitarse hacer cualquier asunción que no esté basada en los datos disponibles. El enfoque es por lo tanto, encontrar la probabilidad de distribución de máxima entropía sujeta restricciones impuestas por la información disponible respecto a la distribución observada de las especies y las condiciones medioambientales del área (Pearson, 2008). Bajo este principio se ha desarrollado un software específico para el modelamiento de nichos ecológicos bajo este mismo nombre Maxent. Este método ha generado gran interés porque en comparación con otros métodos ha demostrado generar predicciones muy precisas únicamente con datos de solo presencia (Elith, et al., 2006).

Phillips *et al* (2006) destacan ciertas ventajas Maxent para el modelamiento de distribución de especies sobre otros modelos. Una de las diferencias importantes entre Maxent y otros modelos, es que en Maxent, las ubicaciones donde no existen registros de ocurrencia no son interpretadas como ausencias o pseudo-ausencias, sino más bien como la representación de una muestra de fondo o background, por lo que Maxent únicamente requiere de datos de presencia e información medioambiental del área de estudio. Los resultados son fáciles de interpretar debido a la forma como las funciones

responden a las variables ambientales. El método puede utilizar variables continuas y categóricas, y los datos de salida son representados como predicciones continuas (como datos de probabilidad en crudo y más comúnmente una probabilidad acumulada en un rango de 0 a 100 que indica el grado de idoneidad de dicho ambiente para la especie). Maxent tiene propiedades que lo hacen muy robusto cuando se tienen cantidades limitadas de datos de entrenamiento (muestras pequeñas). Se ha desarrollado un software específico para el modelamiento de nichos ecológicos basado en este concepto, el cual cuenta con más de 1000 aplicaciones desde el año 2006 (Merow, Smith, & Silander, 2013) y está disponible en internet de forma gratuita (<https://www.cs.princeton.edu/~schapire/maxent/>).

2.3.1.3.1.1 Desarrollo matemático de Maxent

La mayor parte de los resultados de Maxent se han representado sobre una doble perspectiva, la geográfica y la medio-ambiental. En otras palabras, en Maxent la argumentación está dada en ocasiones sobre las celdas o píxeles del espacio geográfico y en otras sobre las variables medio ambientales, siendo ambas perspectivas válidas y complementarias a la vez.

2.3.1.3.1.1.1 Perspectiva geográfica de Maxent

Dado un espacio X que representa una región geográfica de interés. Típicamente X está definida por un conjunto discreto de celdas y se asume que es finito. Sea $\{x_1, x_2, \dots, x_m\}$ el conjunto que representa las localidades donde la especie ha sido observada y registrada (Phillips, Dudik, & Schapire, 2004). Finalmente, sea F el conjunto de variables ambientales del área X , tal como: precipitación, temperatura, altura, y otros.

Problema: Dado el conjunto $\{x_1, x_2, \dots, x_m\}$ seleccionados independientemente para alguna distribución π , se debe construir una distribución π' que aproxime π .

Para construir π' , se hace uso del conjunto $F = \{f_1, f_2, \dots, f_j\}$ donde $f_j : X \rightarrow \mathbb{R}$

Sea \mathbf{f} el vector de características para cualquier función $\mathbf{f} : X \rightarrow \mathbb{R}$.

Sea $\pi(\mathbf{f})$ la esperanza (el valor más esperado de una distribución) de las características \mathbf{f} bajo π .

$$\pi[f] = \sum_{x \in X} \pi(x) f(x) \quad (4)$$

Se asume que la información que se tiene de π está caracterizado por las esperanzas de las características bajo π .

La esperanza promedio de f_j es:

$$\tilde{\pi}[f_j] = \frac{1}{m} \sum_{i=1}^m f_j(x_i) \quad (5)$$

Se espera que $\tilde{\pi}[f_j]$ sea muy cercana a $\pi[f_j]$ y por lo tanto una aproximación de π' bajo cualquier valor de f_j tiene que ser igual a $\tilde{\pi}$, es decir:

$$\pi'[f_j] = \tilde{\pi}[f_j] \quad (6)$$

La igualdad (6) viene dada por el principio de máxima entropía que establece: la mejor aproximación a una distribución no conocida es alguna cuya entropía sea máxima (la más dispersa o cercan a la uniforme) sujeta a restricciones conocidas. Estas restricciones se definen en función del valor esperado de la distribución, la cual es estimada por medio de un conjunto de observaciones de presencia (Villar Hernández, 2014).

La entropía de p sobre X está definida por:

$$H(p) = - \sum_{x \in X} p(x) \ln p(x) \quad (7)$$

Propiedades:

- $H(p) \leq \ln(|X|)$
- Menor entropía, el fenómeno es más determinístico. Mayor entropía el fenómeno es más caótico.
- Menor entropía, poca información del fenómeno. Mayor entropía, mucha información del fenómeno.
- La idea es: estimar π por la distribución π' de máxima entropía sujeta a las restricciones de (6).

En terminología matemática:

$$\pi = \max H(\pi') \quad (8)$$

Sujeta

a:

$$\pi'[f_j] = \tilde{\pi}[f_j] \quad \forall j \in \{1, 2, \dots, n\}$$

Ahora, desde otro punto de vista se considera que todas las distribuciones de Gibbs son de la forma:

$$q_\lambda(x) = \frac{e^{\lambda \cdot f(x)}}{Z_\lambda}, \quad \lambda \in \mathbb{R}^n \quad (9)$$

Donde Z_λ es la constante de normalización para que la suma total sea igual a 1; la cual está definida por:

$$Z_\lambda = \sum_{x \in X} e^{\lambda \cdot f(x)} \quad (10)$$

Pietra *et al* (1997) probó que la maximización de $H(\pi')$ es equivalente a la distribución de Gibbs de máxima verosimilitud, es decir, la distribución q_λ que minimiza $RE(\tilde{\pi} \parallel q_\lambda)$ donde:

$$RE(\tilde{\pi} \parallel q_\lambda) = \sum_{x \in X} \tilde{\pi}(x) \ln\left(\frac{\tilde{\pi}(x)}{q_\lambda(x)}\right) \quad (11)$$

Llamada entropía relativa o divergencia de Kullback-Leibler. Resolviendo dicha ecuación el resultado es:

$$RE(\tilde{\pi} \parallel q_\lambda) = H(\tilde{\pi}) + \tilde{\pi}[-\ln(q_\lambda)] \quad (12)$$

Notar que la función negativa logarítmica de verosimilitud está dada por:

$$\tilde{\pi}[-\ln(q_\lambda)]$$

Se observa que solo difiere de $RE(\tilde{\pi} \parallel q_\lambda)$ por el término $H(\tilde{\pi})$ que es conocido, por lo tanto se usa la segunda como función objetivo para la minimización.

Suponer sin pérdida de generalidad que cada característica f_i se encuentra acotada entre $[0,1]$. Para calcular los λ 's, se utilizó un algoritmo de actualización-secuencial que modifica un solo peso de λ_j (Collins, Schapire, & Singer, 2002).

El enfoque que se expuso anteriormente sirve para calcular la distribución de máxima entropía π' con las restricciones (6). Sin embargo, no se espera que (6) se cumpla, pero sí que este muy cerca. Por lo tanto se amortigua sus restricciones de la siguiente forma:

$$|\pi'[f_j] - \tilde{\pi}[f_j]| \leq \beta_j \quad (13)$$

donde β_j es un estimado de cuan cerca la esperanza empírica $\tilde{\pi}[f_j]$, está de su esperanza verdadera $\pi[f_j]$

Con la restricciones de (13) la función objetivo a minimizar es:

$$\tilde{\pi}[-\ln(q_\lambda)] - \sum_{j=1}^n \beta_j |\lambda_j| \quad (14)$$

Esto se llama las l -regularizaciones. La regularización fuerza a Maxent a enfocarse sobre las características más importantes, y las l -regularizaciones tienden a producir modelos con pocos valores de cero para λ_j (Phillips, Anderson, & Schapire, 2006). Tales modelos son menos probables de tener sobreajuste, porque tienen pocos parámetros; y como regla general, la explicación más simple de un fenómeno es usualmente la mejor (el principio de parsimonia, Navaja de Ockham).

2.3.1.3.1.1.2 Enfoque medioambiental de Maxent

Un observador toma un sitio aleatorio x de un conjunto X de sitios en el área de estudio y reporta 1 si la especie está presente y 0 si está ausente, se denota la variable respuesta como y , entonces $\pi(x)$ es la probabilidad condicional $P(x | y = 1)$, es decir, la probabilidad del observador estando en un localidad x encuentre a la especie. De acuerdo a la regla de Bayes (Elith, et al., 2011).

$$P(y = 1|x) = \frac{P(x|y = 1)P(y = 1)}{P(x)} = \frac{\pi(x) P(y = 1)}{|X|} \quad (15)$$

de acuerdo con la estrategia de muestreo $P(x) = 1/|X|$ para todo x . Aquí, $P(y = 1)$ es el total de la prevalencia de la especie (proporción de los lugares o pixeles ocupados por la especie) en el área de estudio. La probabilidad $P(y = 1|x)$ es la probabilidad de que la especie esté presente en el sitio x . Sin embargo solo con datos de presencia no se puede determinar la prevalencia de la especie (Phillips, Anderson, & Schapire, 2006). Por lo tanto, en lugar de estimar directamente $P(y = 1|x)$, se estima la distribución π . Note que x es un sitio más que un vector de condiciones medioambientales.

Ahora:

$$\begin{aligned} P(y = 1|z) &= \frac{P(z|y = 1)P(y = 1)}{P(z)}, \text{ teorema de Bayes} \\ &= \frac{\sum_{x \in X(z)} P(x|y = 1)P(y = 1)}{\sum_{x \in X(z)} P(x)} \\ &= \frac{\sum_{x \in X(z)} P(y = 1|x)P(x)}{|X(z)|/|X|} \end{aligned} \quad (16)$$

Como $P(x) = 1/|X|$

$$= \frac{\sum_{x \in X(z)} P(y = 1|x)}{|X(z)|}$$

Existe una relación entre el espacio ambiental y el espacio geográfico. Por lo tanto:

$$P(y = 1|z) \approx q_\lambda(x) P(y = 1) |X| \quad (17)$$

Sin embargo, esta aproximación tiene dos dificultades: No se sabe o no se puede estimar $P(y = 1)$ dado que esta cantidad no se determina con valores de presencia (Ward, Hastie, Barry, Elith, & Leathwick, 2009). Segundo, la aproximación podría resultar más grande que la probabilidad, puesto que Maxent no garantiza que $q_\lambda(x)$ es más pequeña que $1/P(y = 1)|X|$.

Para resolver estas dificultades se aplica de nuevo el principio de máxima entropía. En lugar de aplicar el principio para estimar una distribución sobre los sitios, se aplica éste en la distribución conjunta $P(x,y)$ representando ambos un muestreo de la distribución sobre los sitios (asumiendo la igualdad de los puntos para toda la colección y evaluación) y la presencia/ausencia de la especie. En particular se estima $P(x,y)$ por la distribución $Q(x,y)$ de máxima entropía sujeta a restricciones sobre la distribución condicional $P(x|y = 1)$, es decir, las mismas contradicciones que se aplicó para estimar π . Una vez obtenida la distribución conjunta Q , se tiene la información suficiente para derivar la distribución condicional $Q(y = 1|x)$, que resulta:

$$Q(y = 1|x) = \frac{e^H q_\lambda(x)}{1 + e^H q_\lambda(x)} \quad (18)$$

donde q_λ es la estimación de máxima entropía de π y H es la entropía de q_λ .

Para explicar mejor el concepto de máxima entropía, Martínez (2010) planteo el siguiente ejemplo:

Suponiendo la existencia de 3 cajas de manzanas, las cuales están cerradas, se conoce únicamente que en total hay nueve manzanas dentro de ellas (información parcial); a partir de estos supuestos se pretende responder a la pregunta: ¿Cuál es la manera más probable que estén distribuidas las manzanas? La respuesta es la distribución de máxima entropía. Ahora, la pregunta es: ¿Por qué?

Según la fórmula de Shannon (1948) la entropía está definida por:

$$S = - \sum n_j \ln (n_j) \quad (19)$$

Donde n_j es igual al número de manzanas en las cajas. Por tanto la distribución de máxima entropía es (3,3,3). Ejemplo:

Caja 1	Caja 2	Caja 3	Entropía
3	3	3	-9.8
1	5	3	-11.3
0	1	8	-16.6

Se observa que la mayor entropía corresponde a tener las tres cajas con tres manzanas cada una (distribución uniforme y la más probable). Pero se puede poner ciertas restricciones en la forma de organizar las manzanas, por ejemplo: Se pide que además que la distribución sea de máxima entropía cumpla con que el número de manzanas en la segunda casilla sea de 5. Entonces:

Caja 1	Caja 2	Caja 3	Entropía
2	5	2	-10.8
1	5	3	-11.3
4	5	0	-13.6

La distribución que maximiza la entropía es la primera (2, 5, 2), que es una distribución más cercana a la distribución uniforme y la más probable bajo estas condiciones.

Ahora, en lugar de cajas se tienen pixeles (de un área de estudio determinada) y en lugar de manzanas se tienen puntos de presencias de especies, y las restricciones son los

valores empíricos promedio de las llamadas “características” de la información disponible; en este caso variables bioclimáticas. Es decir, los píxeles del área de estudio son el espacio de donde la probabilidad de distribución de máxima entropía es definida. Los píxeles con presencia de una especie (records) constituyen los puntos de muestreo y las características son las variables climáticas y ecológicas (Martínez, 2010).

La información disponible sobre la distribución de los valores bioclimáticos asociados con la presencia de especies, se presenta como un conjunto de valores de variables reales, llamadas "características", y las restricciones son: que el valor esperado para cada característica debe coincidir con su valor empírico promedio (valor promedio de un conjunto de puntos de muestreo tomado de la distribución objetivo o de destino) (Martínez, 2010).

De acuerdo a lo anterior, Maxent estima distribuciones que deben de estar de acuerdo con todo lo que se conoce (aunque sea de manera incompleta) de la información inferida de las condiciones ambientales de las localidades de ocurrencia y evitar restricciones infundadas. Maxent entonces trata de encontrar la distribución de probabilidad de máxima entropía (cercana a la uniforme) sujeta a las limitaciones impuestas por la información disponible sobre la distribución observada de las especies y las condiciones ambientales en el área de estudio (Martínez, 2010).

Maxent computa una distribución de probabilidades basado en las variables ambientales de toda el área de estudio. Si el área es muy grande ($> 600\ 000$ píxeles) se toma una muestra aleatoria de unos 100 000 píxeles “background” para representar las condiciones ambientales de la región (Martínez, 2010).

Además del principio de máxima entropía, Maxent necesita de un algoritmo que le permita encontrar las distribuciones con mayor entropía. Este algoritmo llamado "Sequential update algorithm" (Dudik, Phillips, & Schapire, 2004). Utiliza iteraciones en donde va dando distintos pesos a las variables utilizadas y va ajustándolas. Es

un algoritmo determinístico que según los autores y distintas pruebas empíricas garantiza que convergerá en la distribución de probabilidades de máxima entropía. Al terminar el proceso de iteración Maxent asigna una probabilidad negativa a cada pixel del área total de estudio, que al final deben sumar 1, por lo que se aplica un valor de corrección para hacerlos positivos y que sumen entre todos 100%. Pero como cada pixel presenta valores muy pequeños, Maxent los presenta con un valor que es el resultado de la suma del valor de ese pixel y de todos los demás pixeles con un valor de probabilidad igual. Esos valores pueden ir de 0 a 100 e indican probabilidad de ocurrencia de la especie (Martínez, 2010).

2.3.1.3.2 Modelo de regresión logística.

Los datos ecológicos frecuentemente violan los supuestos del modelo lineal. Los modelos lineales generalizados (GLM) son extensiones del modelo lineal que pueden trabajar con variables respuestas que no provienen de una distribución normal (Franklin, 2009).

En el modelo GLM, las variables predictoras $X_i (i = 1, \dots, p)$ son combinadas para producir un predictor lineal el cual se relaciona con el valor esperado $\mu = E(Y)$ de la variable respuesta Y a través de una función de enlace $g()$, tal como:

$$g(E(Y)) = \mathbf{X}_i^T \boldsymbol{\beta} \quad (20)$$

Donde $\mathbf{X}_i^T \boldsymbol{\beta} = (\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})$ La función $g()$ tiene que ser una función monótona y diferenciable.

Las principales mejoras del modelo GLM sobre el modelo de regresión general son por lo tanto: La capacidad de manejar una clase más grande de distribuciones para la variable respuesta Y . Aparte de la distribución normal, otras distribuciones son la binomial, la Poisson y la Gamma (Guisan et al., 2002).

La relación de la variable respuesta Y con el predictor lineal a través de la función de enlace $g(E(Y))$. Además de garantizar la linealidad, ésta es una forma eficiente de condicionar las predicciones a estar dentro de un rango de valores posibles para la variable respuesta (ver Figura8); por ejemplo, entre 0 y 1 para probabilidades de presencia (Guisan et al., 2002).

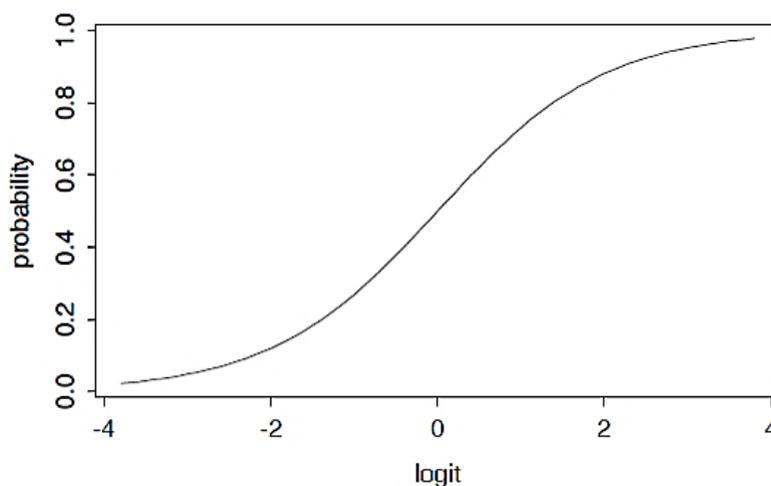


Figura 8. Curva logística.

Un modelo GLM con una familia binomial (liga logit) es conocido como regresión logística y es muy utilizado cuando se modelan distribuciones de especies con respuestas binarias (presencia-ausencia).

En algunas ocasiones se disponen de registros binarios para modelar distribuciones de especies, estos registros denominados también de presencia-ausencia son tratados como éxitos y fracasos respectivamente. La variable respuesta se asocia con un conjunto de variables predictivas y se considera que las respuestas provienen de una secuencia de ensayos independientes Bernoulli, donde cada ensayo tiene su propia probabilidad de éxito que depende de los valores de las variables predictoras.

Definimos la variable respuesta $y_i = 1$ cuando el i -ésimo ensayo resulta exitoso, y $y_i = 0$ cuando no lo es. Por tanto, y_i tiene la distribución *binomial* $(1; p_i)$, donde p es la probabilidad de éxito en el i -ésimo ensayo.

Dicho lo anterior, se requiere encontrar un modelo de regresión para predecir los éxitos utilizando la variable predictora, sin embargo, note que una función lineal de la variable predictora $\mathbf{X}_i^T \beta$ estará en el intervalo $-\infty$ a ∞ mientras que p_i , la probabilidad de éxito, estará siempre entre 0 y 1. Por lo tanto no podemos igualar a la función lineal del predictor directamente con la probabilidad de éxito. En lugar de ello, es necesario encontrar una función de la probabilidad de éxito que también oscile entre $-\infty$ a ∞ que podamos ligar con la función lineal del predictor.

Supongamos que decidimos usar el logaritmo de la razón de probabilidades conocida como función de enlace *logit*. La igualamos a la función lineal del predictor

$$\log\left(\frac{p_i}{1-p_i}\right) = \mathbf{X}_i^T \beta \quad (21)$$

y resolvemos para p_i como función de X_i . Note que p_i es la cantidad que interesa al investigador, es decir, la probabilidad de ocurrencia de la especie.

Es importante recalcar ciertos supuestos que se manejan dentro del modelo logístico como:

- La i -ésima observación tiene la distribución *binomial* $(1, p_i)$. Cada una con su propia probabilidad de éxito.
- El *logit* está vinculado al predictor lineal, una función lineal desconocida de las variables predictoras.

$$\log\left(\frac{p_i}{1-p_i}\right) = \mathbf{X}_i^T \boldsymbol{\beta}$$

Resolviendo para p_i se obtiene:

$$p_i = \frac{e^{\mathbf{X}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{X}_i^T \boldsymbol{\beta}}} \quad (22)$$

$$\mathbf{X}_i^T \boldsymbol{\beta} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_{ip} \quad (23)$$

Que relaciona la probabilidad de éxito con el valor de las variables predictoras, donde las observaciones son independientes unas de otras.

2.3.1.3.3 Evaluación multi-criterio (EMC) y lógica difusa (*Fuzzy*).

La evaluación multicriterio puede definirse como como un conjunto de técnicas orientadas a asistir en los procesos de toma de decisiones, cuyo fin básico es investigar un número de alternativas bajo la luz de múltiples criterios y objetivos en conflicto (Gómez & Barredo, 2005).

