



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

**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 DE DISTRIBUCIÓN ESPACIAL DEL
VECTOR *Aedes Aegypti* TRANSMISOR DEL VIRUS ZIKA
EN EL ECUADOR CONTINENTAL, MEDIANTE LA
APLICACIÓN DE HERRAMIENTAS GEO-INFORMÁTICAS.**

AUTOR: BALSECA CARRERA, MARIO BOLIVAR

DIRECTOR: ING. PADILLA ALMEIDA, OSWALDO VINICIO

SANGOLQUÍ - ECUADOR

FEBRERO 2018



**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 DISTRIBUCIÓN ESPACIAL DEL VECTOR Aedes Aegypti TRANSMISOR DEL VIRUS ZIKA EN EL ECUADOR CONTINENTAL, MEDIANTE LA APLICACIÓN DE HERRAMIENTAS GEO-INFORMÁTICAS*" realizado por el señor *MARIO BOLÍVAR BALSECA CARRERA*, ha sido revisado en su totalidad y analizado por el software anti-plagio, el mismo cumple con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de las Fuerzas Armadas ESPE, por lo tanto me permito acreditarlo y autorizar al señor *MARIO BOLÍVAR BALSECA CARRERA* para que lo sustente públicamente.

Sangolquí, 08 de febrero del 2017

Atentamente,

Director

Ing. Oswaldo Vinicio Padilla Almeida



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

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

AUTORÍA DE RESPONSABILIDAD

Yo, *MARIO BOLÍVAR BALSECA CARRERA*, con cédula de identidad N° 1717667420 declaro que este trabajo de titulación “*MODELAMIENTO DE DISTRIBUCIÓN ESPACIAL DEL VECTOR AEDES AEGYPTI TRANSMISOR DEL VIRUS ZIKA EN EL ECUADOR CONTINENTAL, MEDIANTE LA APLICACIÓN DE 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, 08 de enero del 2017

MARIO BOLÍVAR BALSECA CARRERA

C.C. 171766742-0



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

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

AUTORIZACIÓN

Yo, **MARIO BOLÍVAR BALSECA CARRERA**, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar en la biblioteca virtual de la institución el presente trabajo de titulación "**MODELAMIENTO DE DISTRIBUCIÓN ESPACIAL DEL VECTOR AEDES AEGYPTI TRANSMISOR DEL VIRUS ZIKA EN EL ECUADOR CONTINENTAL, MEDIANTE LA APLICACIÓN DE HERRAMIENTAS GEO-INFORMÁTICAS**", cuyo contenido, ideas y criterios son de mi autoría y responsabilidad.

Sangolquí, 08 de febrero del 2017



MARIO BOLÍVAR BALSECA CARRERA

C.C. 171766742-0

DEDICATORIA

A Dios por brindarme la fortaleza para seguir adelante a pesar de todas las adversidades que se me han presentado a lo largo de mi vida.

A mi papa por apoyarme incondicionalmente en todo, al cual le debo en su mayoría la satisfacción de cumplir esta meta.

A mi mama, mi guerrera, mi fuerza que me ha enseñado el gran valor de luchar por lo que quieres, así sea por minutito más de vida.

A mis hermanas quienes a pesar de sus regaños y a veces no compaginar en ciertas cosas, siempre me han brindado un oportuno consejo y unas palabras de aliento, las amo ñañitas.

A mí porque a pesar de cometer muchos errores y varios tropezones siempre me supe levantar para poder concluir esta meta que me trace hace algún tiempo atrás.

Mario B.

AGRADECIMIENTO

A mi familia por brindarme su apoyo incondicional y siempre motivarme a ser una mejor persona lleno de buenos valores, a mi abuelita, a mis tíos en especial a mi Tía Negrita mi segunda madre gracias por cuidarme como uno más de sus hijos.

A la Universidad de las Fuerzas Armadas-ESPE, especialmente a la Carrera de Ingeniería Geográfica y del Medio Ambiente y todos sus docentes por transmitirme el conocimiento necesario para alcanzar este logro.

A mi tutor de Proyecto Ing. Oswaldo Padilla quien desinteresadamente colaboro para que el proyecto concluya satisfactoriamente.

A los Ingenieros León y Guamán excelentes profesionales con calidad humana que motivan el potencial del estudiante.

A mi amigo Byron el cual ha sido como un hermano para mí, muchas gracias por formar parte de mi familia.

A mis amigos y ahora colegas Pablo y Alfonso, definitivamente la culminación de este proyecto final no se hubiese alcanzado sin su ayuda muchas gracias.

A lo largo de mi paso por la universidad encontré a muchos amigos en el camino a los cuales debo nombrar especialmente a Santiago Jumbo y Gaby Rondal muchas gracias por todo de corazón; a Mateo Navas, Johanna Collaguazo, Bryan Sánchez, Ulises Pupiales, Carolina Tamayo, Alejandro Martínez, Andrea Sambache, Vanessa Tenorio, Daniel López, David Velasco, Irene Quishpe, Victor Villamarín, Daniela Guallichico, Chirstian Bara y Gaby Wolf muchas gracias por brindarme su amistad y apoyo.

A las personas que en su debido momento fueron importantes para mí, pero por circunstancias de la vida nos hemos distanciado, gracias.

Mario B.

TABLA DE CONTENIDO

CERTIFICACIÓN	ii
AUTORÍA DE RESPONSABILIDAD	iii
AUTORIZACIÓN	iv
DEDICATORIA	v
AGRADECIMIENTO	vi
TABLA DE CONTENIDO	vii
LISTA DE TABLAS	x
LISTA DE FIGURAS	xi
RESUMEN	xiii
ABSTRACT	xiv
CAPITULO I	1
GENERALIDADES	1
1.1 Resumen	1
1.2 Antecedentes	1
1.3 Definición del Problema	3
1.4 Justificación	3
1.5 Objetivo general	4
1.6 Objetivos específicos	4
1.7 Metas	4
1.8 Área de Estudio	5
CAPITULO II	6
MARCO TEÓRICO	6
2.1 Resumen	6
2.2 Hábitat y Nicho Ecológico	6
2.2.1 Definición de Nicho Ecológico	7
2.2.2 Tipos de Nicho Ecológico	7

2.2.3	Diagrama de BAM	8
2.3	Modelos de Nicho Ecológico (MNE)	10
2.3.1	Modelado de nicho ecológico y el área de distribución de las especies ..	10
2.3.2	Procesos para el modelamiento del Nicho ecológico	11
2.4	Métodos para el Modelamiento de Nicho ecológico	15
2.4.1	Máxima Entropía (Maxent)	16
2.4.2	Lógica Difusa o FUZZY	20
2.4.3	Regresión Logística.....	28
2.4.4	Splines de Regresión Multivariada Adaptativa (MARS)	31
2.5	Validación de los modelos de Nicho ecológico	34
2.5.1	Curva ROC (Receiver operating characteristic analysis).....	37
2.5.2	Limitaciones de los modelos de Nicho ecológico.....	38
2.6	Evaluación multicriterio (EMC).....	39
2.6.1	Normalización de Variables	40
2.6.2	Método de Análisis Jerárquico AHP (Thomas Saaty)	40
2.7	Sistemas de Información Geográfica (SIG)	42
2.7.1	Base de Datos Geográfica (Geodatabase)	43
2.7.2	Aplicación y uso de los SIG en los Modelos de Especies	44
2.7.3	Uso de SIG y la vigilancia en Salud Pública	44
2.8	Biología del Vector Aedes aegypti	46
2.8.1	Clasificación Taxonómica	46
2.8.2	Ciclo de Vida	48
2.8.3	Hábitat y Alimentación	49
CAPITULO III		51
METODOLOGÍA.....		51
3.1	Resumen.....	51
3.2	Esquema Metodológico	51
3.3	Recopilación de Información	52
3.3.1	Variables Dependientes	52
3.3.2	Variables Independientes.....	55
3.4	Normalización de las variables.....	64

3.5	Modelos de distribución de las especies	67
3.5.1	Maxent (Máxima Entropía)	67
3.5.2	Lógica Difusa.....	73
3.5.3	Regresión Logística.....	76
3.6	Desviación Estándar y Ajuste de los Modelos	81
3.6.1	Desviación Estándar de los modelos.....	81
3.6.2	Ajuste de los modelos	82
CAPITULO IV		83
RESULTADOS Y DISCUSIONES		83
4.1	Resumen.....	83
4.2	Evaluación de los modelos.....	83
4.2.1	Análisis del ajuste de los modelos	83
4.2.2	Análisis de la Curva AUC de los modelos	84
4.2.3	Análisis Gráfico de los modelos	87
4.2.4	Selección del Mejor Modelo.....	90
4.3	Delimitación de zonas idóneas para la presencia del vector.....	91
4.4	Caracterización de las zonas con mayor riesgo epidemiológico	94
CAPITULO V		98
CONCLUSIONES Y RECOMENDACIONES		98
5.1	Resumen.....	98
5.2	Conclusiones	98
5.3	Recomendaciones	99
BIBLIOGRAFÍA		101

LISTA DE TABLAS

Tabla 1. Técnicas para el Modelamiento de Nicho Ecológico.....	15
Tabla 2. Casos Lógica Fuzzy	27
Tabla 3. Ciclo de vida <i>Aedes aegypti</i>	48
Tabla 4. Matriz de Saaty para Uso de Suelo	61
Tabla 5. Variables Predictoras para el modelamiento del <i>Aedes aegypti</i>	63
Tabla 6. Valores máximos y mínimos de las Variables Independientes.....	65
Tabla 7. Análisis de las Variables.....	73
Tabla 8. Desviación Estándar de los Modelos	81
Tabla 9. Ajuste de los Modelos	82
Tabla 10. Análisis del Ajuste de los Modelos	84
Tabla 11. Provincias con mayor riesgo Epidemiológico	93
Tabla 12. Caracterización de la Provincia de Manabí	94
Tabla 13. Caracterización de la Provincia de El Oro.....	95

LISTA DE FIGURAS

Figura 1. Mapa de ubicación del Ecuador.....	5
Figura 2. Diagrama de BAM.....	9
Figura 3. Modelado de Nicho Ecológico	11
Figura 4. Esquema de Probabilidades en el modelo de Maxent.....	17
Figura 5. Probabilidad tomando una variable	22
Figura 6. Puntos importantes del Modelo	23
Figura 7. Probabilidad Creciente	24
Figura 8. Función logística.....	29
Figura 9. Funciones base usados por MARS	33
Figura 10. Matriz de confusión	35
Figura 11. Matriz de confusión representada en el Espacio Geográfico	36
Figura 12. Curva ROC (Receiver operating characteristic analysis).....	37
Figura 13. Escala fundamental de prioridades de Saaty	41
Figura 14. Representación gráfica de un SIG	43
Figura 15. Uso de SIG en análisis de situación de salud	45
Figura 16. Vector <i>Aedes aegypti</i>	46
Figura 17. Familia <i>Aedes aegypti</i>	47
Figura 18. Ciclo de vida <i>Aedes aegypti</i>	48
Figura 19. Esquema Metodológico.....	51
Figura 20. Casos confirmados autóctonos e importados del Virus Zika	53
Figura 21. Presencia de Vector <i>Aedes aegypti</i> en el Ecuador Continental	54
Figura 22. Pseudo-ausencia de Vector <i>Aedes aegypti</i> en el Ecuador Continental	55
Figura 23. Variables Bioclimáticas para el modelamiento del <i>Aedes aegypti</i>	56
Figura 24. Variable Altitud para el modelamiento del <i>Aedes aegypti</i>	57
Figura 25. Variable Inundaciones para el modelamiento del <i>Aedes aegypti</i>	58
Figura 26. Variable Centros Poblados para el modelamiento del <i>Aedes aegypti</i>	59
Figura 27. Presencia del <i>Aedes aegypti</i> en función del Uso de Suelo	60
Figura 28. Variable Uso de Suelo para el modelamiento del <i>Aedes aegypti</i>	62
Figura 29. Variable Índice de Pobreza para el modelamiento del <i>Aedes aegypti</i>	63
Figura 30. Reclasificación de valores de las Variables	66
Figura 31. Datos de presencia en formato .CSV	67