La metodología utilizada para realizar una evaluación multicriterio en términos generales es la siguiente: inicialmente se establece el conjunto de soluciones posibles del problema de decisión analizado, seguidamente basándose en un criterio, se asocia a cada solución posible un código (un valor) que cuantifica el grado de deseabilidad que tiene cada alternativa para el centro decisor, luego, utilizando técnicas matemáticas se procede a determinar entre las soluciones posibles, aquella o aquellas que poseen el mayor grado

de deseabilidad obteniendo así la solución óptima del problema planteado, en base a un determinado criterio (Gómez & Barredo, 2005).

Existen diferentes métodos o técnicas dentro de la EMC, estas se diferencian básicamente en los procedimientos matemáticos o estadísticos que realizan sobre las matrices de evaluación y prioridades, para obtener una evaluación final de las alternativas. Jankowski (1995) clasificó a estos métodos en dos grandes grupos: técnicas compensatorias y técnicas no compensatorias; que se diferencian esencialmente en el grado cognitivo que demandan del centro decisor. Finalmente, en los últimos años se ha incrementado el uso de técnicas compensatorias adaptadas a la filosofía de la lógica borrosa o fuzzy.

La lógica borrosa o teoría de los conjuntos borrosos (fuzzy set theory) desarrollada por el matemático L.Zadeh (1965) es una generalización del álgebra booleana que permite representar el pensamiento humano que no es exclusivo de lo cuantitativo sino también lingüístico cualitativo en lenguaje matemático (Almache, 2013).

Bajo esta lógica subyace la teoría de que el éxito del cerebro humano radica, en parte, en su capacidad de procesar de manera muy eficiente información borrosa e imprecisa. Lo que se pretende con esta teoría es trasladar este éxito a la computación; así, mediante la lógica borrosa es posible trasladar las expresiones naturales de conocimiento y sentido común basadas en el lenguaje en un formalismo matemático preciso (Sainz de la Maza, 2000).

La teoría de conjuntos borrosos representa una clase de elementos u objetos sin límites bien definidos entre los objetos que pertenecen a esa clase y los que no. Cada conjunto borroso está especificado mediante una función de pertenencia (*fuzzy membership function*), que se encuentra en el dominio de los números reales y cumple con la propiedades de normalidad (el valor máximo y mínimo de pertenencia a un conjunto borroso es 1 y 0) y convexidad (los valores de la función de pertenencia no

aumentan a medida que se realiza un desplazamiento tanto a la derecha como la izquierda del valor mayor de la función), y el grado de pertenencia de cualquier elemento a una determinada clase puede ser cualquier número entre 1 y 0, cuanto más se acerque a uno, más claramente pertenece a esa clase. La capacidad de un conjunto borroso de expresar transiciones graduales desde la pertenencia a la no pertenencia, ofrece un gran abanico de posibilidades para la representación geográfica de entidades con límites imprecisos (Gómez & Barredo, 2005).

2.3.1.3.3.1 Construcción de un modelo basado en conjuntos borrosos

Según Openshaw (1997) cualquier modelo basado en la lógica difusa tendrá las siguientes fases:

- a) Especificar o definir la naturaleza del sistema a modelar; la cual, es una de las fases más importantes ya que se requiere un buen conocimiento del tema y toda la información de respaldo para poder expresarlo en términos lingüísticos.
- b) Identificar y etiquetar las funciones de pertenencia para cada variable de entrada. Consiste en convertir los términos lingüísticos en los que han sido definidas las variables en conjuntos borrosos, en términos más concretos se evalúa la posibilidad de que cada elemento o píxel pertenezca a un determinado sistema o fenómeno (Padilla, 2007). La forma de un conjunto borroso es flexible ya que hay disponible un amplio orden de funciones de pertenencia para definir los grados en que cada elemento considerado pertenece al conjunto (Sainz de la Maza, 2000). Entre las funciones más destacadas existen las funciones lineales, en forma de *J* (J-shaped) y las funciones sigmoidales (S-shaped) (ver Figura 10), esta última, es la más utilizada en la teoría de conjuntos borrosos y es la más apropiada para estudios de distribuciones de especies. Para construir cualquier función es necesario especificar la posición de cuatro puntos de inflexión en el eje de las X (donde es representada la variable en cuestión), que gobierna la

forma de la curva. Padilla (2007) resalta la importancia de definir los puntos de inflexión mediante el análisis de la función seno cuadrado (ver ecuación 24).

$$P = f(\text{Sen}^2\alpha) \quad (24)$$

Donde, $\text{Sen}^2 \alpha$ que se encuentra dentro del rango $[0,1,0]$ (ver Figura 9), P es la probabilidad de que ocurra el evento y α es la variables en función del evento; los puntos de inflexión serán:

- Punto X_A : Punto de valor minino en la variable a modelar
- Punto X_B y X_C : Rango de valores donde se encuentran la mayor cantidad de datos de la variable
- Punto X_D : Punto de valor máximo en la variable

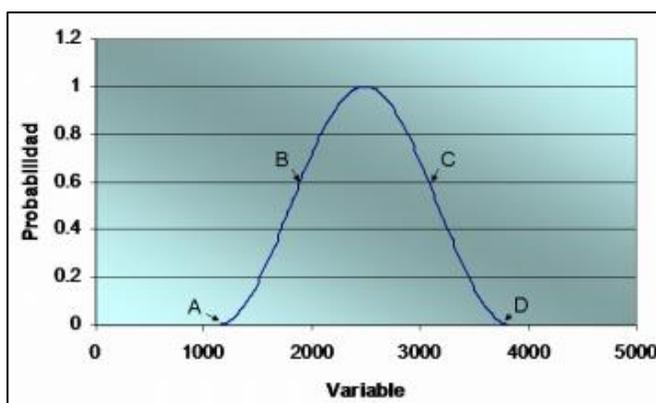


Figura 9. Función $\text{Sen}^2 \alpha$
Fuente: Padilla (2007)

Los puntos de inflexión se calculan con la segunda derivada de la función, por lo tanto:

$$P'' = 2 (\text{Cos}^2\alpha - \text{Sen}^2\alpha) \quad (25)$$

Resolviendo mediante identidades trigonométricas:

$$P'' = 2 (\text{Cos}^2 \alpha - \text{Sen}^2 \alpha)$$

$$\text{Cos}^2 \alpha - \text{Sen}^2 \alpha = 0$$

$$1 - \text{Sen}^2 \alpha - \text{Sen}^2 \alpha = 0$$

$$1 - 2 \text{Sen}^2 \alpha = 0$$

$$\text{Sen}^2 \alpha = 1/2$$

$$\alpha_1 = \pi/4$$

$$\alpha_2 = 3\pi/4$$

Finalmente:

P₁ (X_A ; 0). Punto de valor minino en la variable a modelar

P₂ (X_B ; $\pi/4$). Punto de inicio de la zona en la que existe mayor cantidad de datos de la variable.

P₃ (X_C ; $3\pi/4$). Punto de fin de la zona en la que existe mayor cantidad de datos de la variable

P₄ (X_D ; π). Punto de valor minino en la variable a modelar

P₅ (X_B + X_C / 2 ; $\pi/2$). Corresponde al punto medio de la curva, donde el valor es máximo en la variable.

La elección de una función de pertenencia es una elección subjetiva y esta es una de las críticas más frecuentes a esta metodología; sin embargo, dicha elección, ofrece la oportunidad de plantear una hipótesis sobre la relación de la especie con su entorno (Sainz de la Maza, 2000).

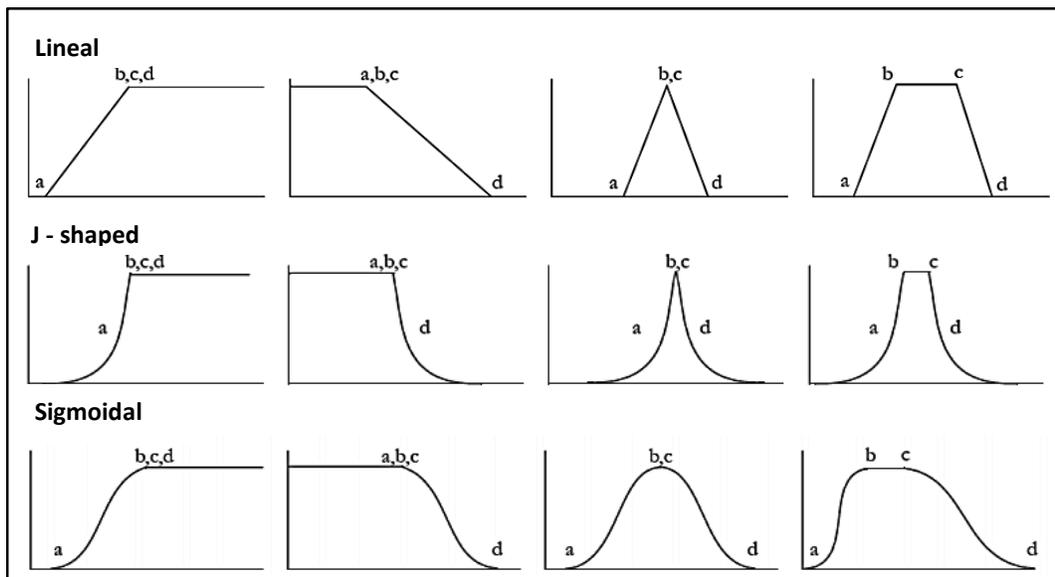


Figura 10. Funciones de pertenencia borrosa.
Fuente: Eastman (2015).

- c) Creación de las reglas: un modelo borroso consta de ciertas reglas del estilo IF-THEN (si-entonces), estas reglas se las denominan “Memoria Asociativa Borrosa” (*Fuzzy Associative Memory, FAM*) y no es más que la transformación de las variables de entrada en las variables de salida.
- d) Aplicación del modelo: Una vez definidas las funciones de pertenencia y las reglas, se ejecuta el modelo. Los operadores que normalmente se aplican con números borrosos son el operador mínimo (equivalente al operador lógico Y) y el operador máximo (equivalente al operador lógico O), adicionalmente se pueden utilizar realizar otras operaciones de superposición y operaciones aritméticas.
- e) Conversión de los resultados a valores rígidos (*crisp*) o defusificación (*defuzzification*): de esta forma los resultados del modelo serán valores numéricos que representen mejor la información contenida en el conjunto borroso resultante.
- f) Evaluación del modelo.

2.3.1.4 Validación de los modelos de nicho ecológico.

La evaluación de cualquier clase de modelo ecológico puede ser enfocada de muchas formas como verificar la lógica de las formulaciones conceptuales, evaluar el ajuste que tiene sobre los datos o su habilidad de predecir (Franklin, 2010). Dentro de los modelos de nicho ecológico la validación de un modelo consiste en contrastar sus predicciones respecto a datos considerados independientes para obtener una medida de su significancia en relación con una cierta hipótesis nula, o su desempeño (capacidad de clasificar correctamente nuevos datos) (Soberón, 2015), además que permite realizar una comparación entre modelos (Pearson, 2008).

Todos los modelos tienen errores de predicción porque son simplificaciones de la realidad (Franklin, 2010) y es imposible controlar todos los factores que gobiernan un fenómeno. La medida de la bondad de un método estará normalmente fundamentada en una evaluación del error que en modelización se mide como la proporción de casos clasificados incorrectamente por el modelo (Fielding & Bell, 1997). Hay dos tipos de error: el de comisión, que consiste en clasificar una ausencia como presencia; y el de omisión, consistente en clasificar una presencia como ausencia (Felicísimo, et al., 2011).

La llamada “Matriz de confusión” (ver Figura 10) resume el desempeño del modelo en términos de errores y aciertos. La matriz de confusión registra las frecuencias de cada uno de los cuatro posibles tipos de predicción obtenidos a partir de los datos de entrenamiento: (a) verdaderos positivos (el modelo predice que la especie está presente y los datos de validación comprueban que realmente está), (b) falso positivo (el modelo predice como presencia pero los datos de testeo muestran como ausencia), (c) falso negativo (el modelo predice ausencia pero el test lo muestra como presencia), (d) verdadero negativo (el modelo predice como ausencia y la evaluación lo muestra, en efecto, como ausencia) (Pearson, 2008).

	Presencias reales	Ausencias reales
Presencias predichas	a <i>(verdadero positivo)</i>	b <i>(falso positivo)</i>
Ausencias predichas	c <i>(falso negativo)</i>	d <i>(verdadero negativo)</i>

Figura 11. Matriz de confusión.

Fuente: Pearson (2008)

Cada elemento de la matriz de confusión puede ser visualizado en el espacio geográfico como se ilustra en la Figura 12.

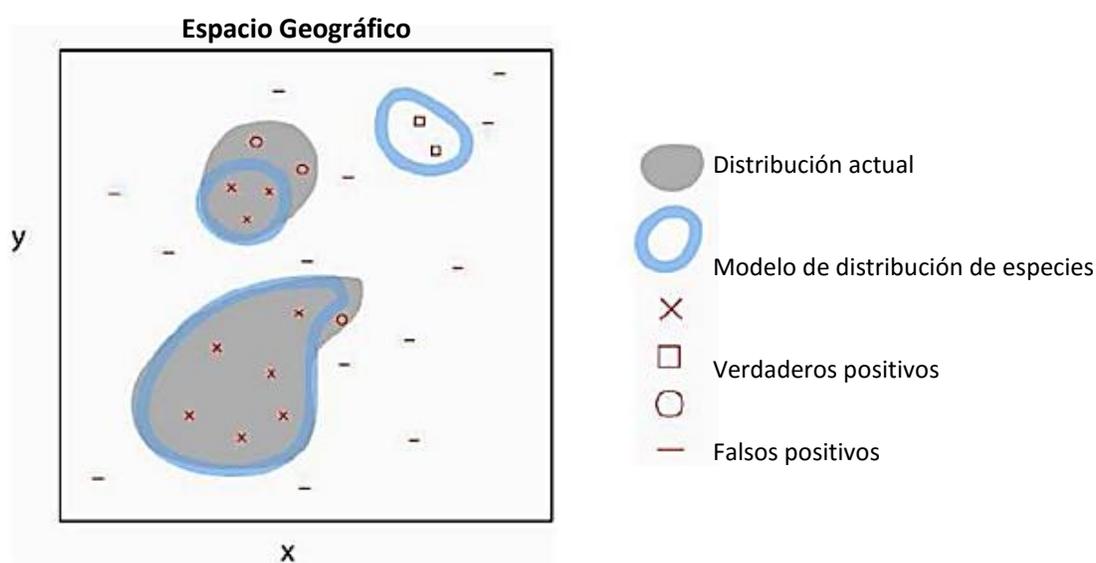


Figura 12. Matriz de confusión representada en el espacio geográfico

Fuente: Pearson (2008)

Fielding & Bell (1997) proporcionan buenas revisiones de los diferentes estadísticos utilizables para evaluar la calidad de los modelos basados en la matriz de confusión: ROC (Receiver Operating Characteristics), AUC (Area Under the Curve), validación cruzada, bootstrapping, kappa de Cohen, y otros (ver tabla 4).

Tabla 4.

Parámetros estadísticos derivados de la matriz de confusión.

Parámetro	Descripción	Fórmula
PCC	Porcentaje correctamente clasificado	$(a + d) / n$
Kappa	Relación entre la precisión de la predicción y la predicción al azar.	$\frac{[(a + d) - (((a + c)(a + b) + (b + d)(c + d))/n)]}{[n - (((a + c)(a + b) + (b + d)(c + d))/n)]}$
Sensibilidad	Proporción de las observaciones de presencia correctamente predichas	$a / (a + c)$
Especificidad	Proporción de las observaciones de ausencia correctamente predichas	$d / (b + d)$
PPP	Proporción de las predicciones de presencia correctamente predichas	$a / (a + b)$
NPP	Proporción de las predicciones de ausencia correctamente predichas	$d / (c + d)$
TSS	True skill statistic	$a / (a + c) + d / (b + d) - 1$

Fuente: Miller (2010)

2.3.1.4.1 Curva ROC (Receiver operating characteristic analysis)

El estadístico más empleado en la bibliografía para la validación de los modelos de nicho ecológico es la curva ROC (Felicísimo , et al., 2011). La curva ROC (Receiver operating characteristic analysis) es un tipo de prueba estadística que fue implementada por los operadores de radar durante la segunda Guerra mundial, pero han sido transportada a muchas áreas en donde es necesario caracterizar el rendimiento de un modelo y tomar decisiones (Martínez, 2010).

La curva ROC es la representación gráfica de la capacidad discriminativa de un modelo para todos sus posibles puntos de corte (ver Figura 13), y necesita que los datos que se evalúan sean de presencia/ausencia. El eje de ordenadas (*eje y*) representa el error de omisión, ya que corresponde a la sensibilidad o casos positivos bien clasificados, y el de abscisas (*eje x*) al error de comisión, ya que representa el complementario de la especificidad (1-especificidad), o falsos positivos. El estadístico derivado es el área bajo la curva ROC o AUC en su acrónimo inglés *Area Under the Curve*, que es equivalente al estadístico U de Mann-Whittney ya que en realidad corresponde a la probabilidad de que, tomado al azar un par de casos, uno presencia y otro ausencia (o pseudo-ausencia), el modelo adjudique a la presencia un valor mayor de idoneidad o probabilidad, y esto para todos los posibles pares en la muestra. El valor de AUC está comprendido entre 0-1; un valor de 1 indica que todos los casos se han clasificado correctamente y uno de 0.5 que el modelo no es diferente de clasificar los casos al azar; valores menores a 0.5 indican que el modelo es realmente malo, ya que clasifica erróneamente más casos que el azar (Felicísimo , et al., 2011).

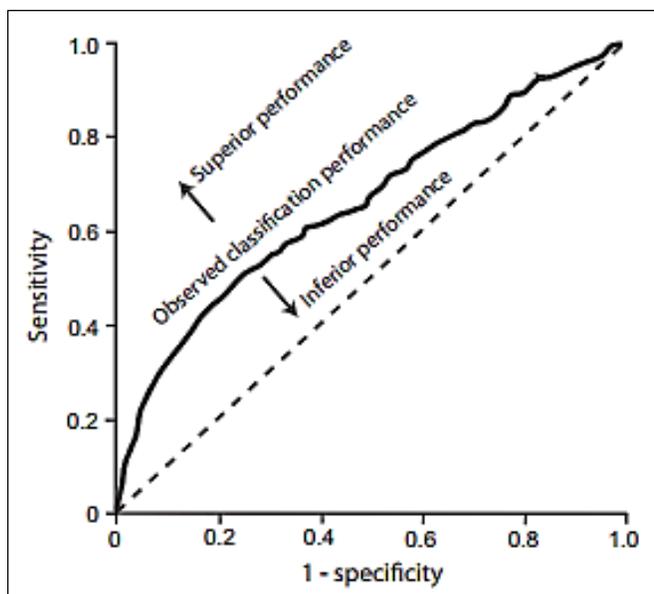


Figura 13. Curva ROC
Fuente: Peterson et al, (2011)

Las mayores ventajas del AUC son: (1) la posibilidad usarlo para comparar cualquier método, sea cual sea el tipo de valores de salida (probabilidad, idoneidad,

puntuaciones, y otros.), ya que no depende de presupuestos paramétricos, y sólo necesita que las distribuciones de estos valores sean monótonas; (2) sus resultados son independientes de la prevalencia; y (3) es una medida independiente del punto de corte, ya que su valor se genera utilizando todos los posibles puntos de cortes. (Felicísimo , et al., 2011)

Finalmente, los MNE son usualmente validados con datos independientes o datos nuevos tomados en campo (usualmente no es factible). El uso de los mismos datos para calibrar y para evaluar, procedimiento llamado “resustitución”, tiende a sobreestimar las predicciones para los nuevos datos; por ello, es muy común dividir o particionar los datos en dos grupos, uno utilizado para el entrenamiento o calibración del modelo (datos de entrenamiento o training) y otro para la validación de las predicciones (datos de testeo o testing) (Miller, 2010). Fielding & Bell (1997) discutieron el particionamiento de los datos para el modelamiento de la distribución de especies y sugirieron como regla principal que la proporción de los datos de testeo debe ser:

$$1/1 + \sqrt{P - 1} \quad (26)$$

Donde P es el número de variables predictoras; de esta forma, si se tienen dos predictores, la proporción será 50:50, si son cinco, será 67:33, y si son diez o más, la proporción será 75:25, para datos de entrenamiento y validación respectivamente.

2.4 Ecología de *Anopheles spp.*

En todo el mundo existen 3 500 especies de mosquitos, sin embargo un reducido grupo de aproximadamente 41 especies son las responsables de la transmisión de la malaria alrededor del mundo y en Sudamérica 9 de ellas son las principales; éste subconjunto se lo conoce como *Anopheles* (Sinka, et al., 2011).

El *Anopheles*, vector transmisor de la malaria, pertenece a la clase Insecta, orden Diptera, de la familia Culicidae, subfamilia Anophelinae, constituida por 3 géneros, aproximadamente 10 subgéneros y 532 especies en el mundo (SNEM, 2013)

Entre las especies más importantes que se hallan en el país son:

- *Anopheles albimanus*
- *Anopheles punctimacula*
- *Anopheles pseudopunctipennis*
- *Anopheles neivai*
- *Anopheles oswaldoi*
- *Anopheles rangeli*
- *Anopheles nuneztovari*
- *Anopheles trinkae*

El *Anopheles albimanus* es el principal vector en el litoral, *Anopheles psedopuctipennis* en los valles interandinos y *Anopheles oswaldoi*, *Anopheles trianulatus*, *Anopheles punctimacula* y *Anopheles trinkae* en la Amazonía (Muñoz & Recalde, 2010). Cabe mencionar que los vectores antes descritos no son los únicos transmisores de la malaria, además existen vectores secundarios (especies que en la naturaleza se han encontrado infectadas con el agente parasitario pero que por su baja densidad no se consideran importantes para la transmisión) o vectores potenciales (especies que han estado involucradas en la transmisión de la enfermedad ya sea en el pasado o en otros países); por ejemplo: *Anopheles fluminensis*, *Anopheles apicimacula*.

De acuerdo con Donnelly *et al* (2002) las especies anofelinas en general tienen las siguientes características: (a) amplia distribución geográfica (b) una alta frecuencia estacional y abundancia local (c) capacidad de colonización y dispersión (d) adaptaciones para explorar diferentes entornos modificados por el hombre (edificaciones, áreas cultivadas, y otros.) (d) hematofagia, se alimentan de sangre preferentemente humana (antropofilia).

El vector *Anopheles albimanus* (ver Figura 14) es uno de los más estudiados debido a su amplio rango de distribución desde América Central, Sudamérica y el Caribe. Esta especie ha sido encontrada en regiones costeras bajas entre 100-500 msnm, sin embargo ha sido reportada en lugares más altos (Guatemala 1 372 msnm, México 1 941 msnm, y Colombia 1000). Se encuentra en zonas con una precipitación anual mayor de 1000 mm y temperaturas medias anuales entre 22 y 28 °C, además su distribución geográfica coincide con áreas que poseen dos estaciones lluviosas con precipitaciones mayores a 1000 mm (Montoya-Lerma, et al., 2011).



Figura 14. *Anopheles albimanus*
Fuente: © James Gathany / CDC

Anopheles pseudopunctipennis ha sido colectada en las ecoregiones costera, piedemonte, llanos y los valles altos, mostrando una ocupación de nichos variable en comparación con cualquiera de las otras especies de anofelinas. Se encuentra a altitudes que van de 0 a 1800 msnm, en sitios con precipitación anual que varía de 654 a 1040 mm en el estado. En cuanto a la temperatura promedio anual, se ha reportado esta especie en áreas con temperaturas de 27-28 °C y otras menos cálidas de 24°C (Osborn, Rubio-Palis, Herrera, Figuera, & Moreno, 2004).

Anopheles oswaldoi es un vector de importante en las zonas selváticas de Perú, Ecuador y Brasil y está directamente relacionado *Anopheles darlingi* (Pinault & Hunter, 2011). Se encuentra en espacios geográficos con altitudes de hasta 1300 msnm,

precipitación anual comprendida entre 1000 y 3500 mm y temperatura media anual que varía entre 23 y 38°C (Osborn, Rubio-Palis, Herrera, Figuera, & Moreno, 2004)

En general, los vectores encargados de la transmisión de la malaria necesitan que se cumplan ciertos parámetros ambientales para poder desarrollarse y poder proveer un espacio adecuado para la producción de esporas del parásito. Se requiere de una temperatura entre 16°C y 33°C, siendo óptimo una temperatura ambiental entre 20°C y 27°C, además de una altitud menor a los 1800 msnm . Climáticamente, la humedad y la lluvia proporcionan un sitio agradable para los sitios de crianza de mosquitos, sin embargo, la lluvia en exceso puede destruir por efecto de flujo estos sitios y exterminar las larvas y crisálidos del mosquito. Es indispensable reconocer como la transmisión de malaria puede variar dependiendo tanto de los patrones de precipitación como la proximidad de los criaderos de mosquitos a las personas. (Soria, 2013).

En cuanto a movilidad, la dispersión de los anofelinos es de 1 kilómetro, aunque algunos pocos individuos pueden dispersarse un poco más. Los factores que influyen en la dispersión son: las características de la especie, la topografía, la dirección y la velocidad del viento, la densidad de mosquitos, la disponibilidad de hospederos y el tamaño de los sitios de reproducción (INS, 2010).

CAPÍTULO III

3 Metodología y Desarrollo

3.1 Esquema metodológico

Siguiendo la propuesta metodológica planteada por Franklin (2010), descrita en el Capítulo II, se creó un esquema propio; descrito en la Figura 15.

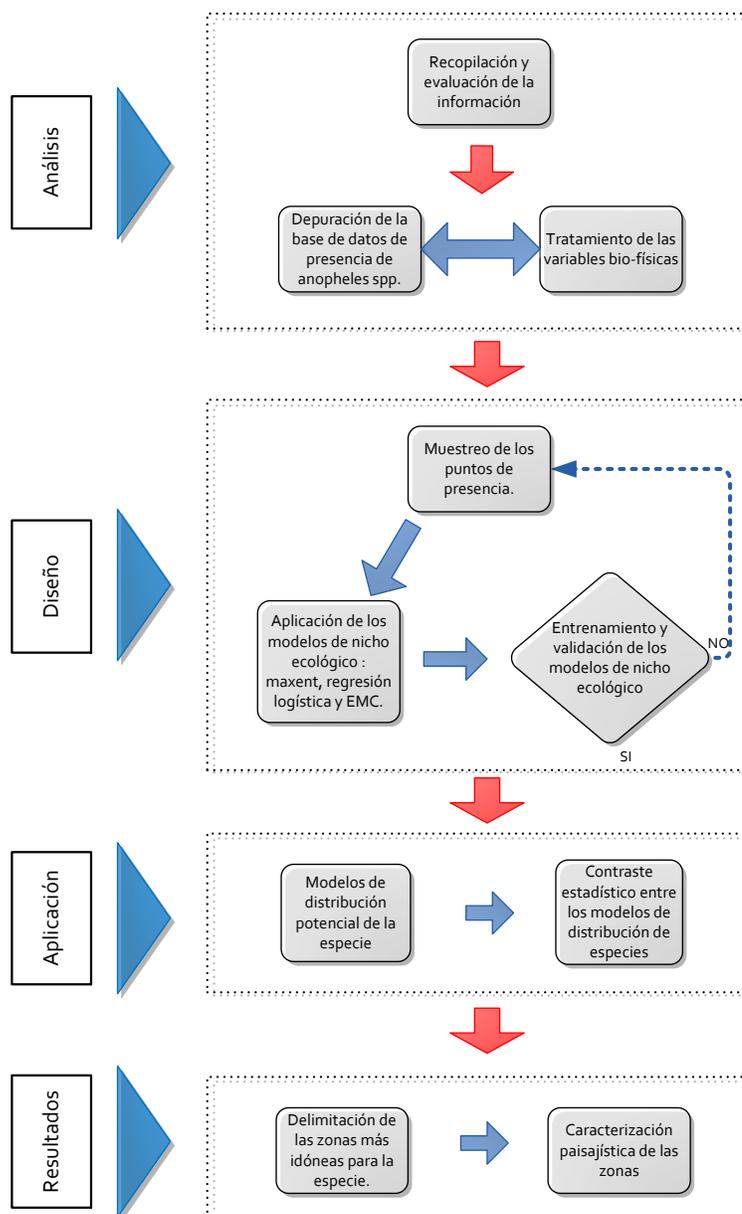


Figura 15. Esquema metodológico para el modelamiento de *Anopheles spp.*

3.2 Estructuración de la información

Para el modelamiento del nicho ecológico de una especie es fundamental contar con dos tipos de datos, los puntos de presencia de la especie y las variables medioambientales que describan a la misma.

3.2.1 Puntos de presencia de *Anopheles spp.*

En el presente proyecto se recopilaron registros de presencia georeferenciados de diferentes especies del género *Anopheles* que son potencialmente transmisoras de la malaria dentro del Ecuador continental. La información se obtuvo de tres fuentes diferentes. A través de dos bases de datos de libre acceso, la primera Global Biodiversity Information Facility-GBIF (<http://data.gbif.org/welcome.htm>) y la segunda Vector Map (<http://www.vectormap.org/>), que contienen colecciones de datos provenientes de museos, institutos y de estudios previos llevados a cabo en todo el mundo, se obtuvieron 395 y 364 registros de presencia respectivamente. Cabe mencionar que son registros históricos comprendidos entre los años 1965 y el año 2007.

La tercera fuente de información fue la investigación *New highland distribution records of multiple Anopheles species in the Ecuadorian Andes* llevada a cabo por Pinault & Hunter (2011), en la cual se colectaron 60 individuos de diferentes especies de anophelinos en todo el Ecuador durante el periodo 2008-2010.

Para el propósito del modelamiento llevado a cabo en este estudio, no es necesario contar con datos de abundancia (número de individuos en cada localidad) por lo que se procedió a depurar los puntos de presencia en función de su localización; mediante el uso de la herramienta *Delete Identical* del software ArcGIS 10.2 se eliminó de la base de datos total de individuos los registros cuyas coordenadas (latitud y longitud) estén repetidas. Por último se realizó una transformación de las coordenadas de los puntos de coordenadas geográficas (latitud y longitud) al sistema de coordenadas proyectadas

UTM (X, Y) en la zona 17 Sur, elipsoide WGS84. De esta forma, se obtuvo una base de datos con un total de 144 registros de presencia de *Anopheles spp.* para el desarrollo de los modelos de nicho ecológico, en la Tabla 5 se resume las especies encontradas con sus respectivas frecuencias.

Tabla 5.

Especies del genero *Anophles* para el modelamiento de nicho ecológico.

Espece	Número de individuos
<i>Anopheles albimanus</i>	26
<i>Anopheles apicimacula</i>	3
<i>Anopheles bambusicolus</i>	2
<i>Anopheles calderoni</i>	6
<i>Anopheles eiseni</i>	3
<i>Anopheles fluminensis</i>	9
<i>Anopheles lepidotus</i>	1
<i>Anopheles neivai</i>	3
<i>Anopheles oswaldoi</i>	23
<i>Anopheles pseudopunctipennis</i>	23
<i>Anopheles punctimacula</i>	14
<i>Anopheles rangeli</i>	16
<i>Anopheles sp.</i>	7
<i>Anopheles squamifemur</i>	1
<i>Anopheles triannulatus</i>	1
<i>Anopheles trinkae</i>	6
TOTAL	144

En la Figura 16 se muestra los puntos de presencia de *Anopheles spp.* en el Ecuador continental

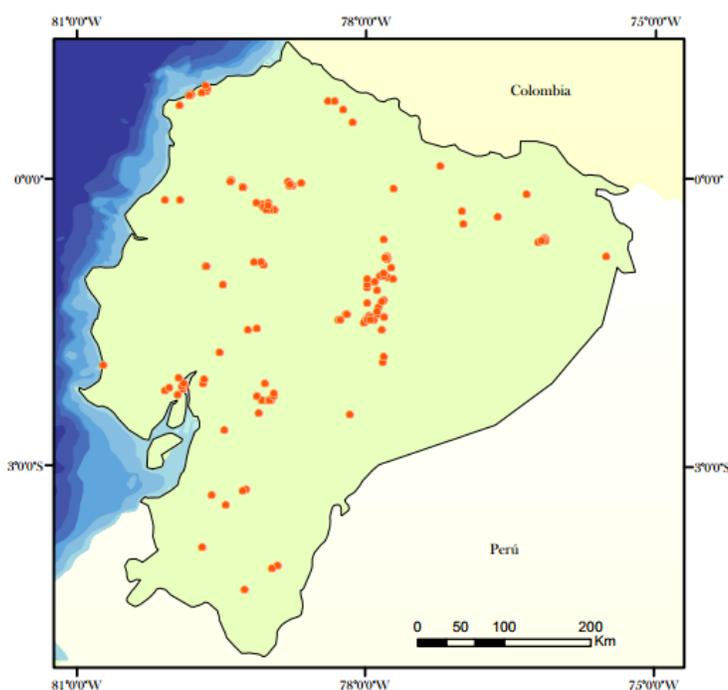


Figura 16. Puntos de presencia de *Anopheles* spp. en el Ecuador

3.2.2 Variables Predictoras

En cuanto a las variables medioambientales y se utilizó un conjunto de datos en formato raster que contienen información climática global con una resolución espacial de 1 km, provenientes de la base de datos de WORLDCLIM, las cuales son usadas para modelamiento espacial. WORLDCLIM es una base de datos desarrollada por el Museo de Vertebrados de la Universidad de California (Berkeley), con la colaboración del CIAT (Centro Internacional de Agricultura Trópica), y Rainforest CRC (Cooperative Research Centre of Tropical Rainforest Ecology and Management). La creación de las capas de datos del WORLDCLIM, fue hecha a través de la interpolación de los promedios de datos climáticos mensuales para cada estación. Las variables incluidas fueron de precipitación total mensual, temperaturas máximas y mínimas anuales, además de variables bioclimáticas derivadas de los valores mensuales de lluvias y temperatura, todas estas variables son biológicamente significativas para el modelamiento ecológico. Estas variables representan tendencias anuales, temporalidad y factores ambientales extremos o limitantes (datos de trimestres más fríos o cálidos, trimestres más lluviosos o

secos). Para los datos de estacionalidad, WORLDCLIM obtuvo información de registros de varios años, que van desde 1950 hasta 2000, incluyendo registros de áreas donde no hubo información válida reciente y registros predominantemente actuales.

Las coberturas fueron descargadas de su sitio web (<http://www.worldclim.org/>) en formato .grid, posteriormente se realizó un corte de las mismas adecuado al área de estudio mediante la herramienta *Extract by Mask* de ArcGIS (ver Figura 17).

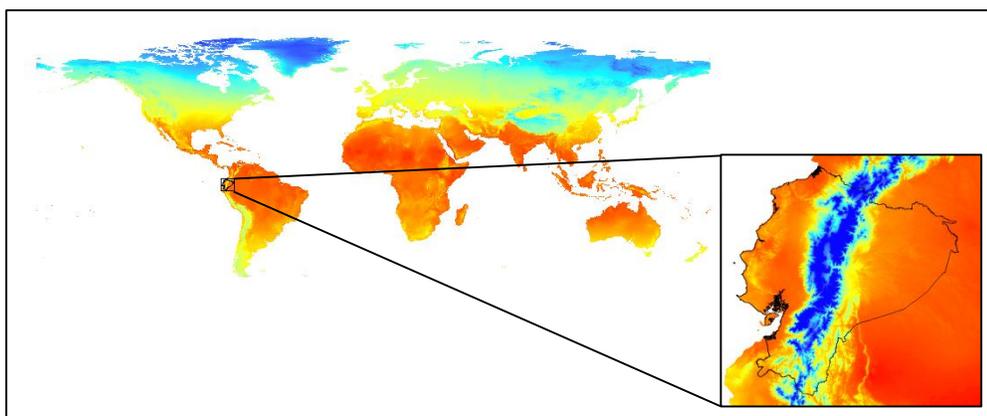


Figura 17. Corte de las variables descriptivas en base al área de estudio.

Todas la 19 variables bioclimáticas generadas por WORLDCLIM además de la variable altitud generada por el SRTM están georeferenciadas en términos de latitud y longitud, por lo que se realizó la conversión del sistema de coordenadas de cada raster al mismo sistema de coordenadas que poseen los puntos de presencia (UTM 17S) mediante el comando *Project Raster*.

Además de las variables medio ambientales se utilizó como variables predictoras coberturas referentes a centros poblados, uso del suelo y zonas de inundación; provenientes del Sistema Nacional de Información (SNI) las cuales fueron rasterizadas bajo los mismos parámetros que las variables medioambientales. De esta forma en todos los modelos se trabajará con una grilla de 802 x 812 píxeles y con una resolución de 1 km, en un sistema de coordenadas proyectadas WGS84 UTM zona 17S. En la Tabla 6 se resume las variables empleadas con su respectiva codificación.

Tabla 6.

VARIABLES DESCRIPTIVAS PARA EL MODELAMIENTO DE *Anopheles spp*

Código	Descripción
BIO 1	Temperatura media anual
BIO 2	Promedio del rango diario (Promedio mensual (max temp - min temp))
BIO 3	Isotermalidad (BIO2/BIO7) (* 100)
BIO 4	Temperatura estacional (Desviación estándar *100)
BIO 5	Max Temperatura del mes más caliente
BIO 6	Mín Temperatura del mes más frío
BIO 7	Rango de Temperatura Anual (BIO5-BIO6)
BIO 8	Promedio de temperatura del trimestre más húmedo
BIO 9	Promedio de temperatura del trimestre más seco
BIO 10	Promedio de temperatura del trimestre más caliente
BIO 11	Promedio de temperatura del trimestre más frío
BIO 12	Precipitación anual
BIO 13	Precipitación del mes más húmedo
BIO 14	Precipitación del mes más seco
BIO 15	Precipitación estacional (Coeficiente de variación)
BIO 16	Precipitación del trimestre más húmedo
BIO 17	Precipitación del trimestre más seco
BIO 18	Precipitación del trimestre más caliente
BIO 19	Precipitación del trimestre más frío
Altitud	Altura referida al nivel medio del mar
Pobla	Centros poblado de todo el Ecuador
Inunda	Zonas de anegación.
Uso	Uso de suelo

Antes de ejecutar los modelos de distribución de especies se realizó un análisis de correlación de las 19 variables medioambientales para seleccionar las variables que no están estrechamente correlacionadas y así evitar información redundante para los modelos. En la Tabla 7 se muestra el coeficiente de correlación de Pearson comparando cada una de las variables entre sí.

Tabla 7.
Matriz de correlación de variables

	bio_1	bio_2	bio_3	bio_4	bio_5	bio_6	bio_7	bio_8	bio_9	bio_10	bio_11	bio_12	bio_13	bio_14	bio_15	bio_16	bio_17	bio_18	bio_19	
bio_1	1.000																			
bio_2	-0.519	1.000																		
bio_3	-0.333	0.532	1.000																	
bio_4	0.533	-0.270	-0.800	1.000																
bio_5	0.973	-0.333	-0.323	0.595	1.000															
bio_6	0.983	-0.648	-0.331	0.464	0.919	1.000														
bio_7	-0.443	0.919	0.157	0.066	-0.230	-0.596	1.000													
bio_8	0.987	-0.568	-0.438	0.602	0.953	0.971	-0.452	1.000												
bio_9	0.985	-0.457	-0.218	0.449	0.964	0.968	-0.424	0.946	1.000											
bio_10	0.996	-0.530	-0.400	0.598	0.974	0.976	-0.424	0.994	0.971	1.000										
bio_11	0.996	-0.537	-0.289	0.466	0.958	0.988	-0.485	0.981	0.984	0.987	1.000									
bio_12	-0.005	0.152	0.451	-0.236	-0.005	0.005	-0.023	-0.084	0.095	-0.036	0.004	1.000								
bio_13	0.039	-0.043	0.100	0.008	0.022	0.054	-0.087	0.041	0.060	0.041	0.047	0.803	1.000							
bio_14	0.010	0.247	0.541	-0.282	0.029	0.006	0.043	-0.106	0.137	-0.035	0.010	0.908	0.496	1.000						
bio_15	0.156	-0.328	-0.684	0.517	0.136	0.139	-0.068	0.289	-0.001	0.213	0.139	-0.739	-0.260	-0.871	1.000					
bio_16	0.038	-0.045	0.105	0.000	0.019	0.053	-0.092	0.038	0.060	0.039	0.047	0.811	0.997	0.506	-0.267	1.000				
bio_17	-0.008	0.249	0.553	-0.299	0.010	-0.009	0.041	-0.122	0.120	-0.052	-0.005	0.914	0.504	0.997	-0.880	0.512	1.000			
bio_18	0.099	-0.195	-0.117	0.150	0.063	0.118	-0.162	0.145	0.074	0.117	0.106	0.605	0.919	0.249	-0.032	0.919	0.264	1.000		
bio_19	-0.050	0.291	0.615	-0.336	-0.031	-0.050	0.061	-0.170	0.085	-0.096	-0.046	0.887	0.464	0.971	-0.889	0.473	0.973	0.198	1.000	

Como resultado de la matriz de correlación se encontró que existe una correlación significativa entre la temperatura (bio_1) con respecto variables comprendidas entre la bio_2 a bio_11 ya que son variables derivadas; así mismo, la precipitación (bio_2) respecto las variables bio_13 a bio_19. A pesar de que las variables en su mayoría tienen una estrecha correlación con la temperatura y la precipitación, ciertas variables como bio_2, bio_3, bio_4 y bio_7 presentan una correlación moderada; dichas variables serán empleadas para el modelamiento ya que pueden aportar información relativa a la especie.

Finalmente, después de un análisis de los datos, en la Tabla 8 se resume las variables que se van a utilizar dentro del modelamiento de *Anopheles spp.*

Tabla 8.

Variables empleadas para el modelamiento de *Anopheles spp.*

Código	Descripción
BIO 1	Temperatura media anual
BIO 2	Promedio del rango diario (Promedio mensual (max temp - min temp))
BIO 3	Isotermalidad (BIO2/BIO7) (* 100)
BIO 4	Temperatura estacional (Desviación estándar *100)
BIO 7	Rango de Temperatura Anual (BIO5-BIO6)
BIO 12	Precipitación anual
Altitud	Altura referida al nivel medio del mar
Pobla	Centros poblado de todo el Ecuador
Inunda	Zonas de anegación.
Uso	Uso de suelo

3.3 Aplicación de los modelos de nicho ecológico

3.3.1 Maxent

3.3.1.1 Ajuste y calibración

Para la aplicación del modelo de máxima entropía o Maxent se empleó el software con el mismo nombre MaxEnt en su versión 3.3.3k desarrollado por Phillips, Dudik & Schapire y disponible en la página web: <https://www.cs.princeton.edu/schapire/maxent>.