Figura 32. Interface principal Maxent	68
Figura 33. Configuración Maxent	69
Figura 34. Curva ROC (Receiver Operating Characteristic analysis)	70
Figura 35. Contribución de las variables predictoras	71
Figura 36. Test Jackknife en función del AUC	72
Figura 37. Curvas de respuesta.....	73
Figura 38. Matriz de las variables en SPSS	76
Figura 39. Configuración de Variables.....	77
Figura 40. Tabla coeficientes de las variables.....	77
Figura 41. Matriz de las variables en MARS	79
Figura 42. Configuración Programa MARS	79
Figura 43. Definición de Funciones Base	80
Figura 44. Análisis de la Curva ROC - Modelo Maxent	85
Figura 45. Análisis de la Curva ROC - Modelo Lógica Difusa	85
Figura 46. Análisis de la Curva ROC - Modelo Regresión Logística	86
Figura 47. Análisis de la Curva ROC - Modelo MARS	86
Figura 48. Análisis de la Curva ROC - General.....	87
Figura 49. Distribución del <i>Aedes aegypti</i> (Modelo Maxent).....	88
Figura 50. Distribución del <i>Aedes aegypti</i> (Modelo Lógica Difusa).....	89
Figura 51. Distribución del <i>Aedes aegypti</i> (Modelo Regresión Logística).....	89
Figura 52. Distribución del <i>Aedes aegypti</i> (Modelo MARS).....	90
Figura 53. Histograma de Probabilidades	91
Figura 54. Probabilidades de Presencia del vector <i>Aedes aegypti</i>	92
Figura 55. Zonas idóneas para la Presencia del vector <i>Aedes aegypti</i>	92

RESUMEN

El vector *Aedes aegypti* es el principal causante del Dengue, la Chikungunya y el virus del Zika, enfermedades que se transmiten por picaduras de hembras infectadas; en el Ecuador el Ministerio de Salud Pública (MSP) ha emitido la alerta epidemiológica de cumplimiento obligatorio para los establecimientos de salud de la Red Pública Integral de Salud. Existen actualmente software geo-informáticos y estadísticos que permiten modelar el nicho ecológico de las especies tales como: Maxent, Regresión Logística, MARS o Lógica Difusa que permiten determinar probabilísticamente zonas idóneas en las que exista presencia de la especie en función de las variables dependientes (presencia y ausencia) e independientes (capas climáticas, geográficas, etc.). Estos modelos requieren previamente contar con insumos de entrada previamente ponderados, normalizados y rasterizados para continuar con el proceso de validación para estimar su rendimiento predictivo a través de varios estadísticos como por ejemplo: la matriz de confusión (errores de comisión y omisión) o la curva ROC que determina el área bajo la curva (AUC). Dando como resultado el método de Maxent con mayor rendimiento predictivo con un valor de $AUC=0,998$, además visualmente describe las zonas con mayor probabilidad del vector asemejándose a la distribución real de la especie en función a los datos de presencia y las variables predictoras; tras el análisis anterior se determinó que, en gran parte del territorio de la región litoral probabilísticamente existe presencia del vector especialmente en las provincias de El Oro y Manabí con valores de 23.22% y 17.96% respectivamente siendo las más vulnerables ante un posible riesgo epidemiológico.

PALABRAS CLAVES

- **MODELO DE NICHOS ECOLÓGICO**
- **MAXENT**
- **REGRESIÓN LOGÍSTICA**
- **LÓGICA DIFUSA**
- **MARS**
- **AEDES AEGYPTI**

ABSTRACT

The vector *Aedes aegypti* is the main cause of Dengue, Chikungunya and Zika virus, diseases that are transmitted by stings of infected females; In Ecuador, the Ministry of Public Health (MSP) has issued the mandatory epidemiological alert for health establishments of the Public Health Integral Network. There are currently geo-software and statistical software that allow modeling the ecological niche of species such as: Maxent, Logistic Regression, MARS or Fuzzy Logic that allow probabilistic identification of suitable areas in which the species is present depending on the dependent variables (Presence and absence) and independent (climatic, geographical, etc.). These models require pre-weighted input, normalized and rasterized inputs to continue the validation process to estimate their predictive performance through several statistics such as the confusion matrix (commission and omission errors) or the curve ROC that determines the area under the curve (AUC). As a result the Maxent method with higher predictive performance with a value of AUC = 0.998, also visually describes the areas with greater probability of the vector resembling the actual distribution of the species as a function of the presence data and the predictor variables; After the previous analysis it was determined that, in a large part of the territory of the coastal region probabilistically presence of the vector especially in the provinces of El Oro and Manabí with values of 23.22% and 17.96% respectively being the most vulnerable before a possible epidemiological risk.

KEYWORDS

- **ECOLOGICAL NICHE MODEL**
- **MAXENT**
- **LOGISTIC REGRESSION**
- **DIFFUSED LOGIC**
- **MARS**
- **AEDES AEGYPTI**

CAPITULO I

GENERALIDADES

1.1 Resumen

El vector *Aedes aegypti* es el principal causante del Dengue, la Chikungunya y el virus del Zika, enfermedades que se transmiten por picaduras de hembras infectadas; en el Ecuador el Ministerio de Salud Pública (MSP) ha emitido la alerta epidemiológica de cumplimiento obligatorio para los establecimientos de salud de la Red Pública Integral de Salud. Existen actualmente herramientas geo-informáticas y estadísticas que permiten modelar el nicho ecológico de la especie en estudio con el fin de identificar los mecanismos y acciones para contrarrestar una posible propagación de epidemias.

1.2 Antecedentes

En el año de 1947 el virus se determinó por primera vez en el país de Uganda, particularmente en los bosques de Zika. Se descubrió en un mono Rhesus cuando se realizaba un estudio acerca de la transmisión de la fiebre amarilla en la selva. En el año 2007 tuvo lugar el primer brote importante de infección por virus Zika en la Isla de Yap (Micronesia) en la que se notificaron 185 casos sospechosos. Posteriormente se registró un brote en la Polinesia Francesa, que inició a final de octubre de 2013 (CONAVE, 2015).

Se registraron alrededor de 10.000 casos de los cuales aproximadamente 70 casos fueron graves, con complicaciones neurológicas (síndrome de Guillain Barré, meningoencefalitis) o autoinmunes (púrpura trombopénica, leucopenia). En el 2014, se registraron también casos en Nueva Caledonia y en Islas Cook (CONAVE, 2015).

En febrero de 2014, las autoridades de salud pública de Chile confirmaron un caso de transmisión autóctona de infección por virus Zika en la isla de Pascua (Chile). Esta aparición coincidió con la presencia de otros focos de transmisión en islas del Pacífico: Polinesia Francesa, Nueva Caledonia, e Islas Cook (CONAVE, 2015).

Las Organizaciones Panamericana y Mundial de la Salud OPS/OMS y la Red de Laboratorios de Diagnóstico de Arbovirus (RELDA) de las Américas acordaron nuevas pautas que permitirán identificar y confirmar casos sospechosos de Zika en los países afectados, mientras la OMS y la comunidad científica buscan desarrollar pruebas más precisas (OMS, 2015).

El Ministerio de Salud Pública del Ecuador (MSP) ha emitido la alerta epidemiológica de cumplimiento obligatorio para los establecimientos de salud de la Red Pública Integral de Salud y la Red Complementaria (es decir, todos los establecimientos públicos y privados), lo que permitirá la detección oportuna de todo paciente que presente síntomas sospechosos, como fiebre inferior a 38,5 °C, inflamación de las articulaciones en manos y pies, manchas rojas en la piel, conjuntivitis.

La fiebre menor de 38,5 °C y la posible presencia de conjuntivitis diferencian los síntomas de la fiebre del Zika de los síntomas del dengue y la chikungunya (MSP, 2015). En el país el Sistema Nacional de Vigilancia y Alerta Temprana para el control del vector de dengue y fiebre amarilla ha propuesto un proyecto que plantea obtener información climatológica, socioeconómica y biológica del mosquito *Aedes aegypti* para desplegarla con Sistemas de Información Geográfica con la finalidad de desarrollar un Sistema de Alerta Temprana para el Control del Vector de dengue y fiebre amarilla, desarrollando modelos matemáticos predictivos para dengue basados en la relación entre los datos entomológicos, epidemiológicos, socioeconómicos y climatológicos (MSP, 2015).

Estos modelos generan predicciones respecto a la distribución y los requerimientos ambientales de las especies, posibilitan la identificación de las variables que mejor predicen los hábitats favorables, permiten testear hipótesis ecológicas acerca de la distribución de los organismos y evaluar los impactos de posibles cambios ambientales. Este tipo de modelado ecológico de distribución (Máxima Entropía, Regresión Logística, etc.) es utilizado en este trabajo para identificar zonas adecuadas para el desarrollo y proliferación del *Aedes aegypti* (Rotela, 2014).

1.3 Definición del Problema

El modelamiento de la distribución espacial de las especies es un conjunto de técnicas y de procesos estadísticos que en los últimos años han demostrado resultados probabilísticos aceptables dentro de diferentes áreas de estudio como herramientas útiles para la gestión y conservación de la biodiversidad; dentro del área epidemiológica estos estudios permiten identificar las zonas con mayor probabilidad de presencia del vector *Aedes aegypti* y el impacto que provoca dentro de su entorno; a sus vez proporciona información necesaria para establecer estrategias de intervención que se deben tomar ante un posible riesgo epidemiológico.

En el Ecuador la investigación sobre la distribución geográfica potencial del vector *Aedes aegypti* es limitada debido a que la información sobre este, es prácticamente nueva y se requiere de análisis en campo y laboratorio para identificarlos debido a que esta especie se adapta fácilmente y coloniza nuevos nichos ecológicos

1.4 Justificación

Las enfermedades vectoriales representan un 17% de la carga mundial estimada de enfermedades infecciosas e incluso en algunos casos letales; en el país existe presencia endémica del mosquito *Aedes aegypti* que está estrechamente relacionada con fenómenos climáticos (temperatura y humedad) causando pérdidas económicas directas e indirectas que afectan principalmente a los estratos bajos de la sociedad (MSP, 2015).