Maxent tiene ciertas restricciones en sus entradas ya que la información es manejada solo en determinados formatos, por ello es necesario realizar una preparación previa de los mismos. En primer lugar es necesario convertir los datos de presencia de la especie, que en principio se encuentran en una hoja de Excel, en un archivo separado por comas (formato .csv) en el cual conste únicamente tres columnas: el nombre de la especie y su ubicación el longitud y latitud (ver Figura 18).

```

ESPECIES,X,Y
Anopheles ,622754.0000000000,9756640.0000000000
Anopheles ,601756.0000000000,9754640.0000000000
Anopheles ,623747.0000000000,9760640.0000000000
Anopheles ,691822.0000000000,9990600.0000000000
Anopheles ,678464.0000000000,9998070.0000000000
Anopheles ,632378.0000000000,10097600.0000000000
Anopheles ,648981.0000000000,10102000.0000000000
Anopheles ,649871.0000000000,10101300.0000000000
Anopheles ,646183.0000000000,9763770.0000000000
Anopheles ,672192.0000000000,9621840.0000000000
Anopheles ,648981.0000000000,9897980.0000000000
Anopheles ,651049.0000000000,10104900.0000000000
Anopheles ,629696.0000000000,10096900.0000000000
Anopheles ,649871.0000000000,9898650.0000000000
Anopheles ,601831.0000000000,9975460.0000000000

```

Figura 18. Puntos de presencia en formato .csv.

Para el manejo de las variables descriptivas, es necesario cambiar el formato de las coberturas de GRID o Tif a ASCII mediante la función *Raster to ASCII*. Una vez que se cuenta con los datos en los formatos necesarios, el siguiente paso es establecer los parámetros necesarios para la ejecución del modelo dentro del programa Maxent. Una vez en el entorno de Maxent (ver Figura 19) se carga los datos de presencia en la sección *Samples* y las variables predictoras en la sección *Environmental layers*, dentro de esta

sección es posible definir si los datos son continuos o categóricos, en este caso se utilizara uso de suelo, zonas de inundación y poblaciones como variables categóricas; además cabe mencionar que Maxent no necesita de ausencias o pseudoausencias ya que éste trabaja con datos de background (fondo) proveniente de las variables descriptivas; posteriormente, se selecciona el destino donde los datos generados por el modelo van a ser guardados. En la opción *Settings* es posible modificar ciertos ajustes para el modelo, en el presente estudio el porcentaje de datos para el entrenamiento va a ser del 75% de los puntos de presencia y el resto (25%) van ser los datos para la validación, por lo tanto en la opción *Random test percentage* se colocó 25 (ver Figura 20). Debido a que no existen estudios ecológicos profundos sobre la especie en cuestión los demás parámetros serán establecidos por default ya que son los más recomendables por los autores.

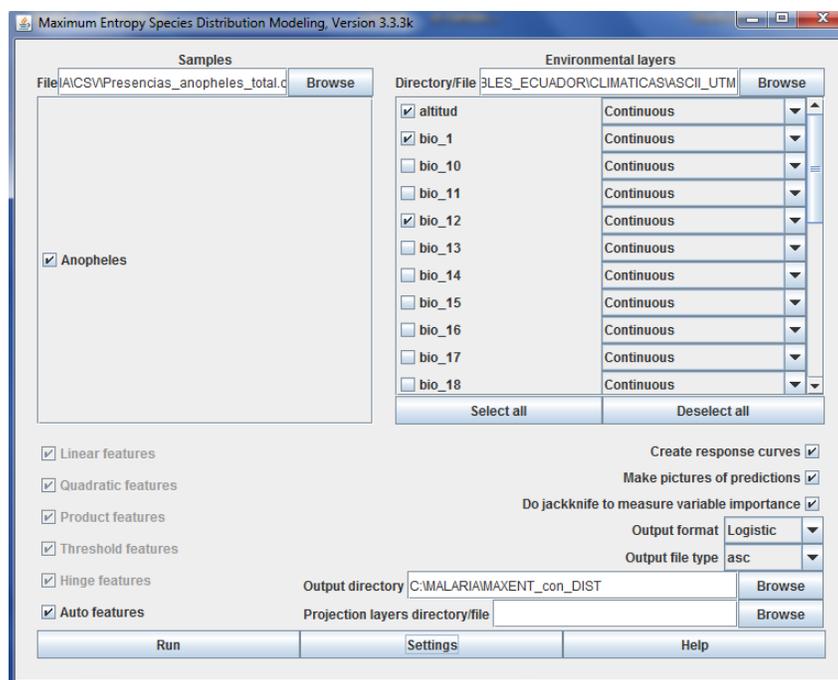


Figura 19. Entorno del software Maxent

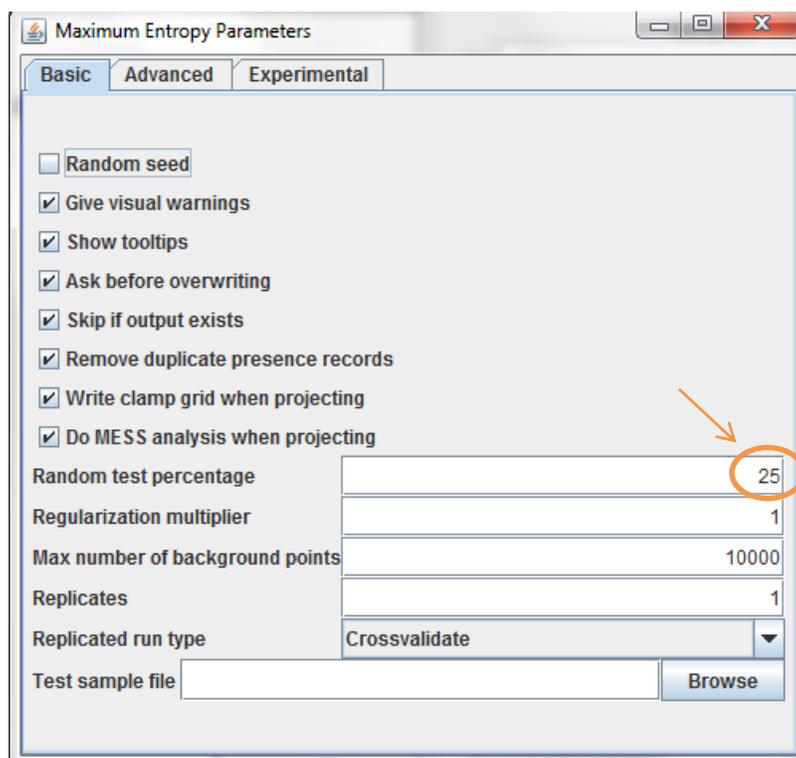


Figura 20. Opciones del software Maxent.

3.3.1.2 Resultados

Finalmente, todas las salidas del modelo se guardan en la carpeta designada, donde se encuentran varios archivos, el más importante es un archivo HTML con un resumen de todos procedimientos realizados por el software como: el análisis de omisión/comisión donde se encuentra las curvas ROC y sus respectivos valores de AUC, curvas de respuesta de cada variable, un análisis de contribución de cada variable y un mapa del resultado de la distribución de la especie generado por el modelo.

Dentro del análisis de omisión/comisión se obtuvo la curva ROC (tanto del entrenamiento del modelo como del test), representada en la Figura 21 con su respectivo valor del área bajo la curva (AUC) de 0.88 para los datos de test y 0.87 para los datos de calibración.

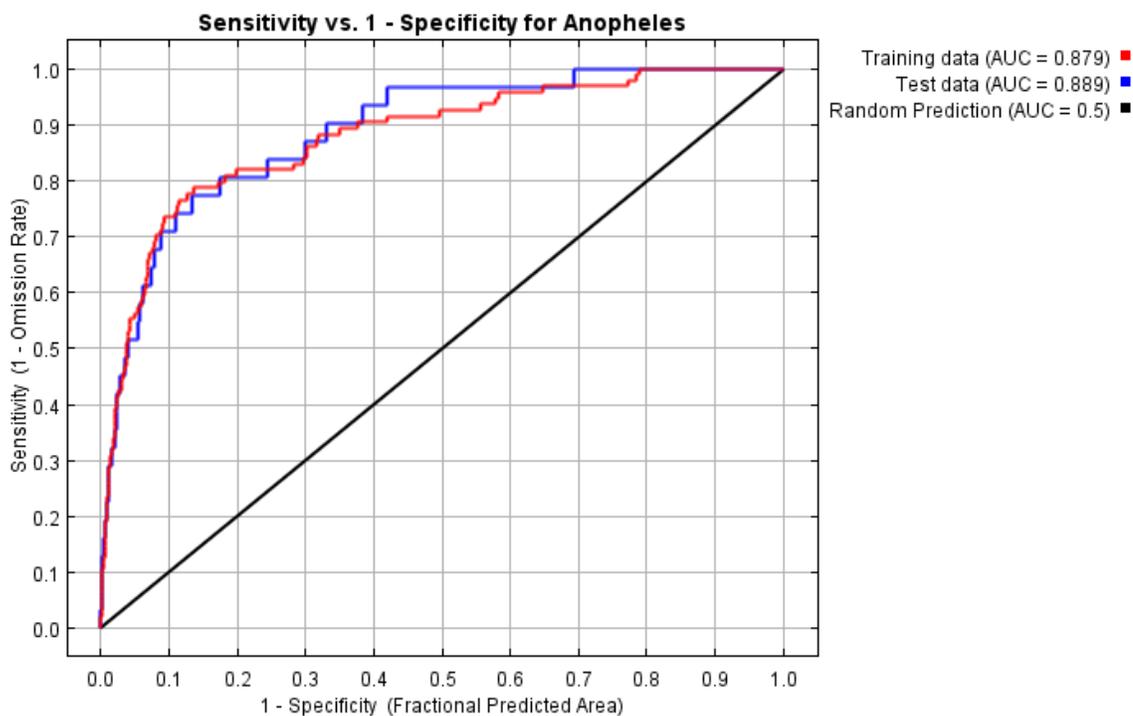


Figura 21. Curva ROC generada por el modelo Maxent.

Maxent también brinda la oportunidad de realizar un análisis individual de cada variable mediante las curvas de respuesta (ver Figura 22), las mismas que representan como cada variable influye sobre el modelo. Las curvas muestran como la respuesta logística de Maxent cambia con una variación en cada variable, manteniendo todos los valores de las demás variables en su valor promedio; además, se muestra cual es el comportamiento de la variable si se ejecutará Maxent solo con dicha variable.

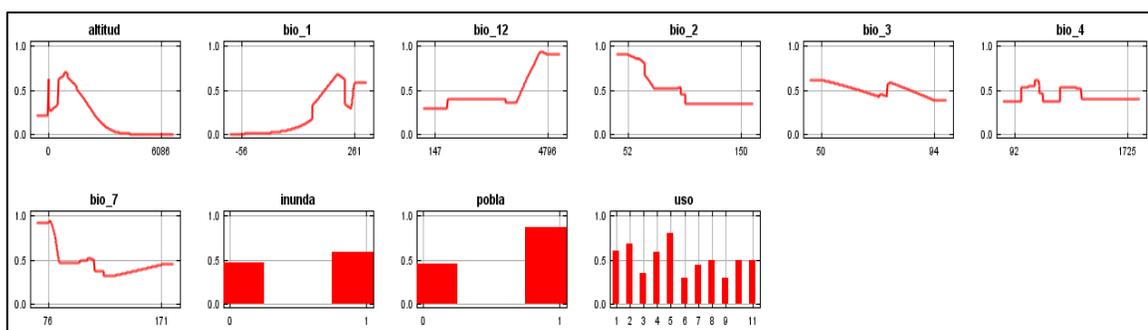


Figura 22. Curvas de respuesta de Maxent.

El jackknife o test de significancia, es un test estadístico que realiza Maxent para determinar la importancia de cada variable para el modelo en conjunto y con respecto a las demás, este es aplicado sobre los datos de entrenamiento (ver Figura 23), de test y respecto al AUC.

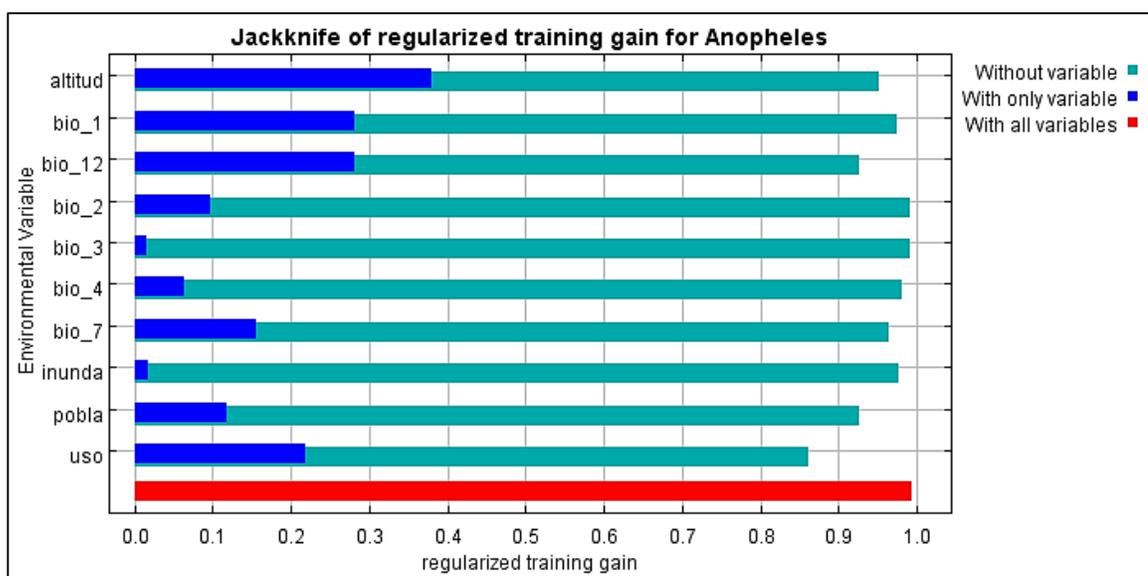


Figura 23. Test Jackknife de Maxent.

En la Figura 23 se muestra en azul la contribución de cada variable por sí sola, mientras que el en color verde muestra la contribución sin la variable en mención y por último la barra roja refleja la contribución de todas las variables en conjunto; de esta forma se observa que las variables altitud, bio_1 y bio_12 son las que más contribuyen individualmente, por otro lado las variables bio_3 e inunda son las que menos contribuyen. Comúnmente este test suele ser interpretado como un instrumento para decidir cuales variables deben ir en el modelo y cuales no en función de su importancia, sin embargo este concepto es equivoco ya que al ver el comportamiento de todo el modelo sin la variable mencionada (barra verde), se puede observar que es relativamente similar para todas las variables.

Por último, se muestra un mapa donde se observa distribución de la especie como resultado del modelamiento (ver Figura 24).

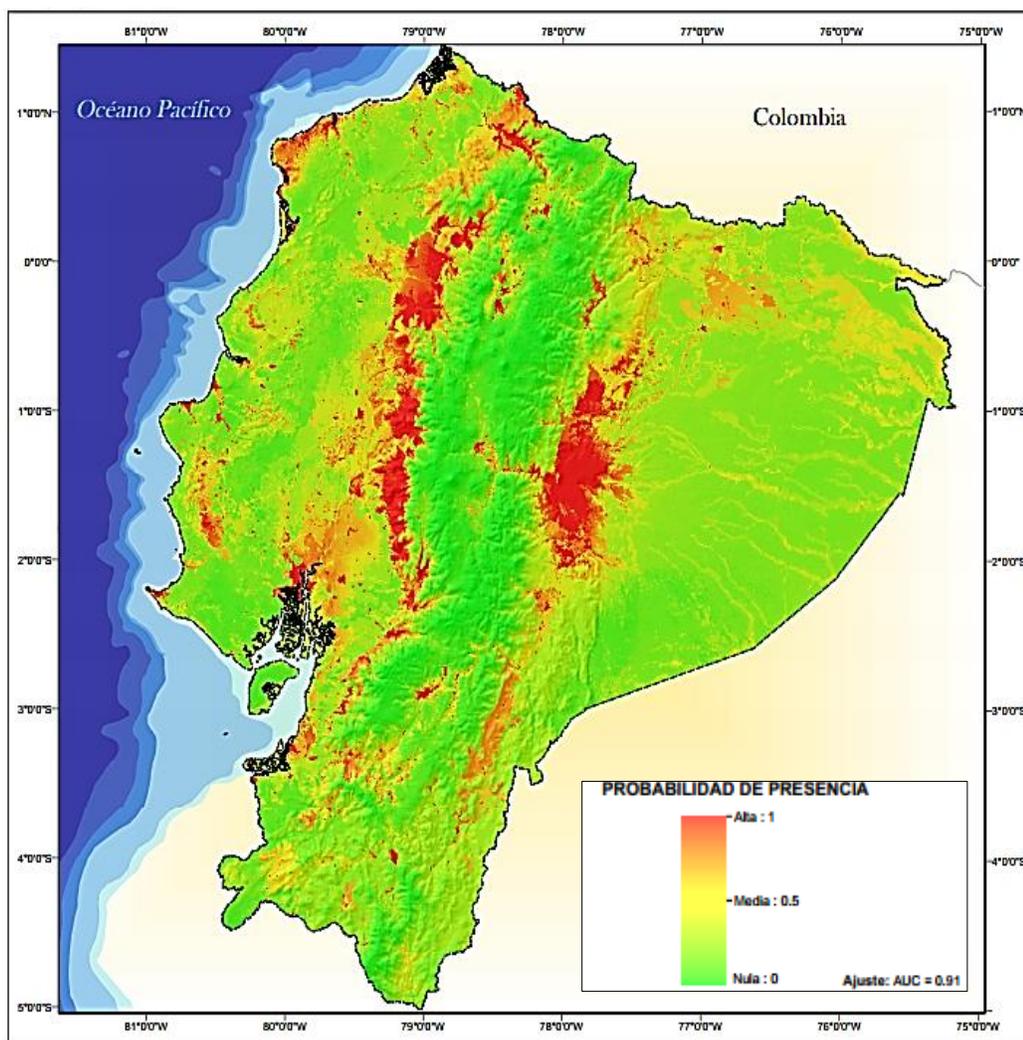


Figura 24. Resultado de Maxent para *Anopheles* spp.

3.3.2 Regresión logística.

La aplicación del modelo de regresión logística se realizó mediante la utilización del software ArcGis y el software estadístico Rstudio basado en el lenguaje de programación R. En primer lugar, se diseñó e implementó un código en R (ver Anexo 1) para obtener la regresión logística requerida en el modelo, donde los insumos necesarios son: los puntos de presencia, puntos de ausencia o pseudoausencia y las variables descriptivas.

3.3.2.1 Ajuste y calibración

Al contar con los registros de presencia y con las variables predictoras, el único insumo faltante son las ausencias, en el caso particular del presente estudio no se cuenta con datos de ausencia; por lo tanto, se generará pseudoausencias mediante puntos aleatorios. Para la generación de puntos de pseudoausencia primeramente se tiene que determinar una zona tampón o buffer alrededor de los puntos de presencia, donde se establece que la especie fuera de esta zona no se encontrará; mediante la herramienta *Buffer* de ArcGis se generó un buffer de 30 km de radio sobre los puntos de presencia (ver Figura 25), dentro de éste radio se asume que las condiciones medio-ambientales permanecen relativamente constantes (Delgado, 2011).

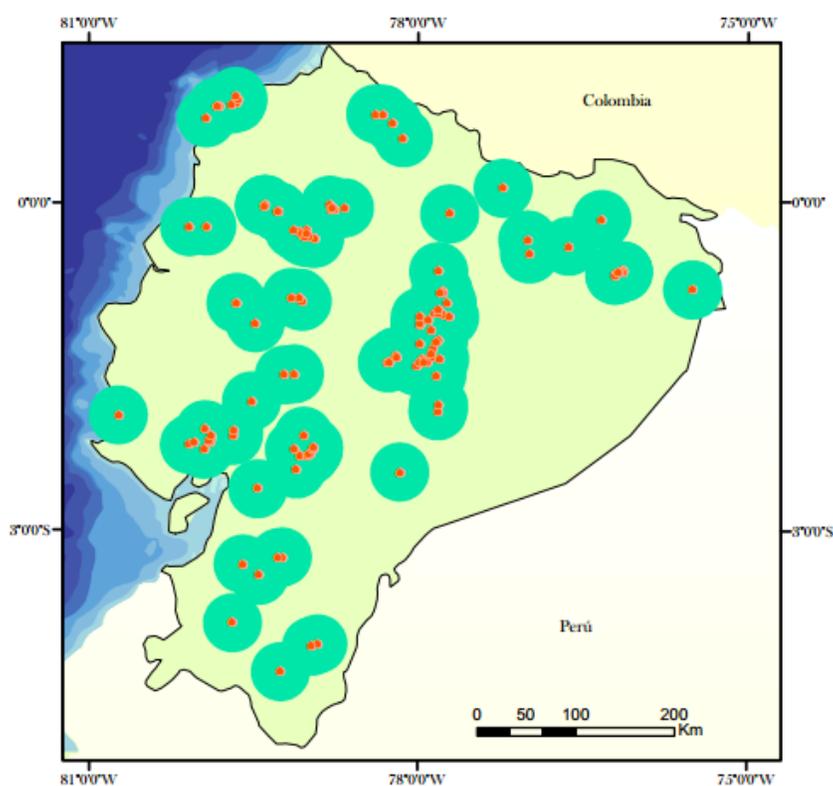


Figura 25. Buffer sobre los puntos de presencia de *Anopheles spp.*

Posteriormente con la herramienta *Erase* se elimina el área cubierta por el buffer para dejar únicamente el área disponible para generar las pseudoausencias (ver Figura 26).