Dentro de las políticas y lineamientos estratégicos del Plan Nacional del Buen Vivir 2013-2017 se establece, promover la investigación en servicios sanitarios, en articulación con el Sistema de Vigilancia Epidemiológica, que permita la detección oportuna de patologías, virus y demás enfermedades, así como la identificación de mecanismos y acciones para contrarrestar una posible propagación de epidemias (PNBV, 2013).

Considerando que el virus del Zika es una enfermedad compleja que se adapta fácilmente y coloniza nuevos nichos ecológicos es necesario utilizar un enfoque interdisciplinario para entender su epidemiología, estudiando la ecología de su vector, 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 del *Aedes aegypti*, transmisor del parásito que causa el virus del Zika, para delimitar las zonas de mayor riesgo biológico para la población, 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.

1.5 Objetivo general

Modelar el nicho ecológico de la especie *Aedes aegypti* vector del virus Zika en el Ecuador continental mediante la aplicación de modelos probabilísticos y herramientas geo-informáticas, que generen un referente para la planificación en el control y prevención ante un posible riesgo epidemiológico.

1.6 Objetivos específicos

- Definir, obtener y estructurar la información espacial con insumos para las variables dependientes e independientes de los modelos.
- Determinar los valores de las variables dependientes e independientes de los modelos en cada punto de muestreo.
- Generar y evaluar entre los modelos de distribución potencial de las especies, el más adecuado para el estudio del *Aedes aegypti* vector del virus Zika.
- Establecer la distribución geográfica y la potencial expansión del vector considerando factores espaciales.
- Delimitar y caracterizar las zonas con mayor riesgo epidemiológico.

1.7 Metas

- Estructurar una Geodatabase de las variables dependientes e independientes de los modelos generados.

CAPITULO II

MARCO TEÓRICO

2.1

2.1 Resumen

Los modelos de Nicho ecológico tales como: Maxent, Regresión Logística, MARS o Lógica Difusa se fundamentan en procesos estadísticos y cartográficos; para determinar probabilísticamente zonas idóneas en las que exista presencia de la especie en función de las variables dependientes (presencia y ausencia) e independientes (capas climáticas, geográficas, etc.). Estos modelos requieren ser validados para estimar su rendimiento predictivo a través de varios estadísticos como por ejemplo: la matriz de confusión (errores de comisión y omisión) o la curva ROC que determina el área bajo la curva (AUC). Otra herramienta útil en el modelamiento de la distribución de las especies son los Sistemas de Información Geográfica que permiten el manejo datos geográficos e incluso la generación de nueva información.

2.2 Hábitat y Nicho Ecológico

Hall (1997) define el hábitat como los recursos y las condiciones presentes en un área que propician su ocupación, e influyen en la sobrevivencia y reproducción de un organismo dado; considera también que la calidad del hábitat es la capacidad del ambiente de proveer las condiciones apropiadas para la persistencia de individuos y poblaciones, por lo que puede ser evaluada a diferentes niveles.

Por otra parte se define al nicho ecológico como algo más extenso que contiene el espacio físico y, por sobre todo, la función desempeñada por el organismo en el ecosistema como su adopción y uso de los recursos a su disposición. Dicho de otra manera los organismos de distintas especies pueden compartir el hábitat, pero es prácticamente imposible que coincidan en el mismo nicho ecológico, de darse esta particularidad se establece un fenómeno de exclusión competitiva o una de las especies deberá cambiar su nicho (Morales, 2015).

2.2.1 Definición de Nicho Ecológico

A principios del siglo XX se definió al nicho como las condiciones ambientales que delimitan la distribución de una especie, además de mencionar que si dos especies están establecidas en una misma fauna, regularmente no tienen las mismas relaciones de nicho. Sin embargo Hutchinson (1957) señala que el nicho está definido como la suma de todos los factores medioambientales que actúan sobre el organismo determinado en una región de un hiperespacio n dimensional compuesto por una serie de variables ambientales, ecológicas e históricas, que permiten la sobrevivencia de una especie.

2.2.2 Tipos de Nicho Ecológico

Dentro de los tipos de nicho ecológico se debe definir previamente que en una determinada comunidad una especie debe cumplir cierta función, explotando ciertas relaciones, pero no viceversa. Solo durante ciertas condiciones ambientales, como la competencia, o la emigración, una especie puede establecer nuevas relaciones teóricamente plausibles y tener de hecho mucha acogimiento en su nueva comunidad (Sillero, Barbosa, & Martínez, 2010).

- **El Nicho Fundamental**

Es un espacio ambiental donde la especie se mantiene en todo momento libre de competidores con n dimensiones que puede mantener una población viable y persistir a lo largo del tiempo, sin inmigración. Cada dimensión es una variable ambiental que influye sobre el nicho. Se puede considerar como un punto en el espacio multidimensional. (Cruz & Camargo, 2012).

- **El Nicho Efectivo**

Es el conjunto de condiciones ambientales (alimentación y espacio) dentro de las cuales un organismo existe en una comunidad. Es el nicho con presencia de competidores en donde algunas especies están sujetas a presiones de competencia por requerimientos, obligándolos a desplazarse del hipervolumen del nicho fundamental; el espacio que ocupan dichas especies, en el cual están bien adaptadas (Cruz & Camargo, 2012).

2.2.3 Diagrama de BAM

Para realizar estudios de los nichos ecológicos se utilizan diferentes técnicas y herramientas que facilitan la investigación de los mismos, una de ellas es el diagrama BAM (Biótico, Abiótico, Movimiento; que representa el espacio geográfico dividido en los tres principales factores que limitan la distribución de una especie: abióticos (A), bióticos (B), e históricos o de dispersión (M) (Sillero & Barbosa, 2010).

- ***Factores abióticos (A)***

Son aquellos que constituyen sus características físico-químicas (temperatura, luz, humedad, etc.), la importancia de estos factores para la vida y el equilibrio ecológico de nuestro planeta es muy grande, ya que estiman la distribución de los seres vivos sobre la Tierra y, además, influyen sobre ellos y sobre su adaptación al medio (Soberón & Peterson, 2005).

- ***Factores bióticos (B):***

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 la dispersión de semillas, polinizadores, y otros.) o negativas (competidores, predadores, parásitos) (Soberón & Hensz, 2012).

- ***Históricos o de dispersión (M):***

Las regiones que son accesibles para la dispersión de las especies además de su capacidad evolutiva 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) (Soberón & Hensz, 2012).

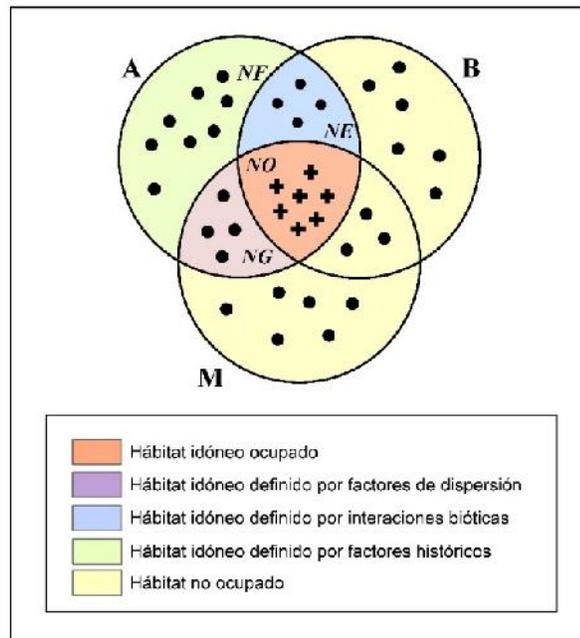


Figura 2. Diagrama de BAM

Fuente: (Sillero et al, 2010)

El hábitat sólo es adecuado para la especie dentro de la zona común a los tres factores (ver Figura 2). Dicha área representa el nicho ocupado (NO); la parte común A y M, el nicho de Grinnell (NG); el espacio común A y B, el nicho de Elton (NE); y toda el área de A es el nicho fundamental de Hutchinson (NF). La especie consigue sobrevivir en aquellas regiones climáticamente favorables, a las que logra dispersar y en las que no hay interacciones bióticas que la excluyan. Las regiones que no cumplan alguna de estas condiciones no son idóneas para la presencia de la especie (Sillero & Barbosa, 2010).

El Nicho ecológico describe la posición de una especie o población en su ecosistema o entre sí, para analizar su comportamiento se utiliza diferentes técnicas como el Diagrama de BAM que divide su espacio geográfico en factores bióticos, abióticos e históricos.

2.3 Modelos de Nicho Ecológico (MNE)

Los modelos de distribución de las especies es un grupo heterogéneo de técnicas y herramientas que se utilizan para analizar de una manera más objetiva la presencia de los organismos, se basan en procedimientos cartográficos y estadísticos que partiendo de datos reales de presencia, permiten inferir zonas potencialmente idóneas en función de sus características ambientales. (Muñoz & Felicísimo, 2012).

El resultado final de un MNE puede ser una representación espacial de los hábitats favorables a la presencia de una especie potencialmente invasora o vectores de enfermedades epidemiológicas (Miller, 2010).

Los modelos de nicho ecológico (MNE) pueden ser divididos de acuerdo al enfoque de investigación siendo correlativo y mecanicista. Dentro del enfoque correlativo este asocia datos estadísticos históricos de la distribución de las especies con las condiciones ambientales necesarias para que las especies sobrevivan, sin la identificación de los factores causales (Peterson, 2003).

Los MNE mecanicistas se basan en respuestas biológicas directas que usan para el modelamiento las observaciones de campo de las especies y las variables predictoras ambientales tales como (la temperatura, la humedad, alimento, y otros). Los datos observacionales de las especies (ocurrencias) pueden ser sólo presencias, o presencias y ausencias, incluso observaciones de abundancia (Guisan & Thuiller, 2005).

2.3.1 Modelado de nicho ecológico y el área de distribución de las especies

Para estimar los modelos de distribución de especies es importante entender la ecología e historia evolutiva de las mismas, que se representa dentro de un contexto geográfico definido por factores como condiciones abióticas, interacciones bióticas, capacidad de dispersión y adaptación a nuevas condiciones (Martínez, 2012).

En los MNE se analizan los datos geográficos de presencia de la especie y se comparan con la información disponible de las variables ambientales de modo que cada punto analizado tendrá información asociada respecto a las características (climáticas, edáficas, topográficas) del sitio (ver Figura 3).

En este caso el nicho modelado se asemeja más al realizado, pues los datos provienen de sitios en la naturaleza donde los organismos se encuentran interactuando unos con otros (Martínez, 2012).

Sin embargo se deben descartar ciertos resultados donde se observan especies en áreas que no proporcionan todos los recursos para la sobrevivencia de la especie, sin embargo existen zonas aledañas que si prestan todas las condiciones necesarias para la presencia de la especie; lógicamente estos eventos ocurren con mayor frecuencia en especies con alta capacidad de dispersión, tales como las aves (Martínez, 2012).

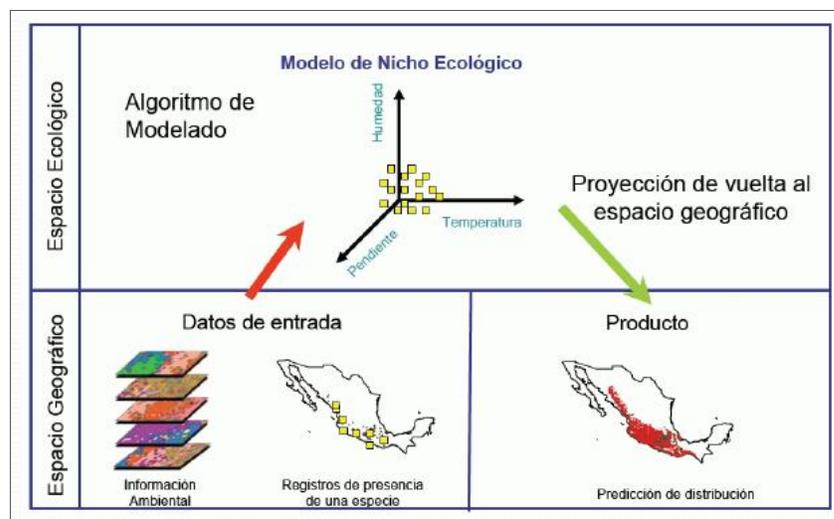


Figura 3. Modelado de Nicho Ecológico

Fuente: (Pearson, 2008)

2.3.2 Procesos para el modelamiento del Nicho ecológico

El modelamiento de nichos ecológicos es un proceso que requiere de una serie de decisiones que están relacionadas con la teoría ecológica en cada paso del proceso, desde la conceptualización, el objetivo del estudio, la clase, la calidad de datos disponibles y la habilidad del modelado de los modelos. (Ghaham, 2012).

El procedimiento para la MNE idealmente implica seis pasos:

- 1) La conceptualización: definición del modelo conceptual (sistema que será simulado), así como la selección del método de modelación, la resolución y extensión del área de estudio

- 2) Preparación de la información
- 3) Desarrollo del algoritmo de modelación y ajuste del modelo
- 4) Evaluación del modelo
- 5) Predicciones espaciales
- 6) Aplicación del modelo (Guisan & Thuiller, 2005).

Aunque el procedimiento puede variar dependiendo de factores como el tipo de algoritmos usados, todos ellos establecen e identifican el espacio ambiental en que se encuentran las especies, el cual es proyectado posteriormente al espacio geográfico para encontrar las áreas en donde se presentan estas condiciones ambientales favorables para las mismas (Alvarado & Serrano, 2013).

Como resultado de los MNE se obtienen las áreas de distribución potencial de las especies, generalmente representadas como mapas de distribución potencial, los cuales son abstracciones de los patrones espaciales y temporales del nicho ecológico de los organismos. Una vez obtenidos los modelos, también pueden ser proyectados a diferentes regiones y/o diferentes periodos de tiempo (Alvarado & Serrano, 2013).

2.3.2.1 Variables Dependientes (Datos de presencia, ausencia o pseudo-ausencia)

Las variables dependientes que determinan la distribución de las especies se miden de forma nominal y ordinal. En la mayoría de los estudios solo se dispone de datos de presencia de la especie, unos pocos estudios cuentan además con datos de ausencia y existen ocasiones en las que se tienen datos de abundancia. Son escasos los ensayos donde existen muestreos diseñados expresamente para estimar la distribución de una especie en particular (Guisan & Thuiller, 2005).

El aspecto más relevante a considerar sobre las variables dependientes 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; la información referente a la distribución de especies colecciones personales de los registros de ocurrencia de las especies que pueden ser obtenidos mediante recorridos de campo por un pequeño grupo de investigadores, además de colecciones de historia natural, museos o herbarios y finalmente recursos en línea disponible actualmente en internet,

como por ejemplo www.gbif.org o concretamente en países como México en el portal de la base de datos REMIB de la CONABIO (<http://www.conabio.gob.mx/>). Se debe señalar que la información puede presentar ciertos inconvenientes como se detallan a continuación:

- Solo registran datos de presencia y no de ausencia.
- Un porcentaje variable pero significativo presenta errores de georreferenciación e identificación taxonómica.
- Están colectados con diferentes fines y por diferentes recolectores, sin una estrategia común de muestreo, por lo que en muchas ocasiones son una representación sesgada de la distribución de la especie (Soberón & Hensz, 2012).

Para mitigar estos errores ciertos autores sugieren remuestrear los datos a pesar que el limitado número de datos disponibles hace imposible esta estrategia en la mayoría de los casos. 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 tienen la capacidad de lidiar con estos problemas (Graham, 2012).

2.3.2.2 Variables independientes (Predictoras)

Las variables independientes tienen un efecto sobre la distribución de las especies, ya sea de una forma directa o indirecta, debido a que tienen una correlación entre el medio abiótico y la presencia de las especies; descrita razonablemente bien por factores climáticos a escalas espaciales diversas y con diferentes resoluciones, gracias al incremento de diversas técnicas de modelamiento espacial y el uso de sensores remotos (Muñoz & Felicísimo, 2012).

Cuando se trabaja con escalas menores cientos de metros o menos, estas condicionan la selección de variables independientes a incluir en la modelización; es probable que la distribución responda a factores como variaciones micro-topográficas en las condiciones ambientales. Lo razonable es incluir variables que cumplan simultáneamente ciertas condiciones (Guisan & Azimermann, 2007).

La primera es que sean potencialmente explicativas, es decir, que tengan una relación potencial con la distribución de la especie, bien como factor limitante o como indicador a través de relaciones indirectas. La segunda es que muestren una variabilidad significativa en la zona de estudio y finalmente las variables independientes no estén excesivamente correlacionadas entre sí, especialmente si se busca una interpretación biológica de los resultados (Muñoz & Felicísimo, 2012).

Las variables comúnmente utilizadas son:

- Variables climáticas, generadas normalmente a partir de la interpolación de datos de estaciones climatológicas y utilizando la elevación; dentro de esta variable se puede incluir el uso de un conjunto de datos espaciales llamado WORLDCLIM; que es un conjunto de datos espaciales resumidos en 19 variables bioclimáticas entre ellas precipitaciones y temperaturas máximas y mínimas diarias.
- Información sobre edafología, litología y geología, que representan la dependencia de la vegetación con el tipo de sustrato.
- Elevación y variables derivadas, tanto topográficas, pendiente, curvatura o rugosidad-, micro-climáticas, radiación solar potencial, como hidrológicas, red de drenaje o flujo potencial
- Variables obtenidas mediante teledetección, como índices de vegetación, clasificaciones de la cobertura y uso del suelo que han sido relativamente poco empleadas aunque tienen un importante potencial.
- 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 (Pahari & Murai, 1999).

Los modelos de Nicho ecológico describen la distribución geográfica real de una especie a través de algoritmos estadísticos y cartográficos; los principales insumos para realizar cualquier modelamiento son las variables dependientes (presencia y ausencia de la especie) y variables independientes (capas climáticas, geográficas, etc.)

2.4 Métodos para el Modelamiento de Nicho ecológico

En la actualidad existe una gran variedad de metodologías para realizar modelamiento de nicho ecológico mediante el uso algoritmos matemáticos y métodos de aprendizaje automático que requieren datos biológicos de la especie, además de la aplicación de variables medioambientales; principalmente los MNE están basados en tres técnicas de clasificación estadística que se muestran en la siguiente Tabla 1:

Tabla 1.
Técnicas para el Modelamiento de Nicho Ecológico

Técnicas	Descripción	Ejemplos
Discriminantes	Necesitan datos biológicos de la especie: presencia y ausencia para construir el estadístico.	<ul style="list-style-type: none"> • Los árboles de clasificación (CART) • Redes neuronales (AN) • Modelos lineales generalizados (GLMs) • Modelos aditivos generalizados GAMS • Modelos de máxima entropía (MAXENT) • Splines de Regresión Multivariada Adaptativa (MARS).
Descriptivas	Requieren datos biológicos de la especie: presencia únicamente	<ul style="list-style-type: none"> • BIOCLIM • BIOMAP • ENFA-Ecological Niche Factor Analysis-Biomapper
Mixtas	Utilizan tanto técnicas descriptivas como discriminantes, realizando sus propias pseudo-ausencias	<ul style="list-style-type: none"> • Desktop-GARP • OM-GARP

Fuente: (Soberón, 2005)

En algunos casos, los algoritmos han sido implementados de forma amigable para el usuario a través de paquetes de software que permiten describir la relación

entre el medioambiente y las especies, que generalmente están disponibles de forma gratuita para posteriormente integrarlo a un SIG para la obtención de productos cartográficos (Miller, 2010).

2.4.1 Máxima Entropía (Maxent)

El modelo de Máxima Entropía o Maxent se basa en conceptos de mecánica estadística y utiliza métodos probabilísticos de aprendizaje automático, el principio de máxima entropía establece que la mejor aproximación a una distribución no conocida es alguna cuya entropía sea máxima es decir la más dispersa sujeta a restricciones conocidas (información medio-ambiental y por la distribución disponible observada de las especies). Estas restricciones están función del valor esperado de la distribución, la cual es estimada por medio de un conjunto de observaciones de presencia de una determinada especie (Phillips & Dudík, 2004). Bajo este concepto se ha desarrollado un software específico para el modelamiento de nichos ecológicos bajo este mismo nombre llamado *Maxent* (Alvarado & Serrano, 2013).

Este método ha generado gran interés porque en comparación con otros métodos ha demostrado generar predicciones muy precisas. La técnica estadística que utiliza es de tipo discriminante, por tanto necesita tanto datos de presencia como ausencia para construir el clasificador o algoritmo. Sin embargo, *Maxent* genera sus propias ausencias (pseudo-ausencias), denominadas “background” o muestra de fondo; no es necesario introducir datos de ausencia (ver Figura 4). El programa está disponible en internet de forma gratuita en (<https://www.cs.princeton.edu/~schapire/maxent/>) (Soberón & Hensz, 2012).

Concretamente, esta técnica estima la probabilidad de presencia de la especie buscando la distribución de máxima entropía (lo más uniforme posible) sujeta a la condición de que el valor esperado de cada variable ambiental según esta distribución coincida con su media empírica (Peterson, 2003).

Maxent necesita de un algoritmo que le permita encontrar las distribuciones con mayor entropía, este algoritmo llamado “*Sequential update algorithm*”, utiliza iteraciones que dan distintos pesos a las variables utilizadas y las ajusta,

convergiendo de esta manera la distribución de probabilidades hasta la máxima entropía.

Maxent tiene propiedades que lo hacen muy robusto cuando se tienen cantidades limitadas de datos de entrenamiento (muestras pequeñas) además de ser usado en múltiples aplicaciones y en todas las escalas (Phillips & Anderson, 2006).