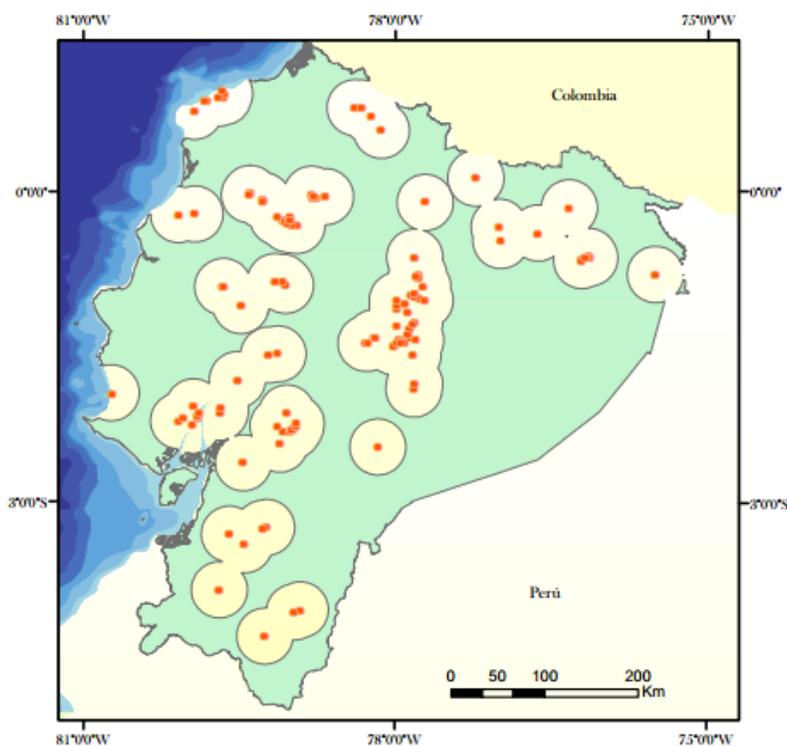


Figura 26. Zona para la creación de pseudoausencias.

Por último, se crean los puntos de pseudoausencias mediante el comando *Create Random Points*, donde se debe establecer el número de puntos aleatorios. Dentro de la bibliografía se recomienda que sea una proporción mayor de puntos de ausencia (Barbet-Massin, Jiguet, Albert, & Tulier, 2012), para la presente investigación se utilizara un conjunto de pseudoausencias de un 10% mayor al de la muestra, teniendo en cuenta que la muestra está constituida por 144 puntos de presencia, el número de pseudoausencias generado será de 159 puntos. De este modo, en la Figura 27 se muestra los dos conjuntos de datos para la regresión logística, los datos de presencia (puntos rojos) y pseudoausencia (puntos azules).

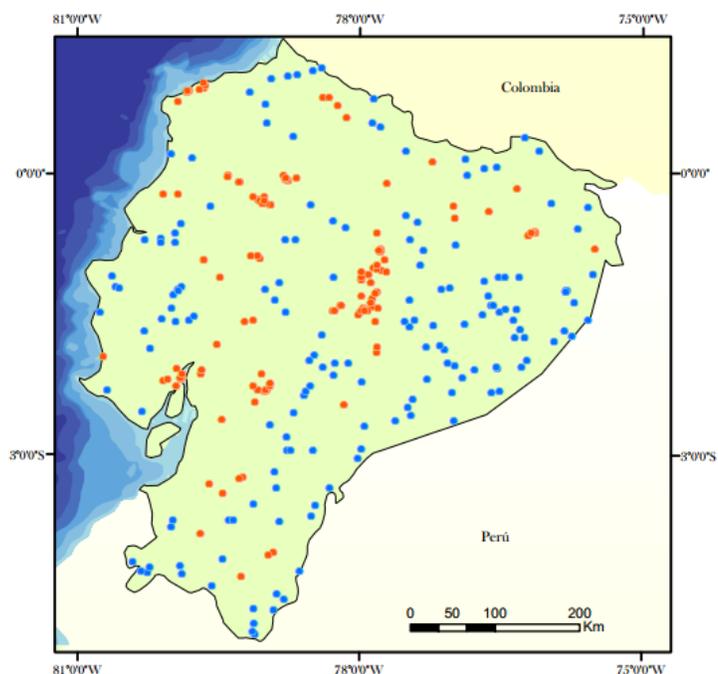


Figura 27. Datos de presencia/ausencia.

Para que la información geográfica representada en las variables sea tratada en R es necesario transformar las coberturas de formato GRID a TIF. Dentro del modelo de regresión logística es preferible tratar con variables de tipo continuo, ya que las coberturas uso de suelo, zonas de inundación y poblados son categóricas, se realizó una cobertura de distancias a las zonas de inundación y a los centros poblados (ver Figura 28) mediante el comando *Euclidean Distance* en ArcGis.

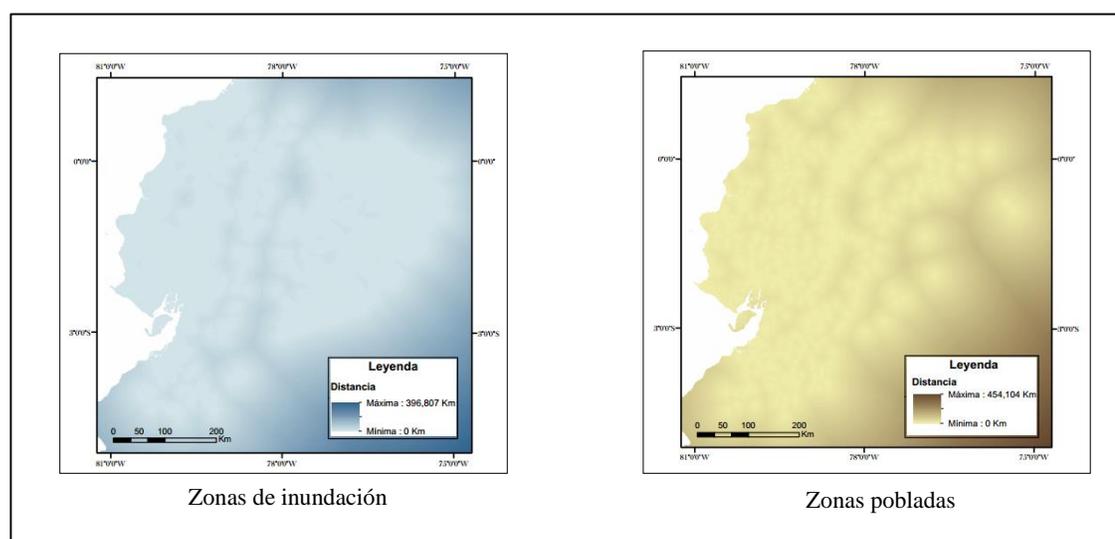


Figura 28. Mapas de Distancias

Por otro lado, para la variable uso de suelo se hizo una ponderación de sus categorías en función del número de puntos de presencia que se encontraron dentro de las mismas (ver Tabla 9); posteriormente se reclasifico el raster Uso de suelo en base a los porcentajes obtenidos (ver Figura 29).

Tabla 9.

Ponderación de la cobertura Uso de suelo

Categoría	Descripción	Frecuencias	Porcentaje
1	Agrícola	31	0.215
2	Agrícola - Conservación y protección	5	0.035
3	Agropecuario forestal	5	0.035
4	Agropecuario mixto	32	0.22
5	Antrópico	10	0.07
6	Conservación y protección	46	0.32
7	Cuerpo de agua	7	0.05
8	Forestal	0	0
9	Pecuario	3	0.02
10	Pecuario - conservación y protección	5	0.035
11	Tierras improductivas	0	0

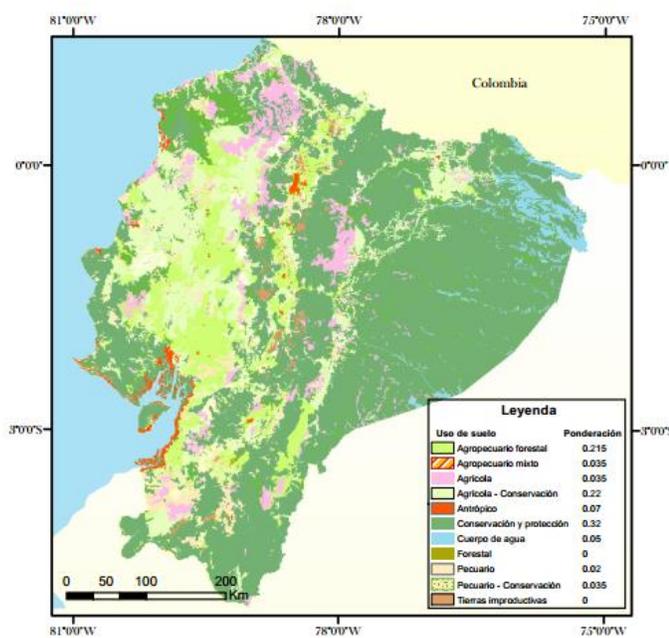


Figura 29. Mapa de uso de suelo ponderado.

Antes de trabajar con las variables dentro del modelo es necesario normalizarlas ya que se encuentran en diferentes unidades; mediante la calculadora raster de ArcGis se empleó la ecuación de normalización (27):

$$X_n = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad (27)$$

Donde X_n es la variable normalizada, X_i es la variable, X_{min} y X_{max} representan los valores mínimo y máximo respectivamente.

Por último es indispensable que los valores de cada variable estén contenidos dentro de cada punto tanto de ausencia como de presencia, para ello se utilizó la herramienta *Extract Multi Values to Points* (ver Figura 30).

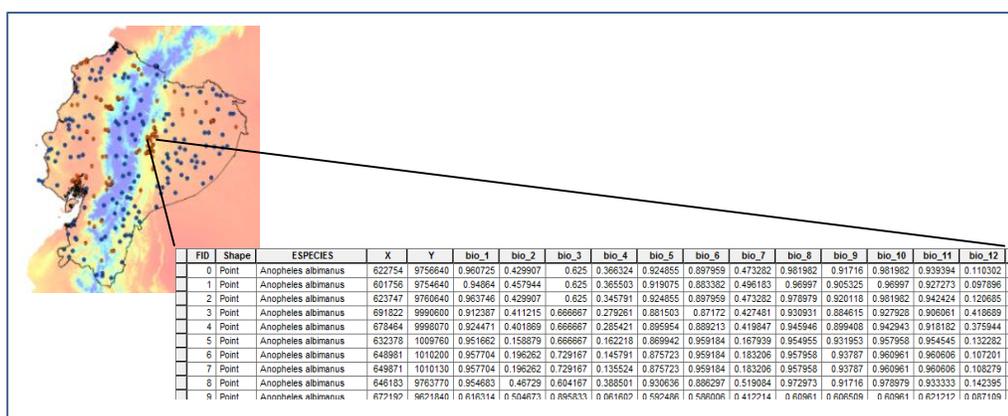


Figura 30. Extracción de los valores de cada variable.

3.3.2.2 Resultados

Una vez que se cuenta con todos los insumos para el modelamiento, en el software R se ingresan las variables mencionadas anteriormente para poder correr el modelo de regresión logística, el cual nos va a permitir calcular las constantes β_n (ver Figura 28) que son los principales parámetros para la construcción de una salida logística.

Deviance Residuals:				
Min	1Q	Median	3Q	Max
-2.5162	-0.7434	-0.1502	0.6874	2.3923
Coefficients:				
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	25.9798	22.9657	1.131	0.257953
bio_1	-10.0463	15.0909	-0.666	0.505590
bio_2	-18.3523	42.5761	-0.431	0.666434
bio_3	-16.6923	19.8374	-0.841	0.400091
bio_4	-24.6976	6.0647	-4.072	4.65e-05 ***
bio_7	21.1059	44.2657	0.477	0.633504
bio_12	5.7360	1.5183	3.778	0.000158 ***
altitud	-11.1354	14.5303	-0.766	0.443464
uso	-0.9466	1.6339	-0.579	0.562352
dist_pobla	-15.0378	4.9288	-3.051	0.002281 **
dist_inund	4.7797	9.2163	0.519	0.604028
--- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

Figura 31. Resumen del modelo logístico en R.

De este modo, los valores de las constantes son multiplicadas por cada variable y siguiendo la fórmula de la regresión logística se obtiene un mapa de probabilidades entre 0 y 1, como se muestra en la Figura 32.

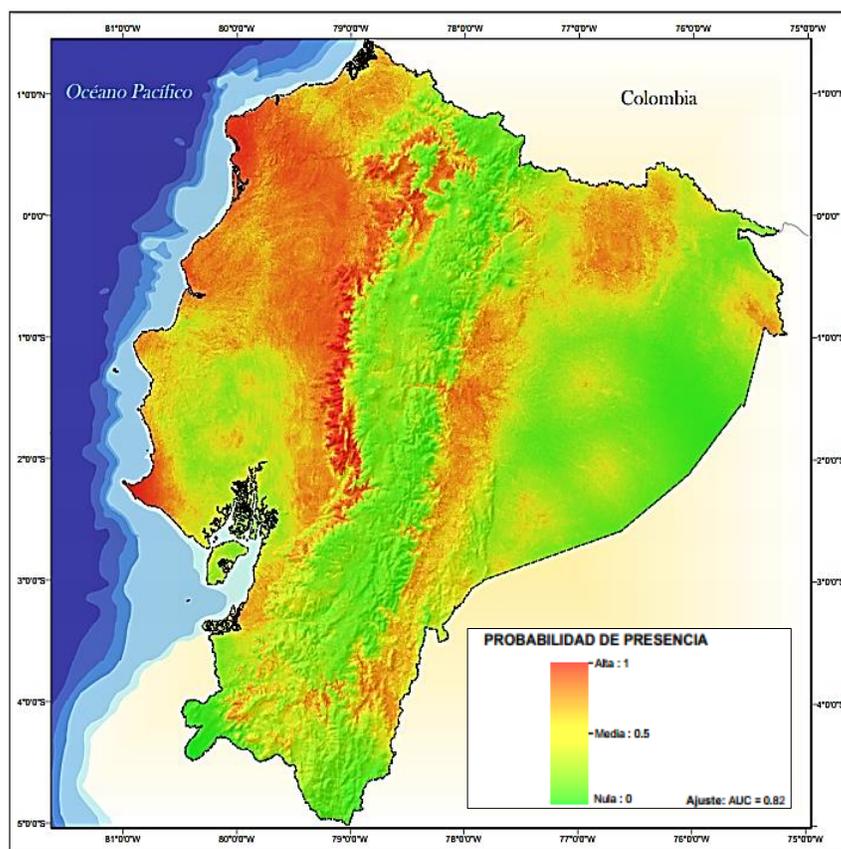


Figura 32. Resultado de la regresión logística en R para *Anopheles* spp.

Adicionalmente, se realizó la prueba estadística de la curva ROC, obteniéndose un valor de AUC de 0,86 (ver Figura 33)

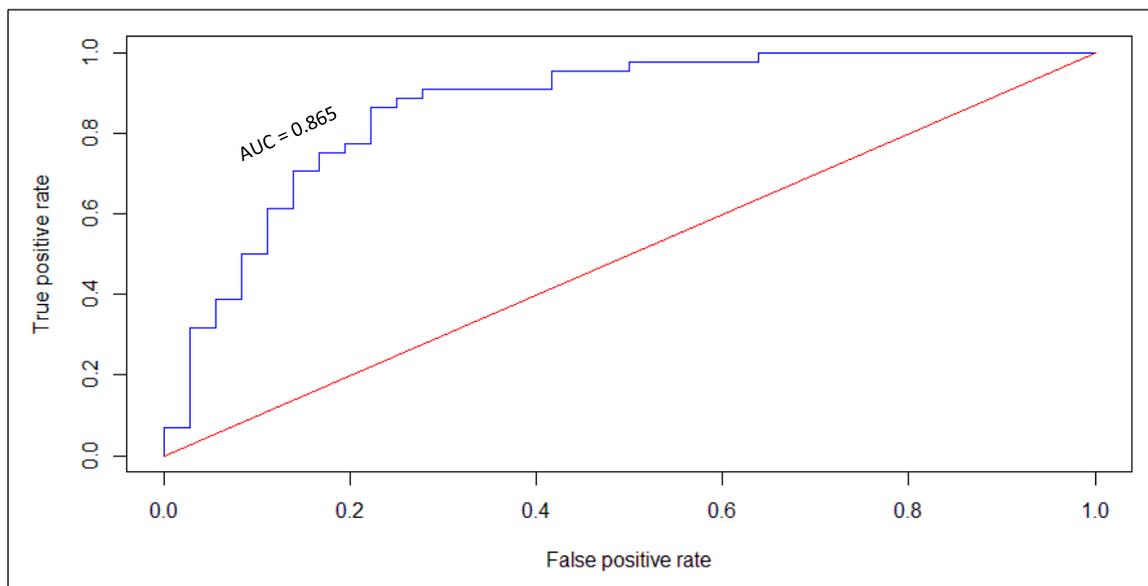


Figura 33. Curva ROC del modelo regresión logística.

3.3.3 Evaluación multicriterio con lógica difusa

El modelo de evaluación multicriterio fue llevado a cabo en su totalidad dentro del software Idrisi ya que cuenta con módulos especializados para llevar a cabo todos los procesos necesarios de una evaluación multicriterio con operadores difusos.

3.3.3.1 Ajuste y calibración

Dentro del entorno de Idrisi existe el módulo Spatial Decision Modeler el cual nos permite organizar y desarrollar todo el esquema necesario para realizar la evaluación multicriterio (ver Figura 34), dentro del mismo es necesario especificar las variables que van a ser utilizadas, para el presente modelo las variables serán: altitud, temperatura media anual (bio_1), precipitación media anual (bio_12), distancia a los centros poblados, distancia a las zonas de inundación y uso del suelo.

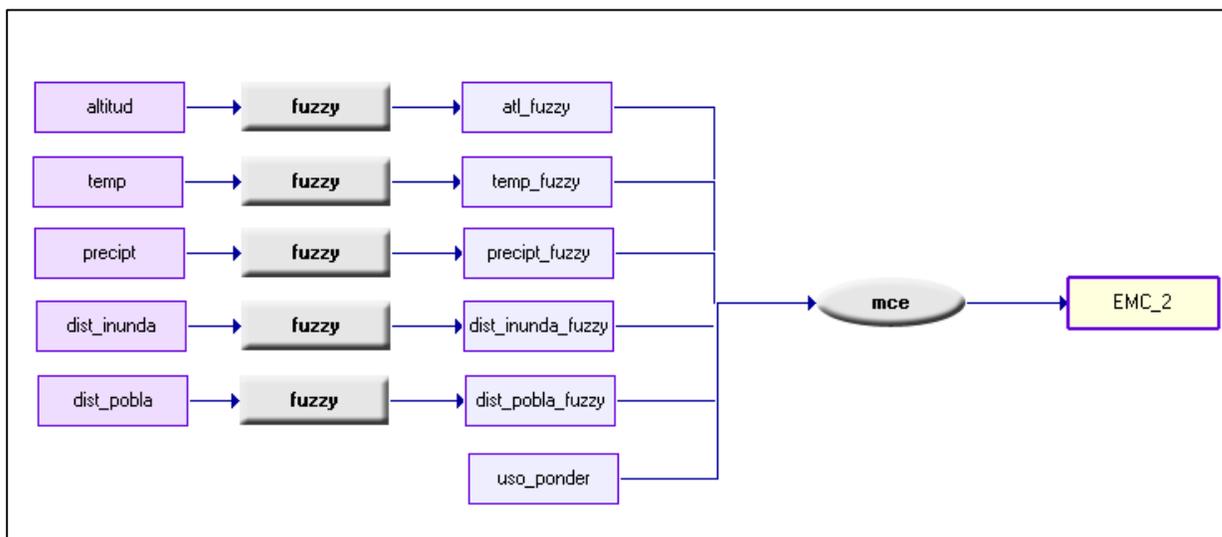
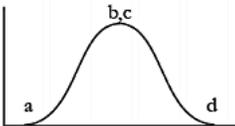
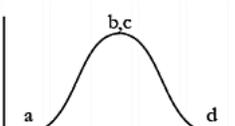
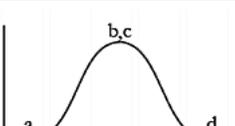
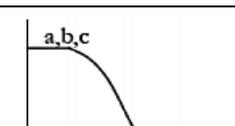
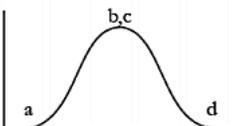


Figura 34. Esquema para la evaluación multicriterio.

El siguiente paso es transformar las variables predictoras en coberturas fuzzy mediante las funciones de pertenencia y sus respectivos puntos de inflexión, los mismos que fueron definidos en base a la revisión bibliográfica descrita en el Capítulo II y la investigación realizada por Fuller et al (2014): *Participatory risk mapping of malaria vector exposure in northern South America using environmental and population data*. En la Tabla 10 se describen las variables empleadas con sus respectivas funciones de pertenencia y puntos de inflexión.

Tabla 10.
Variables y sus respectivas funciones fuzzy.

Variable	Descripción	Función de pertenencia Fuzzy	Puntos de Inflexión
Altitud	Altura referida al nivel medio del mar (m)	 Sigmoide simétrica	$a = 0$ $b = 555$ $c = 875$ $d = 2000$
Temperatura	Temperatura media anual (°C)	 Sigmoide simétrica	$a = 20$ $b = 22$ $c = 26$ $d = 27$
Precipitación	Precipitación media anual (mm)	 Sigmoide simétrica	$a = 654$ $b = 1000$ $c = 1500$ $d = 3500$
Distancia a zonas inundadas	Distancia euclidiana desde las zonas inundadas (m)	 Sigmoide monótona decreciente	$a, b, c = 0$ $d = 3000$
Distancia a áreas pobladas	Distancia euclidiana desde las áreas pobladas (m)	 Sigmoide simétrica	$a = 0$ $b = 1000$ $c = 1000$ $d = 3500$

Para la variable uso de suelo no fue necesario normalizarla ya que previamente se realizó este proceso, por lo tanto puede ir directamente como un insumo para la evaluación multicriterio.