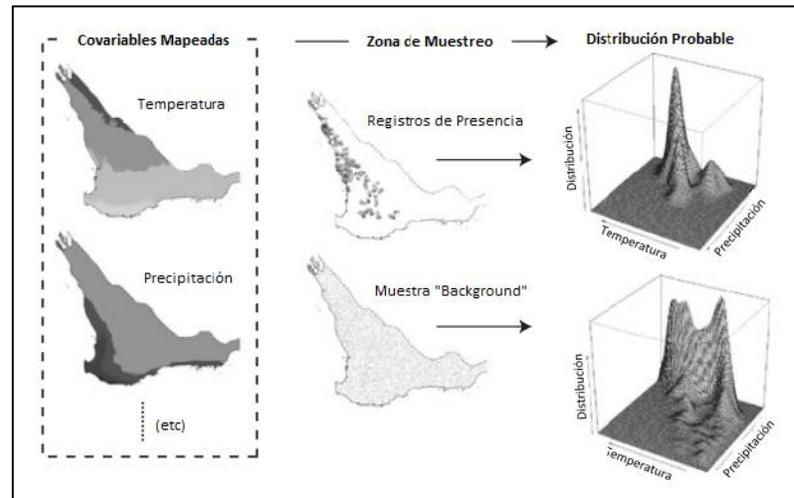


Figura 4. Esquema de Probabilidades en el modelo de Maxent

Fuente: (Phillips & Dudík, 2004)

Algunas de las ventajas de este método son que; se puede utilizar variables continuas y categóricas en muestras pequeñas, incorporando interacciones entre diferentes variables, y que los datos de salida son representados como predicciones continuas permitiendo distinciones entre áreas (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) (Phillips & Dudík, 2004).

2.4.1.1 Descripción matemática de Maxent

Maxent pertenece a la familia de los clasificadores exponenciales o log-lineales, que extrae un conjunto de características (features) de la entrada, las combina de modo lineal y utiliza el resultado como un exponente. Dada una entrada X con features f_i ponderadas mediante w_i , la probabilidad de asignar a X la clase C es:

$$P_{(C|X)} = \frac{1}{Z} \exp\left(\sum_i w_i f_i\right) \quad (1)$$

Donde Z es un factor normalizador para hacer que todo sume 1. Si el resultado pertenece a un conjunto discreto se habla de clasificación, si es un conjunto real se habla de regresión (Shannon, 1948).

Este modelo utiliza un conjunto de celdas discretas definida como X que representa una región geográfica finita, formada por un subconjunto de puntos $x_1 \dots x_m$ los cuales representan las localidades donde fue registrada la especie; suponiendo que los registros de $x_1 \dots x_m$ fueron obtenidos independientemente de X , de acuerdo a una distribución de probabilidad desconocida llamada π , el objetivo es construir una distribución $\tilde{\pi}$ que se aproxime a π . El modelo parte de la premisa de que la distribución π coincide con el concepto biológico de la distribución potencial de las especies. Sin embargo, en la realidad ciertas localidades son mejor evaluadas que otras, por lo que la distribución de π asume un sesgo en el esfuerzo de colecta y puede asignarle un mayor peso a áreas y condiciones ambientales mejor evaluadas (Phillips & Dudík, 2004).

En la construcción de $\tilde{\pi}$ se hace uso de una serie de características $f_1 \dots f_n$ donde: $f_j: X \rightarrow \mathbb{R}$

Estas características consisten en un renglón de variables, donde f denota el vector de las n variables y \mathbb{R} todos los valores que toman las f características en X , mediante una distribución empírica (Phillips & Dudík, 2004):

$$\tilde{\pi}(x) = \frac{|\{1 \leq i \leq m : x_i = x\}|}{m} \quad (2)$$

En general, $\tilde{\pi}$ puede ser distante de π , por otro lado, para una función de f , se espera que $\tilde{\pi}(f)$, el promedio empírico de f , sea muy cercano al esperado que busca una aproximación de $\tilde{\pi}$, en la cual el esperado de f_j sea igual (o al menos cercano) de $\tilde{\pi}(f)$ para cada f_j .

El principio de máxima entropía sugiere que de entre todas las posibles distribuciones, que satisfacen estas limitantes, se elige a aquella con la mayor entropía, es decir la más cercana al uniforme.

La entropía de la distribución p en X es definida:

$$H(p) = - \sum_{x \in X} p(x) \ln p(x) \quad (3)$$

El modelo parte de la idea de estimar π por la distribución de $\hat{\pi}$ en función de la máxima entropía sujeta a la condición de que:

$$\hat{\pi}[f_j] = \tilde{\pi}[f_j] \quad (4)$$

Todas las características de f_j utilizan un algoritmo secuencial actualizado que modifica el peso de f_j por ocasión para encontrar la distribución de la máxima entropía, este algoritmo es fácilmente aplicado cuando el número de características es muy grande. Asume, sin pérdida de generalidad, que cada característica f_j está limitada entre 0,0 - 1,0. En cada secuencia de estimación, se elige la característica f_j a actualizar para la cual $RE(\tilde{\pi}[f_j] q | \lambda f_j)$ es maximizada, donde λ es el actual vector (Phillips & Anderson, 2006).

Se actualiza $f_j \leftarrow f_j + \alpha$ donde:

$$= \ln \left(\frac{\tilde{\pi}[f_j] (1 - \alpha \lambda [f_j])}{(1 - \pi[f_j]) q \lambda [f_j]} \right) \quad (5)$$

La distribución de salida $\hat{\pi}$ es obtenida del cálculo de los pesos, es decir los $q\lambda$, el algoritmo trabaja alternando un peso de f_j por ocasión hasta que maximiza la verosimilitud o encuentra una aproximación, el procedimiento garantiza la convergencia al óptimo de la distribución de la máxima entropía (Phillips & Dudík, 2004).

No se espera que $\tilde{\pi}[f_j]$ sea igual a $\pi[f_j]$, pero al menos cercana, para resolverlo el modelo realiza:

$$|\hat{\pi}[f_j] - \tilde{\pi}[f_j]| \leq \epsilon_j \quad (6)$$

donde ϵ_j es un estimado de que tan cercano se encuentra de $\pi[f_j]$.

Esto se llama las l-regularizaciones, las mismas que fuerzan al modelo de Maxent a enfocarse sobre las características más importantes, y las l-regularizaciones tienden a producir modelos con pocos valores de cero para ϵ_j (Phillips & Dudík, 2004).

2.4.2 Lógica Difusa o FUZZY

Esta lógica es de tipo determinística que utiliza expresiones que no son totalmente ciertas ni totalmente falsas, es decir, es una lógica aplicada a conceptos que pueden tomar un valor indeterminado de veracidad dentro de un conjunto de valores cuyos extremos son la verdad absoluta o la falsedad absoluta. Este método está basado en la teoría de conjuntos que posibilita imitar el comportamiento de la lógica humana utilizando elementos relacionados con la probabilidad, es decir que se produzca o no una determinada condición, además puede definir qué elementos no están correctamente caracterizados por muchos observadores, generando incertidumbre al definir una clasificación específica (Gossweiler, 2002).

La lógica difusa fue investigada por primera vez alrededor de mediados de los años sesenta por el ingeniero Lotfy A. Zadeh en la Universidad de Berkeley (California). En un principio este ingeniero no denominó a esta lógica como lógica borrosa sino que la llamó principio de incompatibilidad. La lógica difusa permite representar el conocimiento común, que es mayoritariamente del tipo lingüístico cualitativo y no necesariamente cuantitativo, en un lenguaje matemático a través de la teoría de conjuntos difusos y funciones características asociadas a ellos.

Este permite trabajar a la vez con datos numéricos y términos lingüísticos, estos son inherentemente menos precisos que los datos numéricos pero en muchas ocasiones aportan una información más útil para el razonamiento humano (López, 2001).

El aspecto central de los sistemas basados en la teoría de la lógica difusa es que, a diferencia de los que se basan en la lógica clásica, tienen la capacidad de reproducir aceptablemente los modos usuales del razonamiento, considerando que la certeza de

una proposición es una cuestión de grado. Más formalmente se puede decir que si la lógica es la ciencia de los principios formales y normativos del razonamiento, la lógica difusa o borrosa se refiere a los principios formales del razonamiento aproximado, considerando el razonamiento preciso (lógica clásica) como caso límite. Así pues, las características más atractivas de la lógica difusa son su flexibilidad, su tolerancia con la imprecisión, su capacidad para modelar problemas no-lineales, y su base en el lenguaje natural (Ortíz, 2002).

Los subconjuntos difusos (o partes borrosas de un conjunto) fueron inventados para modelar la representación humana de los conocimientos (por ejemplo para medir nuestra ignorancia o una imprecisión objetiva) y mejorar así los sistemas de decisión, de ayuda a la decisión, y de inteligencia artificial; 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).

En la lógica clásica una proposición sólo admite dos valores: verdadero o falso. Por ello se dice que la lógica usual es bivalente o binaria. Existen otras lógicas que admiten además un tercer valor posible (lógica tri-valuada) e incluso múltiple valores de verdad (lógica multi-valuada) (Padilla, 2006).

2.4.2.1 Descripción matemática de Lógica Difusa

La lógica aristotélica sirve para explicar ciertos fenómenos y problemas, aunque la gran mayoría de ellos enmarcados en el mundo teórico de la matemática. Por el contrario, la lógica difusa puede usarse para explicar el mundo real, puesto que sigue el comportamiento humano de razonar, sacando conclusiones a partir de hechos observados.

La lógica multi-valuada incluye sistemas lógicos que admiten varios valores de verdad posibles. La lógica difusa permite representar de forma matemática conceptos

o conjuntos imprecisos. Si P es una proposición, se le puede asociar un número $v(P)$ en el intervalo $[0,1]$ tal que:

Si $v(P) = 0$, entonces P es falso.

Si $v(P) = 1$, entonces P es verdadero.

La veracidad de P aumenta con $v(P)$.

Según la clasificación difusa, borrosa o Fuzzy evalúa la posibilidad de que cada elemento o pixel pertenezca a un determinado sistema o fenómeno (Guo et. Al., 2004), (ver Figura 5).

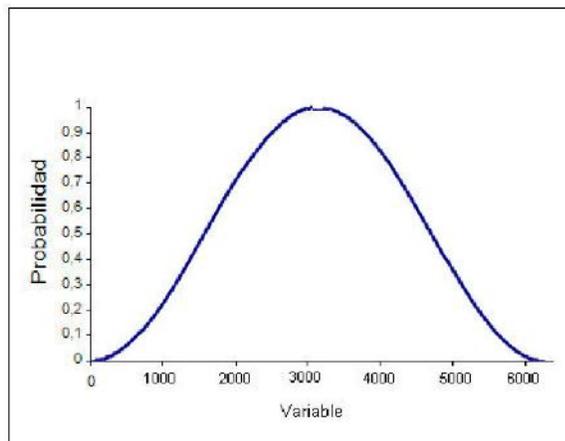


Figura 5. Probabilidad tomando una variable

Fuente: (Padilla, 2006)

Las funciones sinusoidales o cosinusoidales se ajustan bastante bien a este tipo de representaciones, es por esto que utilizaremos ecuaciones en función de seno de alfa y coseno de alfa, (Ruiz, et. Al. 2007) donde alfa estará en función del fenómeno o variable que se desea modelar o representar.

Es así que la probabilidad se define de la siguiente forma:

$$P = f(\text{sen } \alpha) \quad (7)$$

Donde:

P = probabilidad de ocurrencia del evento o fenómeno

α = variable en función del evento o fenómeno

Mínima $P = 0$ o 0%

Máxima $P = 1$ o 100%

Para obtener una curva donde el rango sea de $[0,1]$, usamos la función cuadrada del seno.

$$P = \text{sen}^2 \alpha \quad (8)$$

Puntos importantes en la función:

Estas funciones están controladas por cuatro puntos situados de la siguiente forma (ver Figura 6):

Punto A: Punto de valor mínimo en la variable a modelar

Punto B y C: Rango de valores donde se encuentran la mayor cantidad de datos de la variable

Punto D: Punto de valor máximo en la variable.