Una vez normalizadas las variables mediante las operaciones fuzzy, es necesario calcular el peso que tendrá cada una de las variables dentro del modelo. El método de las Jerarquías Analíticas propuesto por Saaty (1995) es de gran ayuda para la asignación de pesos para las variables, dicha metodología está basada en la comparación por pares de los elementos, que permiten una jerarquización mediante ciertas mediadas de importancia relativa de una variable respecto de otra. En la Tabla 11 se muestra el análisis de jerarquías analíticas.

Tabla 11.

Análisis de Jerarquías Analíticas.

Pairwise Comparison 9 Point Continuous Rating Scale								
1/9	1/7	1/5	1/3	1	3	5	7	9
extremely	very strongly	strongly	moderately	equally	moderately	strongly	very strongly	extremely
Less Important					More Important			
	atl_fuzzy	dist_inunda_fuz	dist_pobla_fuzz	temp_fuzzy	precipit_fuzzy	uso_ponder		
atl_fuzzy	1							
dist_inunda_fuz	2	1						
dist_pobla_fuzz	2	2/3	1					
temp_fuzzy	3	3/2	2	1				
precipit_fuzzy	3	1	2	3/2	1			
uso_ponder	3/2	1/2	1/2	1/3	1/3	1		

Como resultado se obtiene la asignación de pesos para cada variable (ver Tabla 12); además del valor del radio de consistencia, el cual establece una medida operativa de la consistencia en la asignación de los juicios de valor. Donde, para valores mayores o iguales a 0,10 los juicios de valor deben ser revisados ya que no son suficientemente consistentes para establecer los pesos. Si por el contrario el radio de consistencia es inferior a 0,10 son satisfactorios los juicios de valor asignados. En el presente estudio se obtuvo un radio de consistencia de 0,01 el cual es aceptable para continuar con la asignación de pesos.

Tabla 12.

Asignación de pesos para las variables.

Factors	
Factor filename	Factor weight
atl_fuzzy	0.0771
dist_inunda_fuzzy	0.1898
dist_pobla_fuzzy	0.1397
temp_fuzzy	0.2431
precipit_fuzzy	0.2618
uso_ponder	0.0883

3.3.3.2 Resultados

Finalmente, como resultado se obtiene el mapa de distribución potencial de *Anopheles spp* (ver Figura 35).

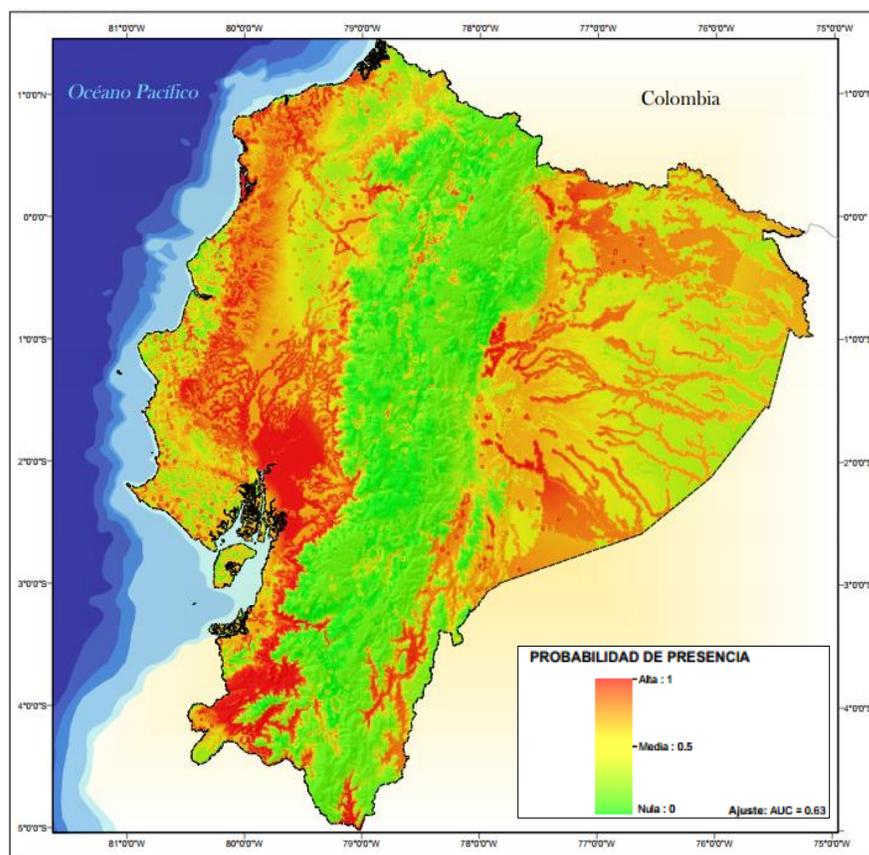


Figura 35. Resultado de la EMC para *Anopheles spp*.

CAPÍTULO IV

4 Análisis y discusión

4.1 Análisis estadístico de los modelos.

Para efectuar la selección del mejor modelo de nicho ecológico para *Anopheles spp.* se realizaron pruebas y comparaciones estadística aplicables a los tres modelos.

4.1.1 Análisis por desviación estándar

La desviación estándar es una medida estadística de dispersión que indica cuánto tienden a alejarse los valores concretos respecto al valor promedio en una distribución de datos. En la Tabla 13 se resumen los valores de desviación estándar extraídos de cada cobertura raster resultante de cada uno de los modelos.

Tabla 13.

Análisis por desviación estándar

Modelo	Min	Max	Media	Desv.Estandar
Evaluación Multicriterio	0	1	0.3235	±0.2801
Regresión Logística	0.0031	0.9992	0.4351	±0.2583
Maxent	0	0.9906	0.1853	±0.1592

Como resultado del análisis por desviación estándar, se obtuvo que el modelo de máxima entropía (Maxent) arrojó el menor valor lo que significa que el rango de valores del modelo no varía con más de ± 0.15 del valor promedio. Seguido del modelo de regresión logística 0.25 y posteriormente evaluación multicriterio con 0.28.

4.1.2 Análisis por rangos de probabilidad

Para el presente análisis se realizó una reclasificación de las coberturas de probabilidad de cada uno de los modelos en intervalos del 20 % (0, 0.20, 0.40, 0.60,

0.80, 1), posteriormente se empleó los 144 puntos de presencia para identificar con qué frecuencia son identificados los puntos en cada una de las clases antes mencionadas.

- Evaluación Multicriterio (EMC)

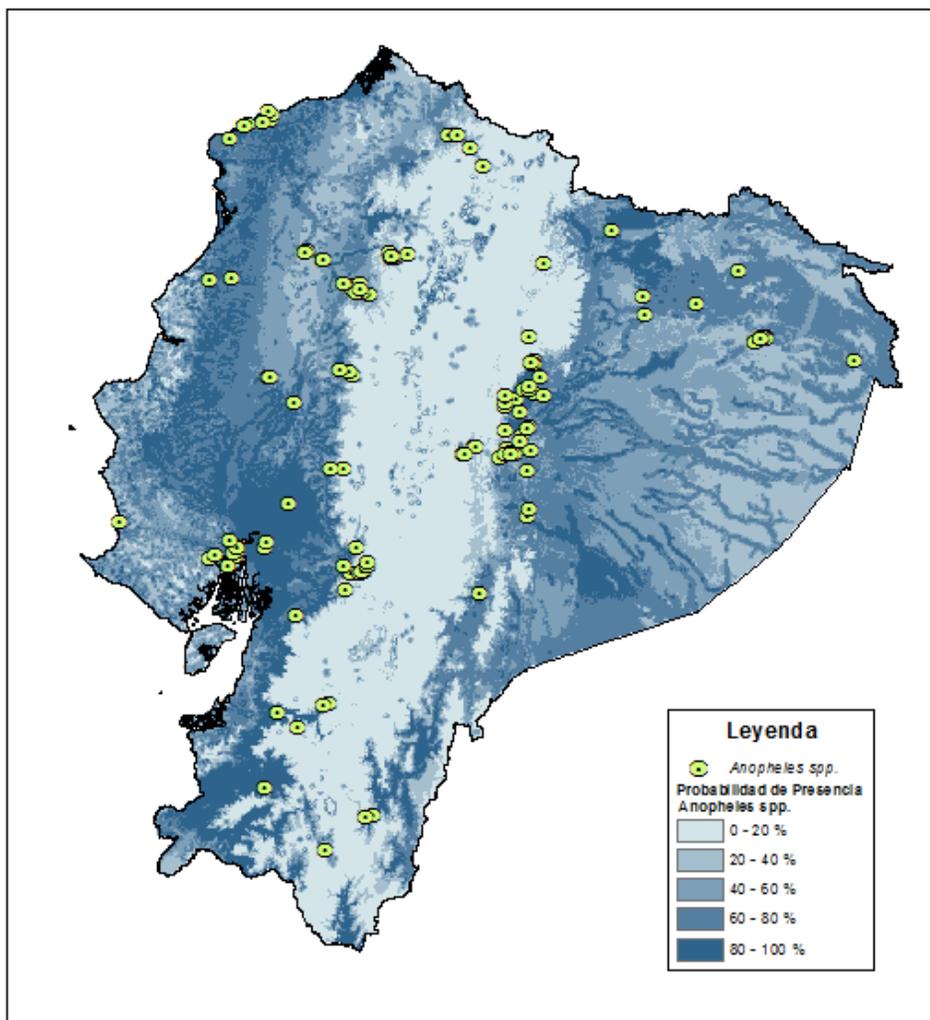
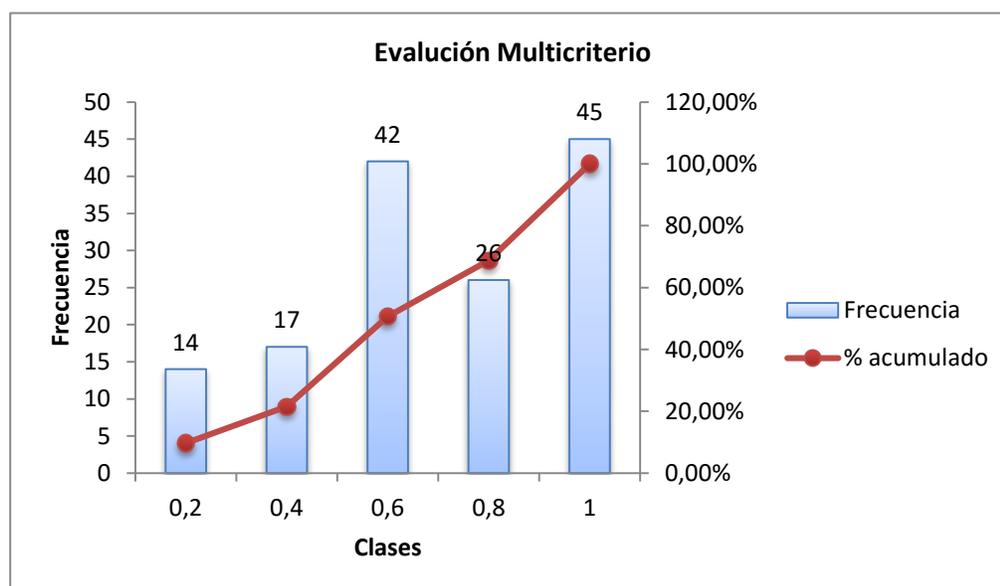


Figura 36. Evaluación Multicriterio - Rangos de probabilidad

Tabla 14.

Análisis por rangos de probabilidad - EMC

Clases		Frecuencia	% Individual	% Acumulado
Lím. Inferior	Lím. Superior			
0	0.2	14	9.72%	9.72%
0.2	0.4	17	11.81%	21.53%
0.4	0.6	42	29.17%	50.69%
0.6	0.8	26	18.06%	68.75%
0.8	1	45	31.25%	100.00%
Total		144	100%	
Media		0.6145		
Desv. Estandar		±0.2756		

**Figura 37.** Histograma de frecuencias - EMC

- Regresión logística

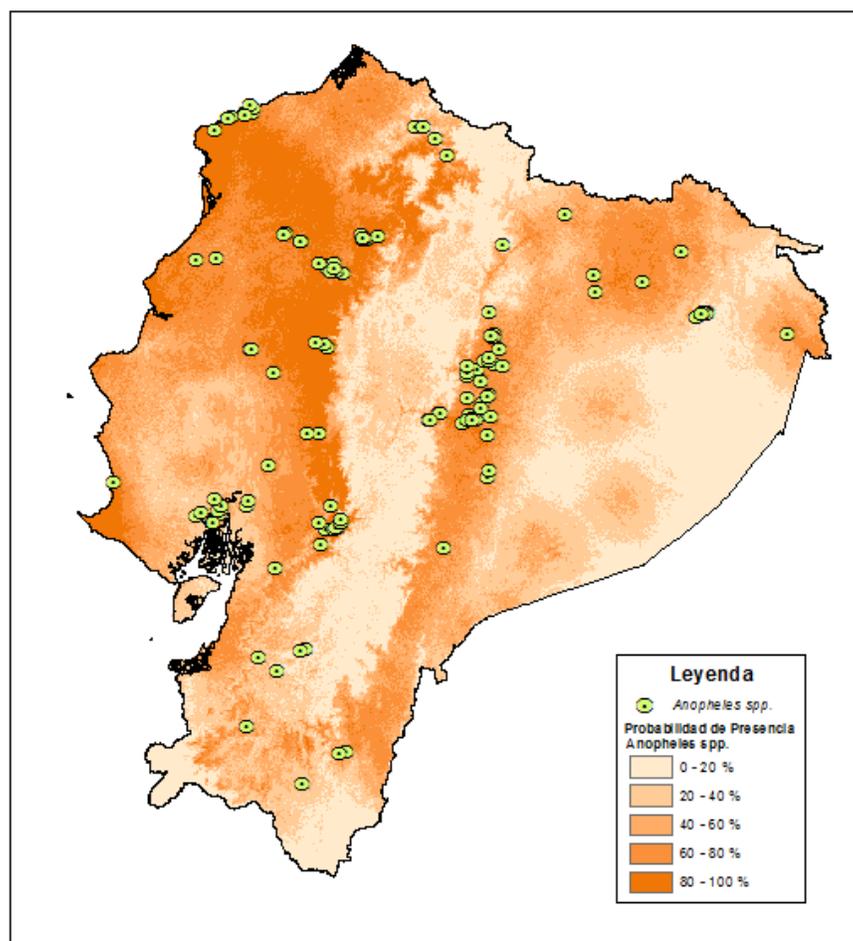


Figura 38. Regresión Logística - Rangos de probabilidad

Tabla 15.

Análisis por rangos de probabilidad – Regresión Logística

Clases		Frecuencia	% Individual	% Acumulado
Lím. Inferior	Lím. Superior			
0	0.2	8	5.56%	5.56%
0.2	0.4	24	16.67%	22.22%
0.4	0.6	17	11.81%	34.03%
0.6	0.8	52	36.11%	70.14%
0.8	1	43	29.86%	100.00%
Total		144	100%	
Media		0.6318		
Desv. Estandar		±0.2446		

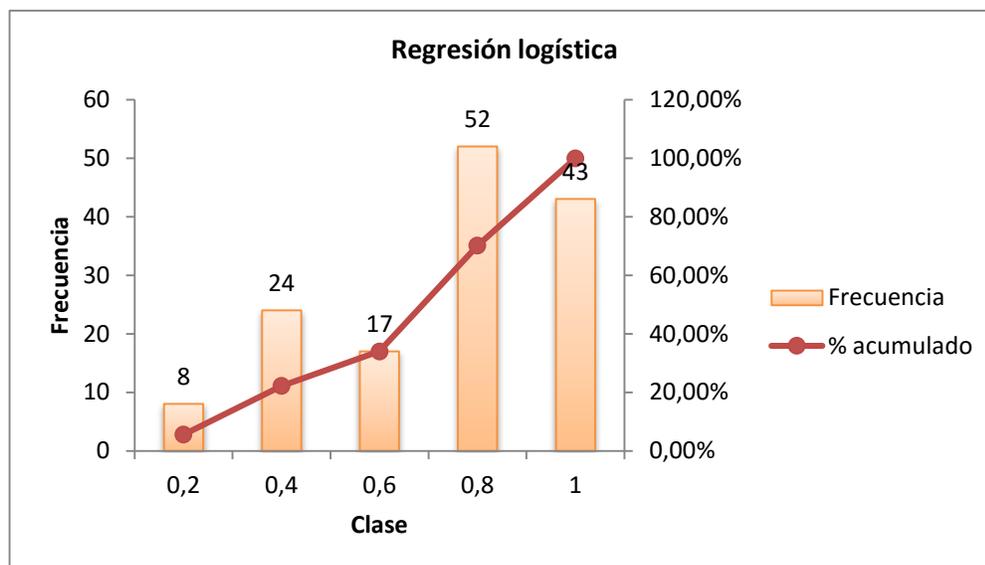


Figura 39. Histograma de frecuencias – Regresión Logística

- Maxent

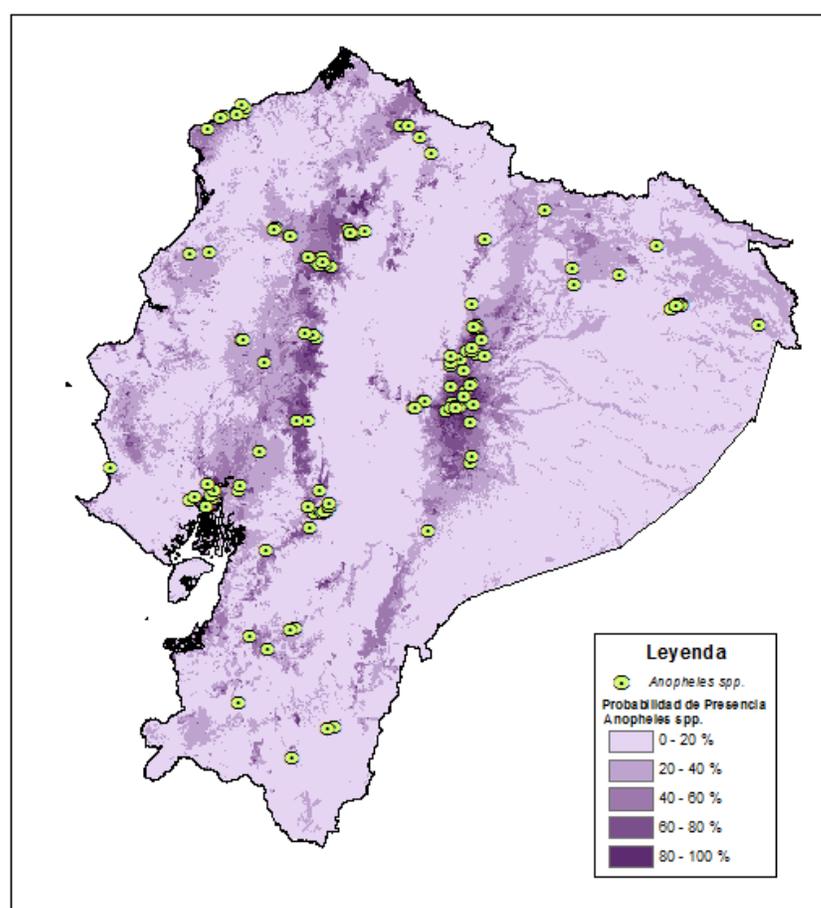
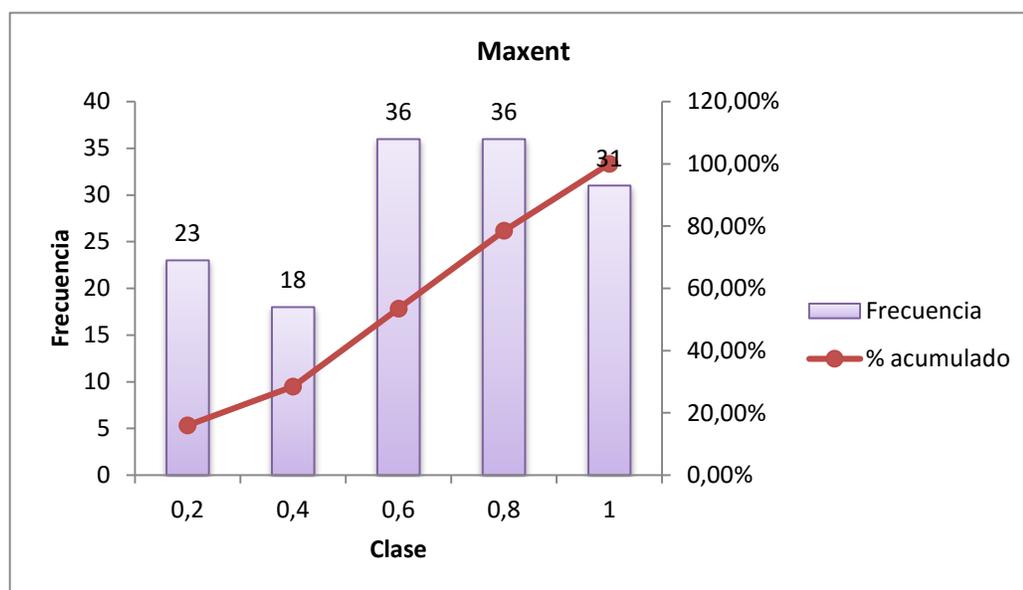


Figura 40. Maxent - Rangos de probabilidad

Tabla 16.

Análisis por rangos de probabilidad – Maxent

Clases		Frecuencia	% Individual	% Acumulado
Lím. Inferior	Lím. Superior			
0	0.2	23	15.97%	15.97%
0.2	0.4	18	12.50%	28.47%
0.4	0.6	36	25.00%	53.47%
0.6	0.8	36	25.00%	78.47%
0.8	1	31	21.53%	100.00%
Total		144	100%	
Media		0.5494		
Desv. Estandar		±0.2652		

**Figura 41.** Histograma de frecuencias – Maxent

De los parámetros estadísticos obtenidos, se observa que dentro del rango de probabilidad de 0.6 a 1 se encuentra el 49.31% de los puntos de presencia para el modelo de evaluación multicriterio, dentro del mismo rango se encuentra el 65.97% de los datos para el modelo de regresión logística y para el modelo Maxent se presentan 46.53% de

los puntos de presencia. Por otro lado, dentro de los puntos de presencia se analizó la desviación estándar de dichos puntos respecto a las coberturas raster de probabilidad; donde, el modelo de EMC presento una desviación típica de 0.27, el modelo de regresión logística 0.24 y Maxent 0.26 respecto al valor medio.

4.1.3 Análisis por la Curva ROC.

Para los tres modelos empleados se realizó un análisis de la curva ROC con su respectivo valor del área bajo la curva (AUC).

- Evaluación Multicriterio (EMC)

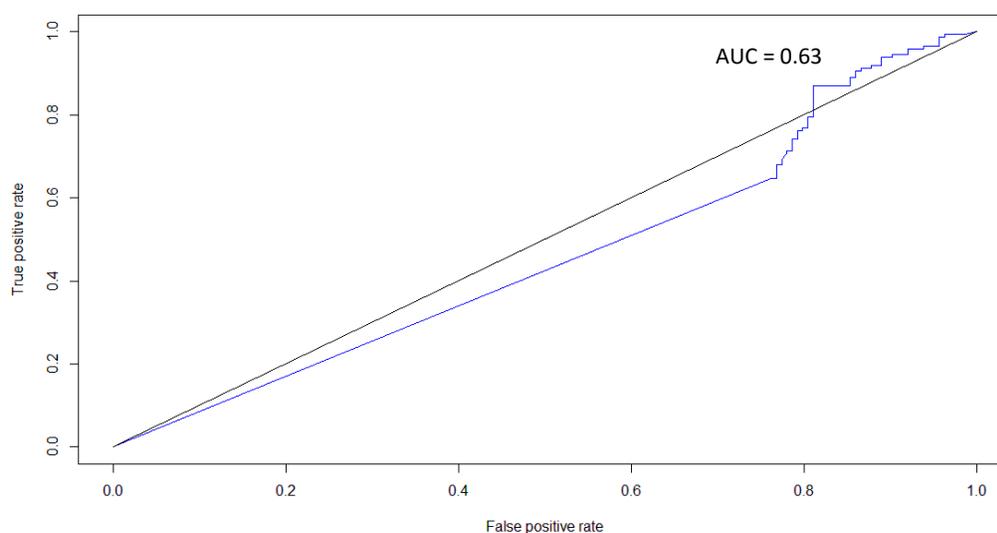


Figura 42. Análisis de Curva ROC – EMC.

- Regresión Logística

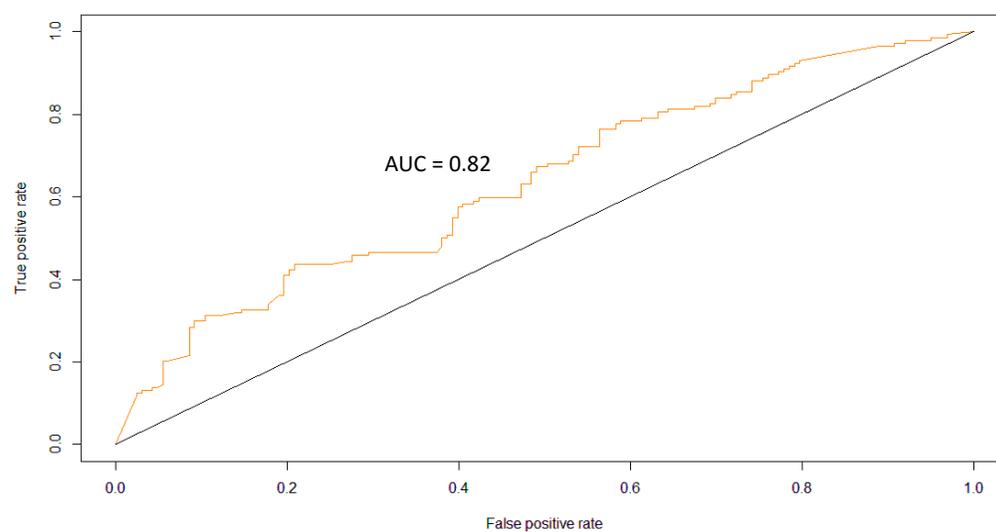


Figura 43. Análisis de Curva ROC – Regresión Logística

- Maxent

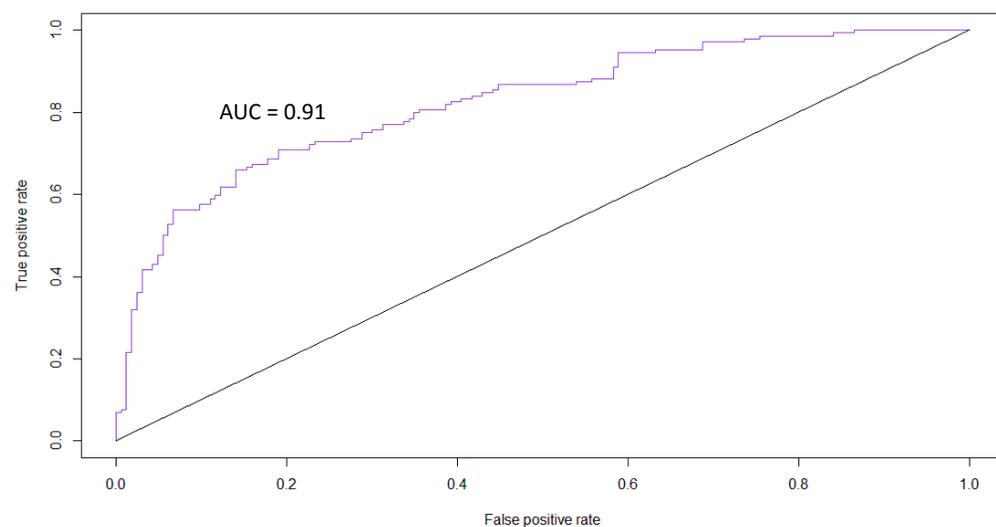


Figura 44. Análisis de Curva ROC – Maxent

Por último, si se superponen los gráficos (ver Figura 45) es evidente que el modelo Maxent posee un mayor valor de AUC con 0.91 lo que significa que está haciendo buenas predicciones ya que éste valor está muy cercano a 1, de igual forma el valor de

AUC para regresión logística es aceptable (AUC=0.82); y finalmente, el valor de AUC = 0.63 de evaluación multicriterio establece que el modelo no está haciendo un buen trabajo ya que sus predicciones están muy cercanas a un modelo al azar (AUC \leq 0.5).

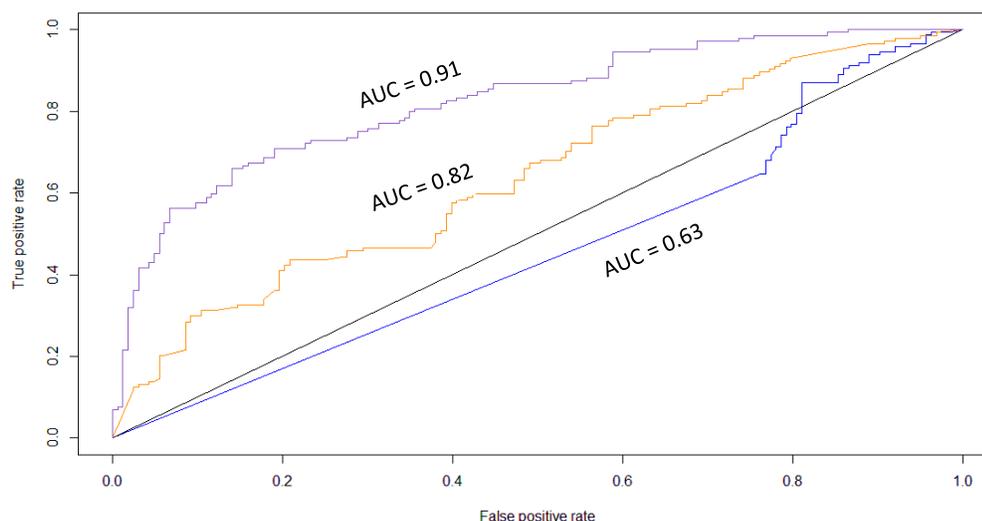


Figura 45. Análisis de Curva ROC general

4.2 Selección del mejor modelo y delimitación de las zonas idóneas para la presencia.

Después de haber realizado la evaluación estadística de los modelos empleados en la presente investigación, se ha determinado que el modelo que mejor describe la distribución de *Anopheles spp* es el modelo Maxent, presentando un buen desempeño de predicción con un valor AUC=0.91 en el análisis de curva ROC; y además, el valor de desviación estándar fue menor que los demás modelos con ± 0.15 , representando una menor variación con respecto al valor medio del modelo. Posteriormente, se realizó una reclasificación de los valores de probabilidad resultantes del modelo Maxent (ver Figura 46) bajo el umbral de presencia de la especie sobre 60 %, lo cual permitirá discriminar dos categorías, las zonas libres de la presencia de la especie (de 0 a 60 %) y las zonas idóneas para la presencia de la especie *Anopheles spp* (de 60 a 100 %), por lo tanto las de mayor riesgo biológico.

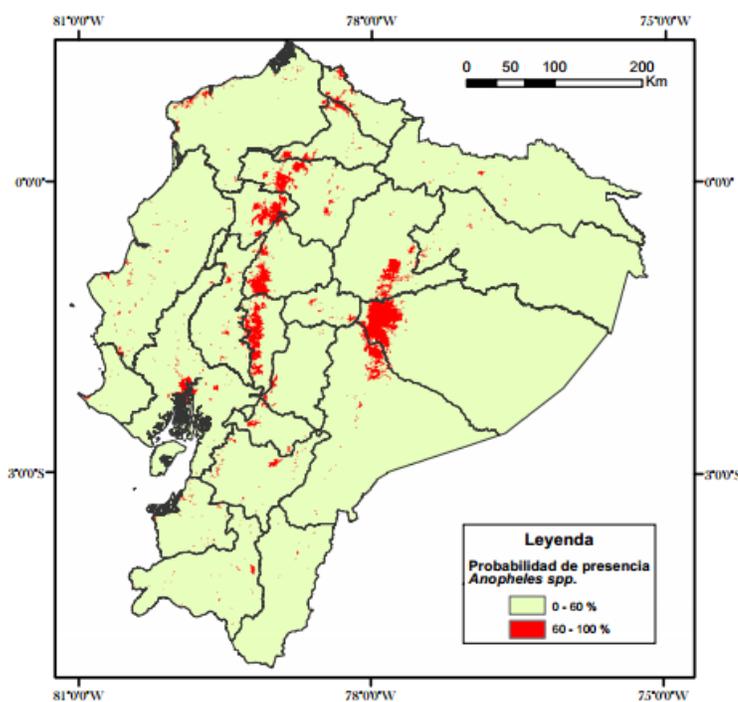


Figura 46. Zonas idóneas para la presencia de la especie *Anopheles spp.*

La áreas idóneas para la presencia de *Anopheles spp.* ocupan un superficie de 821.15 km², lo que significa el 3.24 % del territorio continental, dentro de esta porción del territorio se encuentran localidades importantes para el desarrollo nacional como las ciudades de Guayaquil, Manta, Santo Domingo de los Tsáchilas, Esmeraldas, Puyo y Tena; además de ciertas zonas en la región costera, en las zonas bajas de la cordillera y una zona amplia en el Oriente ecuatoriano.

4.3 Caracterización paisajística.

Dentro de todas las zonas delimitadas, donde la probabilidad de presencia de la especie es alta (>60%), existen diversas localidades donde es posible realizar un análisis paisajístico que permita conocer características del territorio que ayuden a describir mejor los factores asociados a la distribución de *Anopheles spp.* Sin embargo, se ha seleccionado la provincia de Esmeraldas, específicamente el sector noroccidente de la provincia, donde se encuentra su capital Esmeraldas y centros poblados cercanos como Atacames, Tonsupa y Muisne, ya que la provincia es una de las más vulnerables ante

amenazas de origen natural como deslizamientos, tsunamis, inundaciones y sismos; así mismo, existe una marcada problemática ambiental y social debido a la deforestación, contaminación y extrema pobreza. Para llevar a cabo este análisis, se creó una base de datos geográfica con información temática de la provincia de Esmeraldas que pueda ayudar a describir mejor la presencia de *Anopheles spp* en el área, por ello se han obviado coberturas relativas a la geología, geomorfología y características físicas del suelo y otras; ya que no juegan un papel realmente importante para el fenómeno.

4.3.1 Uso de suelo de la provincia de Esmeraldas

En la Figura 47 se muestra el mapa de uso de suelo de la provincia de Esmeraldas, donde, el 29.16 % de la superficie es destinada tanto a la producción pecuaria como a la conservación, además un 24,62 % destinado a la conservación y protección; siendo estas dos categorías las más representativas en función de la superficie del territorio.

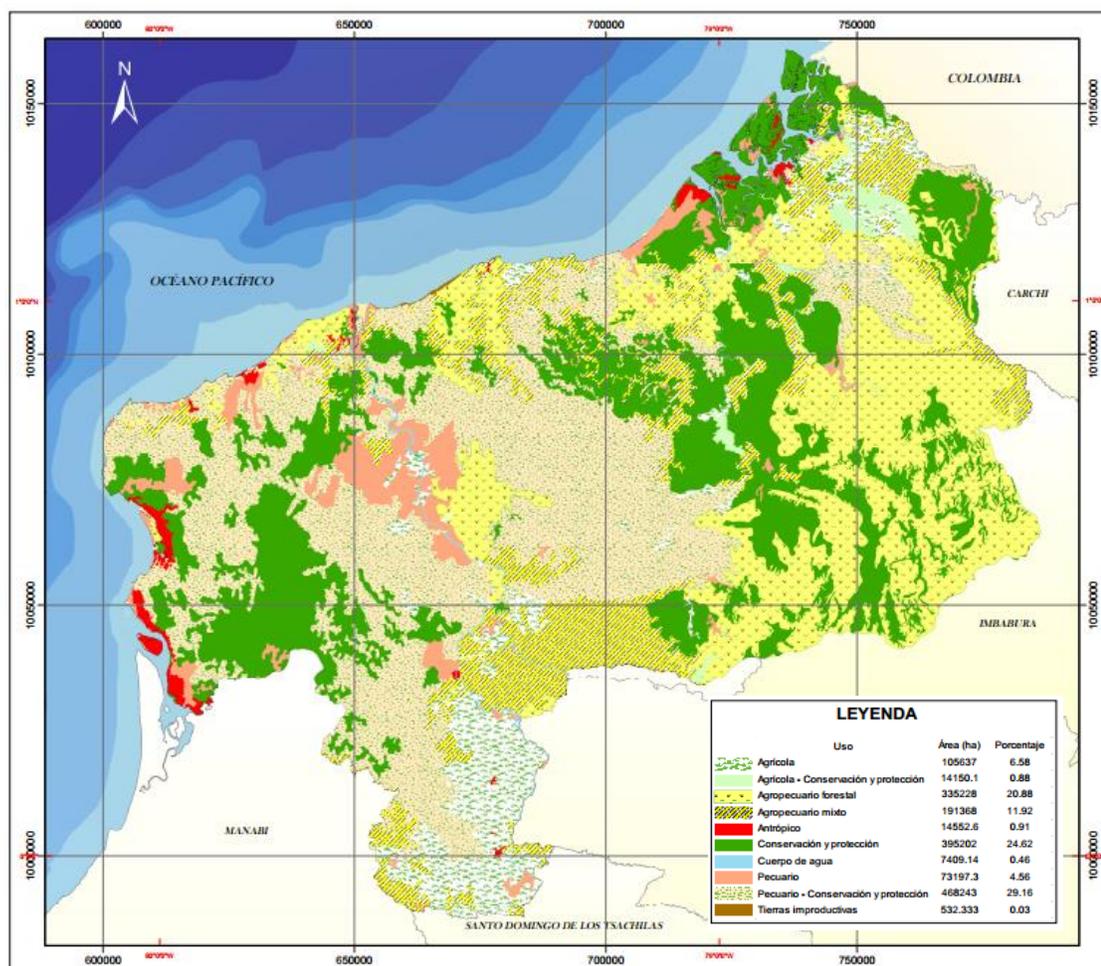


Figura 47. Uso de suelo de la provincia de Esmeraldas

4.3.2 Zonas de Vida de la provincia de Esmeraldas.

Las zonas de vida o formaciones ecológicas, propuesto por Leslie Holdridge (1979) y modificado por Cañadas (1983) para el Ecuador, es un sistema de zonificación cuyo objetivo es el de determinar áreas donde las condiciones ambientales sean similares, con el fin de agrupar y analizar las diferentes poblaciones y comunidades bióticas con características semejantes, para así aprovechar mejor los recursos naturales sin deteriorarlos y conservar el equilibrio ecológico. Dentro de la provincia de Esmeraldas se encuentran varias zonas de vida (ver Figura 48), donde destacan el Bosque húmedo tropical (43.60% de la superficie) y el Bosque seco tropical (19.67% de la superficie), en

los cuales se desarrollan diferentes tipos de manglar como el mangle rojo (*Rizophora mangle*), cuya altura a veces supera los 10 y 15 metros; en las zonas de sedimentación, además del mangle rojo, se encuentran árboles de mangle iguanero (*Avicennia nitida*) formando bosques densos que fluctúan entre 5 y 10 metros; además de una zona de transición entre el bosque seco y el bosque húmedo cuya característica es la formación de un pantano estacionario dulceacuícola propicio para la actividad camaronera.

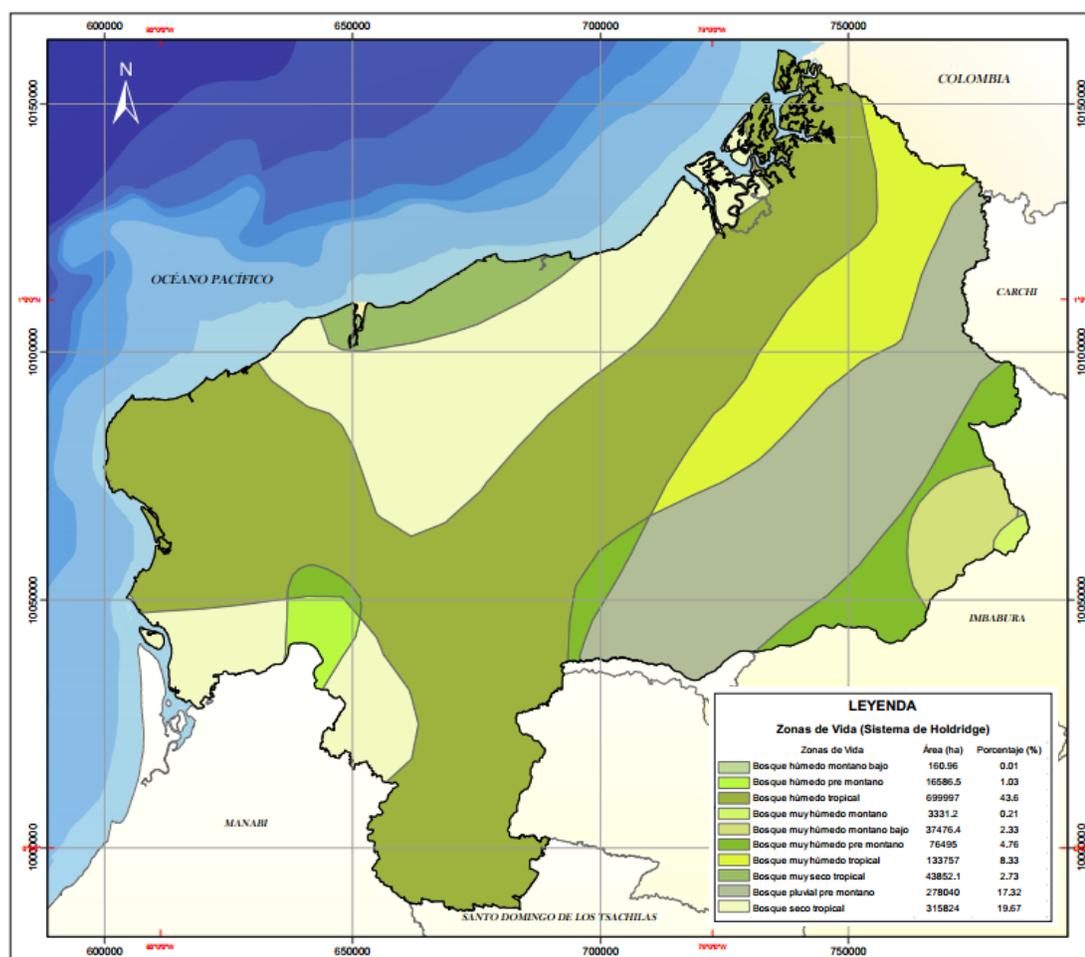


Figura 48. Zonas de vida de la provincia de Esmeraldas

4.3.3 Factores Climáticos de la provincia de Esmeraldas.

En la Figura 49, está representado el régimen térmico anual de la provincia donde predomina una temperatura entre los 24 a 26 °C, llegando a ser mayor en ciertas regiones (26-28°C).