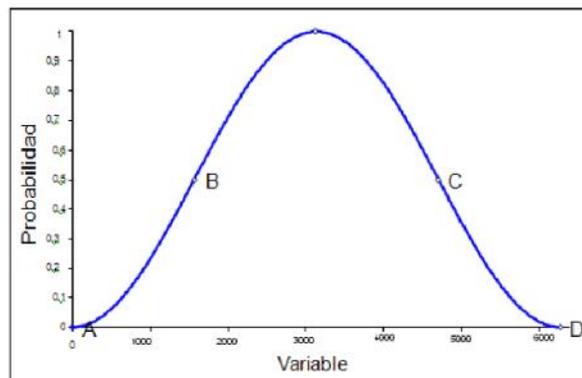


Figura 6. Puntos importantes del Modelo

Fuente: (Padilla, 2006)

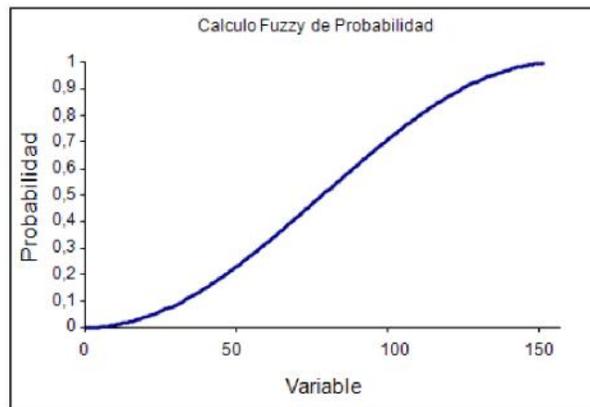


Figura 7. Probabilidad Creciente

Fuente: (Padilla, 2006)

Puntos de inflexión de la curva: estos puntos pueden ser tomados como un caso particular para situar los puntos B y C de las variables. Se tienen tres casos de los que se va a tomar como ejemplo el de la curva simétrica (ver Figura 7).

$$P = \text{sen}^2 \alpha$$

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

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

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

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

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

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

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

$$\alpha_1 = \frac{1}{4}\pi$$

$$\alpha_2 = \frac{3}{4}\pi$$

Finalmente:

$$P_1 (A; 0)$$

$$P_2 \left(B; \frac{1}{4}\pi \right)$$

$$P_3 \left(C; \frac{3}{4}\pi \right)$$

$$P_4 \left(D; \pi \right)$$

$$P_5 \left(\frac{B+C}{2}; \frac{1}{2}\pi \right), \text{ donde } P_5 \text{ corresponde al punto medio de la curva.}$$

Zonas Bajo la Curva

De la ecuación general de la recta se tiene:

$$y - y_1 = \frac{(y_2 - y_1)}{(x_2 - x_1)} (x - x_1) \quad (9)$$

El eje x será la variable y el eje y será.

Zona I (F₁, F₂)

$$P_1 (A, 0)$$

$$P_2 \left(B; \frac{1}{4}\pi \right)$$

$$y - y_1 = \frac{(y_2 - y_1)}{(x_2 - x_1)} (x - x_1)$$

$$- 0 = \left(\frac{\frac{1}{4}\pi - 0}{B - A} \right) (V - A)$$

$$= \left(\frac{1}{4}\pi \right) \left(\frac{V - A}{B - A} \right) \quad (10)$$

Zona Ia (P₂, P₅)

$$P_2 \left(B; \frac{1}{4}\pi \right)$$

$$P_5 \left(\frac{B+C}{2}, \frac{1}{2}\pi \right)$$

$$y - y_1 = \frac{(y_2 - y_1)}{(x_2 - x_1)} (x - x_1)$$

$$-\frac{1}{4}\pi = \left(\frac{\frac{1}{2}\pi - \frac{1}{4}\pi}{\frac{B+C}{2}} \right) (V - B)$$

$$= \frac{1}{4}\pi + \left(\frac{\frac{1}{2}\pi(V - B)}{C - B} \right) \quad (11)$$

Zona II (P_5, P_3)

$$P_5 \left(\frac{B+C}{2}, \frac{1}{2}\pi \right)$$

$$P_3 \left(c, \frac{3}{4}\pi \right)$$

$$y - y_1 = \frac{(y_2 - y_1)}{(x_2 - x_1)} (x - x_1)$$

$$-\frac{1}{2}\pi = \left(\frac{\frac{3}{4}\pi - \frac{1}{2}\pi}{c - \frac{B+C}{2}} \right) \left(V - \frac{B+C}{2} \right)$$

$$= \frac{1}{2}\pi + \left(\frac{\frac{1}{4}\pi(2V - B - C)}{C - B} \right) \quad (12)$$

Zona III (P_3, F_4)

$$P_3 \left(c, \frac{3}{4}\pi \right)$$

$P_2(\varepsilon, \pi)$

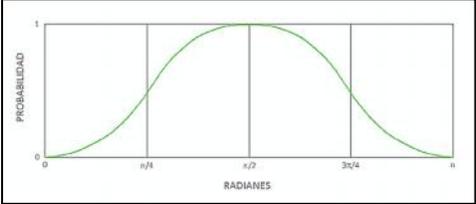
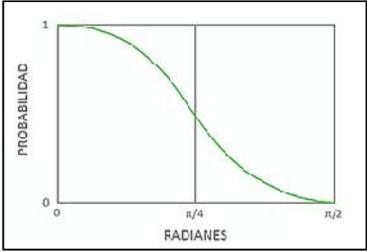
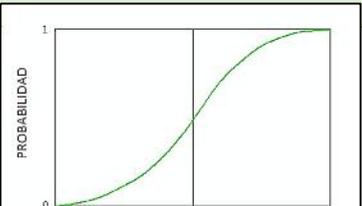
$$y - y_1 = \frac{(y_2 - y_1)}{(x_2 - x_1)} (x - x_1)$$

$$-\frac{3}{4}\pi = \left(\frac{\pi - \frac{3}{4}\pi}{D - C} \right) (V - C)$$

$$= \frac{3}{4}\pi + \left(\frac{\frac{1}{4}\pi(V - C)}{D - C} \right) \quad (13)$$

El análisis de la probabilidad de la ocurrencia de un evento se realiza mediante la aplicación de funciones sinusoidales o cosinusoidales a las variables dependientes que se ajustan a este tipo de representaciones donde los puntos de inflexión de la curva determinan un caso en particular como se muestra en la Tabla 2.

Tabla 2.
Casos Lógica Fuzzy

Casos Fuzzy	Descripción	Rango
<p>Primer Caso</p> 	<p>La probabilidad de ocurrencia crece desde 0 hasta un punto máximo 1, y posteriormente decrece nuevamente a 0.</p>	<p>Grados: 0° a 180°</p> <p>Radianes: 0 a π</p> <p><u>Función</u> Seno</p>
<p>Segundo Caso</p> 	<p>La probabilidad de ocurrencia, es inversamente proporcional al evento analizado</p>	<p>Grados: 0° a 90°</p> <p>Radianes: 0 a $\pi/2$</p> <p><u>Función</u> Coseno</p>
<p>Tercer Caso</p> 		

La probabilidad de ocurrencia, es directamente proporcional al evento analizado.	Grados: 0° a 90°
	Radianes: 0 a $\pi/2$
	<u>Función</u>
	Seno

Fuente: (Padilla, 2006)

Existen diversas aplicaciones en las que se puede utilizar este método, en función a la ocurrencia de un determinado fenómeno en el cual se involucren más de dos variables; la probabilidad de localizar o modelar este elemento corresponde al producto de todas esas variables como plantea Padilla (2006) a través de la siguiente fórmula (14):

$$P_t = \frac{PV_1 + PV_2 + PV_3 + \dots + PV_n}{n} \quad (14)$$

Donde:

P_t : Es la probabilidad total de obtención de un fenómeno (clasificación de una imagen)

$PV_1 + \dots + PV_n$: Son las probabilidades individuales, obtenidas mediante la metodología de operadores difusos propuesta, de cada variable.

n: Número total de variables.

2.4.3 Regresión Logística

Los modelos de regresión logística son modelos estadísticos de uso tanto explicativo como predictivo, que permiten estudiar mediante un análisis bivariado o multivariado si una variable binomial depende, o no, de otra u otras variables (no necesariamente binomiales), mediante el uso de su transformación logarítmica (*logit*) como una función lineal que garantiza que las predicciones estén dentro de un rango de valores posibles para la variable respuesta, que no son más que un conjunto de variables predictivas que pueden ser cuantitativas (que se denominan covariables) y se considera que las respuestas provienen de una secuencia de ensayos, donde cada

uno de ellos tiene su propia probabilidad de éxito que depende de los valores de las variables predictoras (ver Figura 8); por ejemplo, entre 0 y 1 para probabilidades de presencia (Guisan et al, 2002).

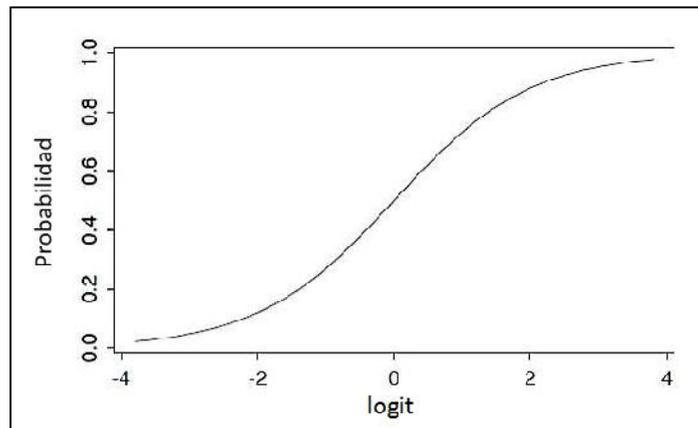


Figura 8. Función logística

Fuente: (Villar, 2014)

El objetivo primordial que resuelve esta técnica es el de modelar cómo influye en la probabilidad de aparición de un suceso, habitualmente dicotómico, la presencia o no de diversos factores y el valor o nivel de los mismos. También puede ser usada para estimar la probabilidad de aparición de cada una de las posibilidades de un suceso con más de dos categorías (politómico), debido a que 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 (Villar, 2014).

Por sus características, los modelos de regresión logística permiten dos finalidades:

- Cuantificar la importancia de la relación existente entre cada una de las covariables y la variable dependiente, lo que lleva implícito también clarificar la existencia de interacción y confusión entre covariables respecto a la variable dependiente.
- Clasificar individuos dentro de las categorías (presente/ausente) de la variable dependiente, según la probabilidad que tenga de pertenecer a una de ellas dada la presencia de determinadas covariables. No cabe duda que la regresión logística es una de las herramientas estadísticas con mejor capacidad para el

análisis de datos en investigación clínica y epidemiología, de ahí su amplia utilización (Villar, 2014).

2.4.3.1 Descripción matemática de Regresión Logística

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* $(1; \pi_i)$, 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 $X_i^T \beta$ estará en el intervalo $-\infty$ a ∞ mientras que π_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 (Guisan et al, 2002).

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

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = X_i^T \beta \quad (15)$$

Se deben considerar ciertos supuestos que se manejan dentro del modelo logístico como:

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

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = X_i^T \beta$$

Resolviendo para π_i se obtiene:

$$\pi_i = \frac{1}{1 + e^{-X_i^T \beta}}$$

$$X_i^T \beta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \dots + \beta_P X_{Pi} \quad (16)$$

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

2.4.4 Splines de Regresión Multivariada Adaptativa (MARS)

Los Splines de Regresión Adaptativa Multivariada (MARS) fueron desarrollados por Friedman en el año de 1991 con el fin de ajustar modelos con relaciones aditivas y posibles interacciones, mediante el uso de un algoritmo basado en las ideas de particionamiento recursivo y regresión multietapa que emplea funciones tipo *spline* (Base de Suavizamiento) para lograr la aproximación a una función de regresión arbitraria. El procedimiento consiste en particionar el dominio de definición de la función en diferentes regiones, ajustando en cada una de ellas una función tipo *spline*, (Friedman, 1991).

MARS es un método estadístico de regresión no lineal que no hace ninguna suposición sobre la relación funcional entre las variables respuesta e independientes así como los supuestos que requiere la Regresión Lineal Múltiple, por tal motivo, este modelo construye un conjunto de funciones base que están definidas en pares a ambos lados de un nudo (Villar, 2014).

Un nudo es un valor de una variable que define un punto de inflexión a lo largo del rango del predictor, por tanto, dos nudos definen una sección. Los coeficientes se estiman de tal forma que definen la pendiente de cada sección, muchos nudos potenciales se identifican automáticamente (Villar, 2014).

Un desafío práctico de usar MARS es que comúnmente se utiliza software basados en ajuste por mínimos cuadrados, que son apropiados para datos donde los errores tienen distribución normal, sin embargo, este enfoque no es apropiado para datos binomiales (es decir: presencia/ausencia). Otra dificultad técnica, es que la mayoría del software disponible solo permite el uso de variables predictoras continuas o cuantitativas (no categóricas). Así, algunos estudios no han considerado ninguna variable categórica (Franklin, 2009).

La aplicación de los modelos MARS en el modelamiento de Nichos Ecológicos es relativamente reciente, y derivado de ello son pocos los estudios donde han explorado esta herramienta estadística para modelar distribuciones de especies, sin embargo Franklin (2009) detalla ciertas ventajas que ofrece MARS, como por ejemplo computacionalmente son más rápidos que los Modelos Aditivos Generalizados (GAM) y por tanto, son modelos prácticos para trabajar con grandes bases de datos, donde se tengan muchas observaciones y un gran número de variables predictoras. La predicción espacial resulta mucho más simple cuando se utiliza el enfoque MARS en comparación con los árboles de clasificación (Espíndola, 2013).

2.4.4.1 Descripción matemática de MARS

Los MARS utilizan expansiones de funciones de base lineal por partes de la forma $(x - t)_+$ y $(t - x)_+$. El signo “+” representa la parte positiva, por lo que (ver Figura 9):

$$(x - t)_+ = \begin{cases} x - t, & \text{si } x > t \\ 0, & \text{de otro modo} \end{cases}$$

y

$$(t - x)_+ = \begin{cases} t - x, & \text{si } x < t \\ 0, & \text{de otro modo} \end{cases}$$

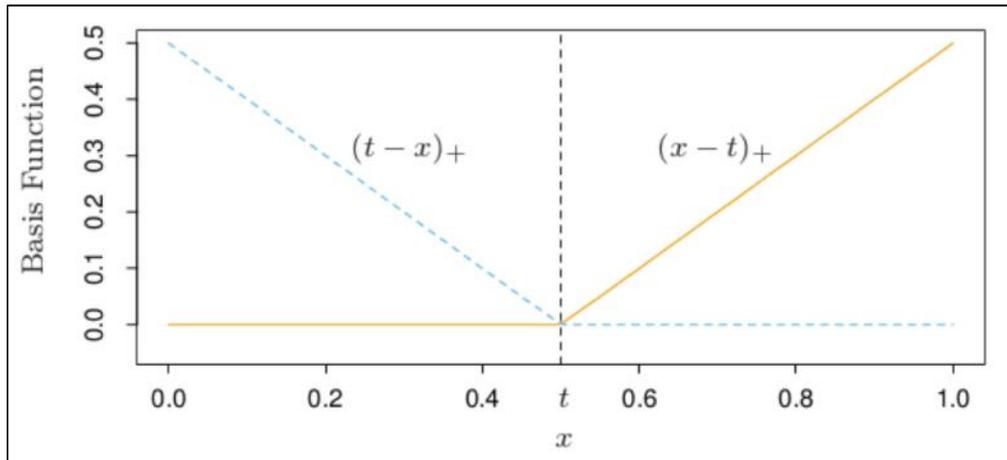


Figura 9. Funciones base usados por MARS

Fuente: (Villar, 2014)

Cada función es lineal por partes con un nudo en el valor t , lo que también se conoce como splines lineales. La idea es formar pares para cada valor observado x_{ij} de dicha entrada. Por lo tanto, la colección de funciones base:

$$C = \left\{ (X_j - t)_+, (t - X_j)_+ \right\}_{t \in \{x_{1j}, x_{2j}, \dots, x_{Nj}\}} \quad \text{con } j = 1, 2, \dots, p \quad (17)$$

La estrategia para construir el modelo se asemeja a una regresión lineal por partes, pero en lugar de utilizar las entradas originales, se permite utilizar las funciones de la serie C y sus productos. Así el modelo tiene la forma

$$f(X) = \beta_0 + \sum_{m=1}^M \beta_m h_m(X) \quad (18)$$

Donde cada $h_m(X)$ es una función en C , o un producto de dos o más de tales funciones. Los coeficientes β_m se estiman minimizando la suma de cuadrados de los residuales, como en la regresión lineal estándar. Sin embargo, el verdadero arte radica en la construcción de las funciones h_m . Se puede comenzar con la función constante $h_m(X) = 1$ en el modelo y por tanto, todas las funciones en el conjunto C son funciones candidatas. En cada etapa se considera como un nuevo par de función

base F_{η} a todos los productos de una función hm en el modelo del conjunto M con uno de los pares reflejados en C (Villar, 2014).

Este modelo generalmente se sobre ajusta a los datos y es necesario aplicar el procedimiento de selección de variables *backward*. El término cuya extracción cause el menor incremento en el error cuadrático de los residuales se elimina del modelo en cada etapa, produciendo un modelo estimado \hat{f}_{λ} . Para estimar el valor óptimo de λ se utiliza el criterio de validación cruzada generalizada (Villar, 2014). Este criterio se define como

$$GCV = \frac{\sum_{i=1}^N (y_i - \hat{f}_{\lambda}(x_i))^2}{(1 - M(\lambda)/N)^2} \quad (19)$$

El valor $M(\lambda)$ es el número efectivo de parámetros en el modelo: esto representa tanto el número de términos en el modelo, más el número de parámetros utilizados en la selección de las posiciones óptimas de los nudos.

Los métodos para el modelamiento del nicho ecológico se dividen en tres categorías en función a los insumos que requieren y son: discriminantes, descriptivos y mixtos; tienen como finalidad, encontrar probabilísticamente zonas idóneas en las que exista presencia de la especie, en base a las variables dependientes e independientes que forman parte de cualquier modelo, elaborado a través de los métodos de Maxent, Regresión Logística, MARS o Lógica Difusa.

2.5 Validación de los modelos de Nicho ecológico

La validación de los modelos de nicho ecológico se enfoca en la evaluación del rendimiento predictivo de los mismos; consiste principalmente en distinguir entre áreas adecuadas para la especie de las inadecuadas, de esta manera se establece un umbral de decisión que contrarresta las predicciones del modelo con respecto a los datos considerados independientes en función de una cierta hipótesis nula, además permite realizar una comparación entre modelos (Pearson, 2008).

En los Modelos de Nicho Ecológico, la validación resulta más adecuada si se utilizan datos nuevos e independientes, es decir, datos que no fueron utilizados para

estimar los parámetros del modelo. Sin embargo, en la mayoría de las veces, no es posible ni factible coleccionar nuevos datos. En estos casos, un enfoque bastante frecuente es dividir los datos en dos partes, una parte se utiliza para entrenamiento, y la otra se usa para validar las predicciones (Miller, 2010).

La mayoría de los modelos de distribución de las especies presentan errores de predicción, por tal motivo se realiza la validación de los mismos, la cual consiste en probar si los ajustes hechos al modelo concuerdan con la realidad, enfocados en medidas de desempeño. La mayor parte de las medidas modernas están basadas en la así llamada matriz de confusión que resume el desempeño del modelo en términos de errores y aciertos (Pearson, 2008).

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 (ver Figura 10):

	Presencias Reales	Ausencias Reales
Presencias predichas	<i>a</i>	<i>b</i>
Ausencias Predichas	<i>c</i>	<i>d</i>

Figura 10. Matriz de confusión

Fuente: (Pearson, 2008)

- **Positivas verdaderas (a):** es decir las predicciones de presencia de la especie donde la especie es realmente presente
- **Positivas falsas (b):** que corresponde a los puntos donde la presencia fue predicha pero en realidad la especie no está presente
- **Negativas falsas (c):** donde se predijo la ausencia pero en realidad la especie está presente
- **Negativas verdaderas (d):** cuando se predijo ausencia y que realmente la especie es ausente.

Cada elemento de la matriz de confusión puede ser visualizado en el espacio geográfico como se ilustra en la Figura 11 donde se representa la región ocupada en la realidad (la "verdad") con el color azul, y la predicción, con el color amarillo, los puntos de prueba independientes son usados para verificar el modelo y pueden caer en cualquiera de las cuatro regiones, las estrellas están bien clasificadas y los círculos son errores, finalmente los puntos oscuros son errores de omisión, y los círculos abiertos son errores de comisión (Phillips & Dudík, 2004).

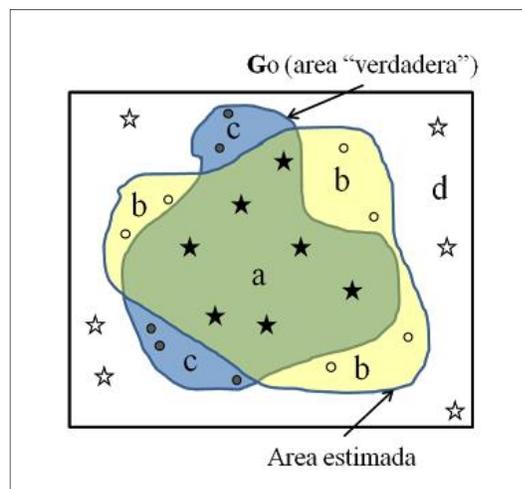


Figura 11. Matriz de confusión representada en el Espacio Geográfico

Fuente: (Pearson, 2008)

Errores de omisión: Clasificar equivocadamente un dato de gran fuerza, que es una ocurrencia, tal vez respaldada por un ejemplar físico, es decir predice la no presencia de la especie donde realmente está (falso negativo) (Villar, 2014).