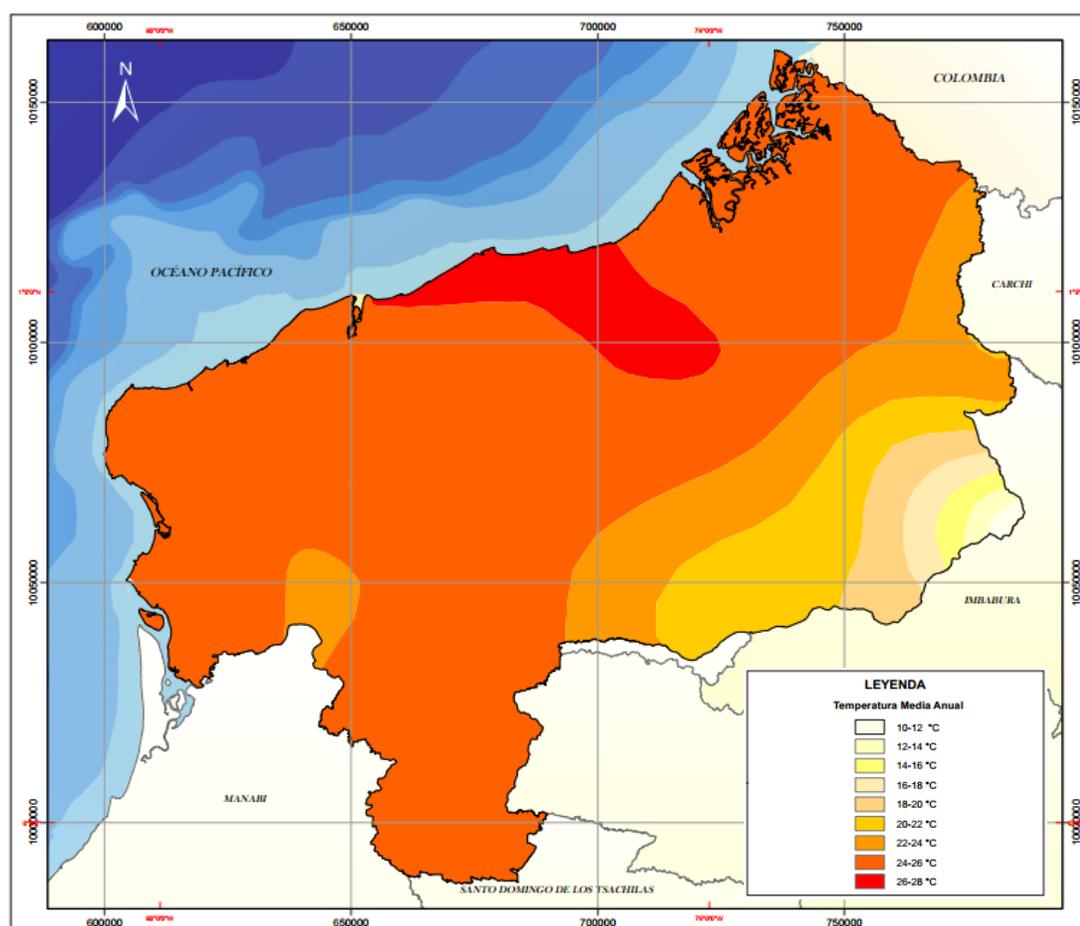


Figura 49. Temperatura Media Anual de la provincia de Esmeraldas.

Es posible caracterizar a la provincia de Esmeraldas como húmeda, con varios rangos de pluviosidad anual por sobre los 2000 mm (ver Figura 50); llegando a superhúmeda con pluviosidad anual de hasta 7000 mm o más.

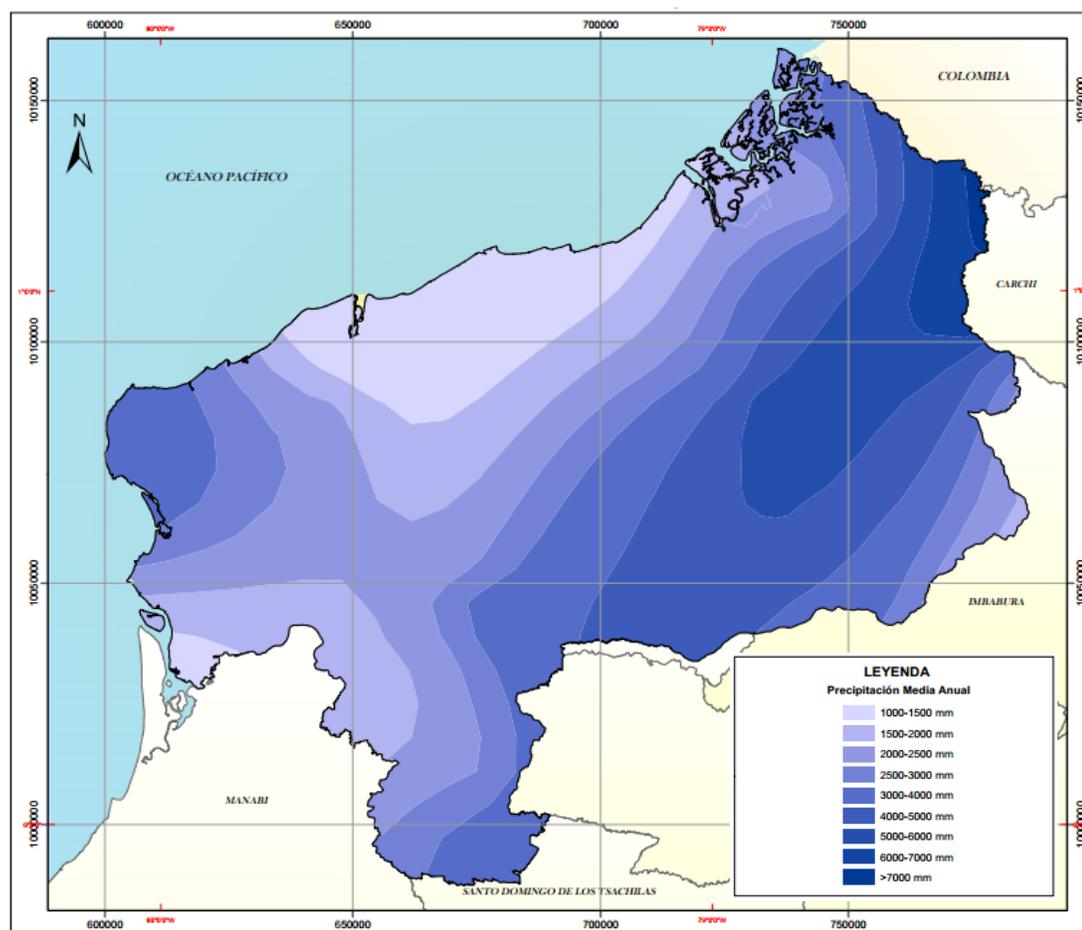


Figura 50. Precipitación Media Anual de la provincia de Esmeraldas.

CAPÍTULO V

5 Conclusiones y Recomendaciones

5.1 Conclusiones.

- Las zonas donde la probabilidad de presencia de *Anopheles spp.* es alta (>60%) comprenden: el litoral ecuatoriano, específicamente la zona costera de la provincia de Esmeraldas, parte de la provincia de Santa Elena, Santo Domingo de los Tsáchilas, la ciudad de Guayaquil y poblados aledaños; así mismo, en parte de la amazonia como la ciudades del Puyo, Tena y zonas selváticas. Finalmente, las zonas bajas de la Cordillera de los Andes.
- Uno de los puntos críticos para mejorar la calidad de los resultados del modelo es la selección de variables y la cuantificación de su contribución al fenómeno que tratan de evaluar, por ello la importancia de un análisis previo de las variables medioambientales y geoespaciales.
- El modelo que mejor describe el nicho ecológico o distribución potencial de las especies anofelinas en el Ecuador es el modelo Maxent, ya que demostró estadísticamente tener mejor desempeño que los demás, seguido del modelo de regresión logística que también mostro un buen ajuste. Por último, el modelo de evaluación multicriterio con lógica difusa cuyo resultado de predicción resultó ser muy generalizado, derivando en una pobre predicción (underfitting); no obstante, es muy útil para un análisis exploratorio del fenómeno.
- Maxent es uno de los modelos de distribución de especies más completos por su extenso soporte matemático, por su versatilidad para manejar todo tipo de información, y por la facilidad de aplicación dentro de su propio entorno.

- Se encontró que la provincia de Esmeraldas posee factores potencialmente influyentes para la presencia de *Anopheles spp.* como el tipo clima, el tipo de zona ecológica donde se encuentra y el tipo de uso de suelo predominante dentro de la provincia.
- El presente estudio pone en evidencia la importancia de realizar investigaciones geoespaciales enfocadas en el campo biológico y epidemiológico ya que es un buen aporte para la planificación territorial y la salud pública.

5.2 Recomendaciones

- Es pertinente contar con un muestreo de campo específico para el modelamiento de especies, ya que los datos de la mayor parte de investigaciones son obtenidos generalmente mediante información secundaria, como es el caso de los puntos de presencia, o mediante información virtual como los puntos de pseudoausencia, lo que implica un sesgo previo al modelamiento; especialmente si se utilizan modelos de regresión en los cuales es indispensable contar con puntos de ausencia comprobados para un buen desempeño.
- Es necesario generar coberturas climáticas óptimas como insumo para los modelos, ya que los datos climáticos existentes para Ecuador son limitados y poco confiables, lo que conlleva a recurrir a información externa y de menor resolución.
- Es preponderante contar con información actualizada y profunda sobre los factores ecológicos relativos a la especie para obtener una predicción más precisa en base a las variables que mejor describan al fenómeno.
- Es evidente la necesidad de realizar un estudio a profundidad en las zonas que presentan mayor riesgo, ya que existen diversos factores que influyen la

presencia de *Anopheles spp.* en el territorio, muchos de ellos relativos a la población y su entorno como factores sociales y económicos.

6 Anexos

- 6.1 Código en lenguaje R para regresión logística**
- 6.2 Mapa de distribución de *Anopheles spp.* – Evaluación Multicriterio**
- 6.3 Mapa de distribución de *Anopheles spp.* – Regresión logística**
- 6.4 Mapa de distribución de *Anopheles spp.* – Maxent**
- 6.5 Mapa de Uso de Suelo de la provincia de Esmeraldas**
- 6.6 Mapa de Zonas de Vida de la provincia de Esmeraldas**
- 6.7 Mapa de Factores Climáticos de la provincia de Esmeraldas – Temperatura Media Anual**
- 6.8 Mapa de Factores Climáticos de la provincia de Esmeraldas – Precipitación Media Anual**

7 Bibliografía

- Almache, J. (2013). Lógica clásica y lógica difusa: Facetas que las caracterizan. *Estoa No.2*, 91-101.
- Barbet-Massin, M., Jiguet, F., Albert, C., & Tulier, W. (2012). Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, 327-338.
- Collins, M., Schapire, R., & Singer, Y. (2002). Logistic regression, adaboost and bregman distances. *Machine Learning No.48*, 253-285.
- Delgado, T. (2011). *Evolución de la diversidad vegetal en Ecuador ante un escenario de cambio global (Tesis doctoral)*. Madrid: Universidad Complutense de Madrid.
- Donnelly, M., Simard, F., & Lehmann, T. (2002). Evolutionary studies of malaria vectors. *Trends in Parasitology*, 75-80.
- Dudik, M., Phillips, S., & Schapire, R. (2004). Performance guarantees for regularized maximum entropy density estimation. *Proceedings of the 17th Annual Conference on Computational Learning Theory*.
- Eastman, R. (2015). *IDRISI 18: Terrset edition*. Worcester, MA, USA: Clark Labs, Clark University.
- Elith, J., Graham, C., Anderson, R., Dudik, M., Ferrier, S., Antoine, G., . . . Zimmermann, N. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography N. 29*, 128-151.
- Elith, J., Phillips, S., Hastie, T., Dudik, M., Young, E. c., & Yates, C. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions No.17*, 43-57.
- Elton, C. (1927). *Animal Ecology*. London: Sidgwick and Jackson.
- Felicísimo, Á., Rubén, M., Muñoz, J., Felipe, B., Sánchez, J., Salvatierra, C., & Herrera, M. (2011). *FORCLIM, Bosques y cambio global. I Modelos de distribución de especies: Fundamentos de las IDE Argentina*. Madrid: Programa CYTED.
- Fielding, A., & Bell, J. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 38-49.

- Franklin, J. (2010). *Mapping species distribution: Spatial inference and prediction*. Cambridge - Inglaterra: Cambridge University Press.
- Fuller, D., Troyo, A., Alimi, T., & Beier, J. (2014). Participatory risk mapping of malaria vector exposure in northern South America using environmental and population data. *Applied Geography No.48*, 1-7.
- Gómez, M., & Barredo, J. (2005). *Sistemas de Información Geográfica y la evaluación multicriterio*. Madrid: RA-MA.
- Graham, C., Ferrier, S., Huetteman, F., Moritz, C., & Peterson, A. (2004). New developments in museum-based informatics and applications in biodiversity. *Trends in Ecology & Evolution No.19*, 497-503.
- Graham, C., Loiselle, B., Velásquez-Tibatá, J., & Cuesta, F. (2012). Modelos de distribución de especies y el desafío de pronosticar distribuciones futuras. *Cambio climático y la biodiversidad en los Andes Tropicales*, 349-367.
- Grinnell, J. (1924). Geography and Evolution. *Ecology No.5*, 225-229.
- Guisan, A., & Zimmermann, N. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling No.135*, 147-186.
- Hijmans, R. L., Cameron, S., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology No.25*, 1965-1978.
- Hutchinson, E. (1953). The Concept of Pattern in Ecology. *Proceedings of the Academy of Natural Sciences of Philadelphia Vol 105*, 1-12.
- Hutchinson, G. (1957). Concluding remarks. *Symp. Quant. Biol.*, 22 (págs. 415-427). Cold Spring Harbor.
- IEE. (2014). *Proyecto: Modelos matemáticos de dinámica ecológica sobre Chagas, Leishmania y Malaria mediante el desarrollo y aplicación de tecnología espacial*. Quito: Instituto Espacial Ecuatoriano.
- INEC. (2010). *Censo de población y vivienda en el Ecuador 2010*. Quito: Instituto Nacional de Estadísticas y Censos.
- INEC, I. N. (2014). *Encuesta de condiciones de vida 2013-2014*. Quito-Ecuador: INEC.

- INS, I. N. (2010). *Gestión para la vigilancia entomológica y control de la transmisión de malaria*. Bogotá-Colombia: Ministerio de la Protección Social.
- Jaynes, E. (1957). Information theory and statistical mechanics. *Physical review*, 620.
- Kleinbaum, D., & Klein, M. (2010). *Logistic Regression: A Self-Learning Text*. Nueva York-EEUU: Springer.
- Martínez Ainsworth, N. (2013). El nicho ecológico: útil concepto aún en debate. *Cienciorama No.1*, 1-11.
- Martínez, N. (2010). *Apuntes sobre modelación de nichos ecológicos*. México D.F: Laboratorio de ecología de la Universidad Nacional Autónoma de México.
- Mateo, R., Felicísimo, A., & Muñoz, J. (2011). Modelos de distribución de especies: Una revisión sintética. *Revista Chilena de Historia Natural Vol.84*, 217-240.
- Merow, C., Smith, M., & Silander, J. (2013). A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography No.36*, 1058-1069.
- Milesi, F., & de Casenave, J. (2005). El concepto de nicho en ecología aplicada: del nicho al hecho hay mucho trecho. *Ecología Austral 15*, 131-148.
- Miller, J. (2010). Species Distribution Modeling. *Geography Compass No.46*, 490-509.
- Montoya-Lerma, J., Solarte, Y., Gloria, G.-C., Quiñones, M., Ruíz-López, F., Wilkerson, R., & González, R. (2011). Malaria vector species in Colombia - A review. *Memórias del Instituto Oswaldo Cruz Vol.106*, 223-238.
- MSPC, M. d. (2010). *Guía Para la Atención Clínica Integral del Paciente con Malaria*. Bogotá - Colombia.
- Muñoz, Á., & Recalde, C. (2010). *Reporte Metodológico sobre el Experimento de Predicibilidad de Malaria en el Litoral Ecuatoriano*. Quito-Ecuador: PROYECTO INAMHI-MAE-SCN-PRAA-PACC.
- Osborn, F., Rubio-Palis, Y., Herrera, M., Figuera, A., & Moreno, J. (2004). Caracterización Ecoregional de los Vectores de Malaria en Venezuela. *Boletín de malariología y salud ambiental*, 77-92.
- Padilla, O. (2007). Fundamento Teórico Para Modelización De Variables Mediante Operadores Difusos. *Revista Geoespacial No. 4*.

- Palazón, J., & Calvo, J. (2009). *Modelos de Regresión*. Murcia-España: Universidad de Murcia.
- Pearson, R. (2008). *Species' Distribution Modeling for Conservation Educators and Practitioners*. American Museum of Natural History: Center for Biodiversity and Conservation & Department of Herpetology.
- Pearson, R. (2014). *Nerc: Advanced Training Short Course: Species Distribution Modelling*. Londres - Inglaterra: University College London.
- Pearson, R., & Dawson, T. (2003). Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography*, Vol 12, 361-371.
- Peterson, A., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Mayer, E., Nakamura, M., & Bastos Araújo, M. (2011). *Ecological Niches and Geographic Distributions (Monographs in population biology No. 49)*. New Jersey: Princeton University Press.
- Phillips, S., Anderson, R., & Schapire, R. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 231-259.
- Phillips, S., Dudik, M., & Schapire, R. (2004). A Maximum Entropy Approach to Species Distribution Modeling. *In Proceedings of the Twenty-First International Conference on Machine Learning*, 655-662.
- Pietra, S., Pietra, V., & Lafferty, J. (1997). Inducig features of random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence No.19*, 380-393.
- Pinault, L., & Hunter, F. (2011). New highland distribution records of multiple Anopheles species in the Ecuadorian Andes. *Malaria Journal*, 236-246.
- Prugnolle, F., & Yalcindag, E. (2012). Paludismo:¿Cómo han llegado a América? *Actualité Scientifique No.394*, 2-3.
- Riquelme, J. (2013). Técnicas para el análisis de patrones de uso del espacio en felinos americanos (Mammalia: Carnivora). *Gestión Ambiental No.25*, 53-84.
- Rotela, C., De Elia, E., Elorza, C., Horlent, N., Lamaro, A., Lanfri, M., . . . Torrusio, S. (2014). *Epidemiología panorámica: introducción al uso de herramientas geoespaciales aplicadas a la salud pública*. Ciudad Autónoma de Buenos Aires:

- Comisión Nacional de Actividades Espaciales; Ministerio de Planificación Federal Inversión Pública y Servicios Ministerio de Salud de la Nación.
- Sainz de la Maza, M. (2000). SIG y Modelización de Fauna Protegida: Delimitación de Hábitat Potencial del Urogallo (*Tetrao Urogallus*) mediante Lógica Borrosa (Fuzzy Logic) en el Parque Natural del Cadí-Moixeró (Pirineo oriental). *Tecnologías Geográficas para el Desarrollo Sostenible*, 352-366.
- Silero, N., Barbosa, M., Martínez-Freira, F., & Real, R. (2010). Los modelos de nicho ecológico en la herpetología ibérica: pasado, presente y futuro. *Boletín de la Asociación Herpetológica No.21*, 2-23.
- Sinka, M., Rubio-Palis, Y., Manguin, S., Patil, A., Temperley, W., Gething, P., . . . Hay, S. (2011). The dominant Anopheles vectors of human malaria in the Americas: occurrence data, distribution maps and bionomic précis. *Parasites & Vectors No.3*, 1-26.
- SNEM, S. N. (2013). *Proyecto Vigilancia y Control de Vectores para la Prevención de la Transmisión de Enfermedades Metaxénicas en el Ecuador 2013 - 2017*. Guayaquil: Ministerio de Salud Pública.
- Soberón, J. (2007). Grinnellian and Eltonian niches and geographic distribution of species. *Ecology Letters No.10*, 1115-1123.
- Soberón, J. (08 de Octubre de 2015). *Nichos y Áreas de Distribución*. Obtenido de CONABIO: <http://nicho.conabio.gob.mx/>
- Soberón, J., & Peterson, A. (2005). Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics No.2*, 1-10.
- Soria, E. (2013). *Eliminación de Malaria en el Ecuador: Análisis Descriptivo de la Prevalencia de Casos de Malaria en el Área IX Sucumbíos -Orellana entre los años 2004 – 2012 (Tesis de pregrado)*. Quito-Ecuador: Universidad San Francisco de Quito.
- Vaughan, I., & Ormerod, S. (2003). Improving the quality of distribution models for conservation by addressing shortcomings in the field collection of training data. *Conservation Biology No.17*, 1601-1611.

- Vilalta y Perdomo, J. (2005). Como enseñar la autocorrelación espacial. *Economía, sociedad y territorio*.
- Villar Hernández, B. (2014). *Modelos bayesianos para la distribución de especies con registros de solo presencias*. Montecillo, México: Tesis de maestría no publicada. Universidad Autónoma Chapingo,.
- Ward, G., Hastie, T., Barry, S., Elith, J., & Leathwick, J. (2009). Presence:only Data and the EM Algorithm. *Biometrics No.65*, 554-563.