Errores de comisión: Es mucho más importante, pues no predice lugares de presencia que pueden ser de importancia crucial para la supervivencia de la población, en donde la especie puede estar ausente por una variedad de razones, o simplemente no haber sido detectada, es decir predice la presencia de la especie donde no está (falso positivo) (Pearson, 2008).

2.5.1 Curva ROC (Receiver operating characteristic analysis)

La curva ROC (Receiver Operating Characteristic analysis), es un estadístico que representa de manera gráfica la capacidad discriminativa de cualquier modelo para todos sus posibles puntos de corte, necesita que los datos a ser evaluados sean de presencia/ausencia para definir de esta manera el umbral o criterios necesarios para predicción de la especie (Stockwell & Peterson, 2002). La curva ROC (ver Figura 12) se obtiene al relacionar la sensibilidad que como se dijo anteriormente es la fracción de verdaderos positivos (Eje y), con la “1- especificidad” que es la fracción de falsos positivos (Eje x), por facilidad de cálculo se utiliza la expresión “1- especificidad” para que la sensibilidad y la especificidad varíen en la misma dirección cuando se defina el umbral (Pearson, 2008).

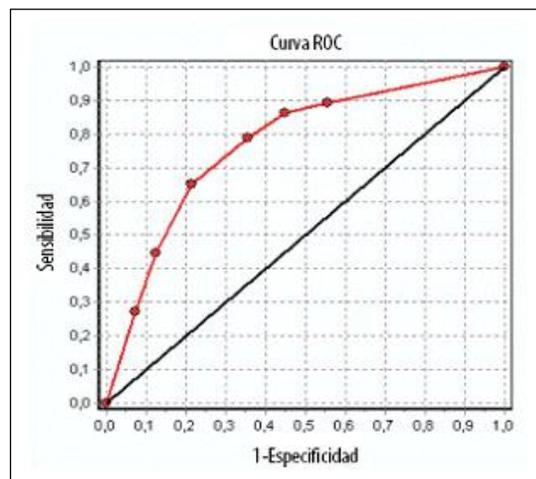


Figura 12. Curva ROC (Receiver operating characteristic analysis)

Fuente: (Pearson, 2008)

El estadístico derivado es el área bajo la curva ROC, o AUC (siglas del nombre inglés "Area Under Curve"), que proporciona una medida completa de la capacidad predictiva además de evaluar la bondad del ajuste en los Modelos de Nicho Ecológicos, la define Pearson (2008) como:

$$AUC = \int_0^1 ROC(p) dp \quad (20)$$

El AUC se calcula sumando el área debajo de la curva ROC y toma valores entre 0 a 1, donde valores inferiores a 0,5 indican que el modelo es realmente malo,

ya que clasifica erróneamente más casos que el azar; valores de AUC de 0,5 a 0,7 se considera como bajo rendimiento del modelo; valores entre 0,7 a 0,9 presume un rendimiento moderado del modelo, y valores mayores a 0,9 estiman un nivel alto del modelo, es decir, indica que todos los casos se han clasificado correctamente (Muñoz & Felicísimo, 2012).

Los valores de AUC no se ven afectados por cambios en la prevalencia de la especie y por tanto es una estadística confiable en la comparación de modelos; algunos estudios han mostrado que el AUC no disminuye al aumentar la prevalencia de la especie (Franklin, 2009). Dentro de las ventajas que proporciona el cálculo del AUC está la posibilidad de comparar varios métodos, sea cual sea el tipo de valores de salida debido a que sólo necesita que las distribuciones de estos valores (Graham, 2012).

2.5.2 Limitaciones de los modelos de Nicho ecológico

Debido a que la naturaleza es compleja y heterogénea además de que sufre constantes cambios por factores antropogénicos, es considerable no esperar que los modelos nos reflejen con precisión los mecanismos inherentes a un proceso espacio-temporal tan complejo como es la distribución de las especies. A continuación se describe brevemente las limitaciones más comúnmente destacadas por varios autores (Peterson, 2003).

- Ausencia de la componente espacial en el proceso, es decir encontrar zonas potenciales separadas por amplias barreras climáticas cuando, en realidad, la especie está presente solamente en una de esas zonas y la otra podría considerarse aislada e inaccesible.
- Ausencia de la componente temporal, es decir cambios geográficos, geológicos climáticos en el área de estudio dentro de una transición entre un periodo temporal y otro.
- Dificultad de hacer modelos fiables para especies que se adaptan en distintas sub-poblaciones morfológicamente indistintas (Guisan & Thuiller, 2005).

Algunas de estas limitaciones pueden evitarse al restringir los resultados por regiones biogeográficas e introducir parámetros auxiliares como distancia de

dispersión, fragmentación del terreno o poder reproductivo o también se puede elaborar modelos basados en datos fisiológicos de las especies y de la interacción de éstas con el medio, una estrategia novedosa y con gran futuro en la modelización (Ghaham, 2012).

La validación de los modelos de nicho ecológico se realiza a través de varios estadísticos tales como la matriz de confusión (errores de comisión y omisión) y la curva ROC que determina el área bajo la curva (AUC), evaluando de esta manera el rendimiento predictivo de los métodos aplicados, lo que permite realizar una comparación entre modelos y definir cuáles son las limitación que existen entre ellos.

2.6 Evaluación multicriterio (EMC)

La evaluación multicriterio es un conjunto de técnicas que facilitan procesos de toma de decisiones, mediante el análisis de factores y restricciones para obtener alternativas; estas técnicas se basan en la ponderación y comparación de variables previamente inventariadas y clasificadas que influenciarán sea de manera positiva o negativa sobre la actividad objeto de decisión, cuyo resultado final deberá ser evaluado y jerarquizado (Gómez & Barredo, 2005).

La evaluación multicriterio y el análisis espacial dentro del ámbito territorial deberán ser desarrolladas conjuntamente ya que aspectos ambientales, económicos, sociales e incluso políticos juegan un papel importante que generan posibles escenarios en conflicto donde la toma de decisiones debe ser abierta y flexible de acuerdo a una visión de la realidad objetiva (Moreno, 2001).

Los componentes de la EMC tienen como objetivo fundamental definir los lineamientos a los que se van a regir los parámetros evaluados, mediante la normalización, ponderación y jerarquización de los resultados obtenidos en función a la regla de decisión aplicada (Moreno, 2001).

2.6.1 Normalización de Variables

Mediante la normalización de datos se establecen valores que pueden ser comparados y operables entre sí para ser medidos dentro de una misma escala comparable en tipo, rango de extensión, unidad de medidas, eventual posición del cero, dispersión, etc. ya que dentro de una matriz es importante que los datos sean homogéneos por lo cual es necesario aplicar operadores o metodologías específicas (Gómez & Barredo, 2005).

La normalización permite eliminar problemas de cálculo debido a la diferencia de escalas que pueden presentarse dentro de una matriz. Sin embargo no la describe como un proceso necesario, pero para muchos de los métodos compensatorios suele ser esencial, además permite comparar atributos y como resultados se obtienen valores sin unidad de dimensión. (Molero & Rodriguez, 2007). La siguiente ecuación establecida por (Barba & Romero, 1987) permite normalizar los valores obtenidos para cada variable:

$$X_n = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad (21)$$

Donde X_n es la variable normalizada, X_i es la variable, X_{min} y X_{max} representan los valores mínimos y máximos respectivamente

2.6.2 Método de Análisis Jerárquico AHP (Thomas Saaty)

El método de análisis jerárquico propuesto por Thomas Saaty plantea un procedimiento de comparación por pares de los criterios que parte de una matriz cuadrada en la cual el número de filas y columnas está definido por el número de criterios a ponderar, de cada uno de ellos con los demás, y posteriormente se establece el eigenvalor principal, el cual constituye los pesos (w_j) que a su vez proporciona una medida cuantitativa de la consistencia de los juicios de valor entre pares de factores (Saaty, 1980). La metodología resultante de Saaty debe ser: simple en su construcción, adaptable a las decisiones individuales y en grupo, en consonancia con nuestros pensamientos, valores e intuiciones, orientada a la búsqueda del consenso y que no requiera una especialización suprema para su aplicación (Marquez, 1999).

Saaty (1980), propone la utilización de una escala fundamental para establecer los valores (juicios) considerando un rango de valores entre 1/9 y 9 (ver Figura 13) evita el problema que se plantea cuando se realizan comparaciones relativas, o si se prefiere razones, entre elementos con valores que van de cero a infinito como en las fórmulas matemáticas habituales. Al utilizar en el proceso de cálculo las potencias de los juicios, los valores obtenidos tienden rápidamente a tomar valores fuera del rango de nuestra habilidad de interpretación de esos números.

VALOR	DEFINICION	COMENTARIOS
1	Igual Importancia	El criterio A es igual de importante que el criterio B
3	Importancia Moderada	La experiencia y el juicio favorecen ligeramente al criterio A sobre el B
5	Importancia grande	La experiencia y el juicio favorecen ligeramente al criterio A sobre el B
7	Importancia muy grande	El criterio A es mucho más importante que el B
9	Importancia extrema	La mayor importancia del criterio A sobre el B esta fuera de toda duda
2,4,6 y 8	Valores intermedios entre los anteriores, cuando es necesario matizar	
Recíprocos de lo anterior	Si el criterio A es de importancia grande frente al criterio B las notaciones serían las siguientes. Criterio A frente a criterio B 5/1 Criterio B frente a criterio A 1/5	

Figura 13. Escala fundamental de prioridades de Saaty

Fuente: (Saaty, 1980)

Una vez obtenidos los pesos para cada uno de los criterios el siguiente paso consiste en calcular el eigenvalor principal de dicha matriz el cual nos va a proporcionar una medida cuantitativa de los juicios de valor asignados a los elementos de la matriz de comparación por pares. Esta medida se consigue mediante la expresión planteada por (Saaty, 1980):

$$C.I = \frac{\lambda_{max} - n}{n - 1} \quad (22)$$

Donde

C.I: es el índice de consistencia

λ_{max} : representa al eigenvalor principal de la matriz de comparación por pares

n: es el número de filas o de columnas de la matriz

La evaluación multicriterio es un conjunto de técnicas que facilitan procesos de toma de decisiones las cuales se basan en la ponderación y normalización de cada una de las variables.

2.7 Sistemas de Información Geográfica (SIG)

Es una integración organizada de hardware, software, datos geográficos y personal, diseñada para capturar, almacenar, manejar, analizar, modelar y representar en todas sus formas la información geográficamente referenciada con el fin de resolver problemas complejos de planificación y gestión. También puede definirse como un modelo de una parte de la realidad referido a un sistema de coordenadas terrestre y construido para satisfacer unas necesidades concretas de información (Sastre, 2010).

Una de las razones principales para utilizar un SIG es estructurar la información espacial dentro de un sistema que permite separar la información en diferentes capas temáticas y la almacena independientemente, permitiendo trabajar con ellas de manera rápida y sencilla, facilitando al usuario la posibilidad de analizar la información existente, con el fin de generar otra nueva que no se podría obtener de otras formas de análisis, como representa en la Figura 14 (Avilés & Lizzete, 2009).

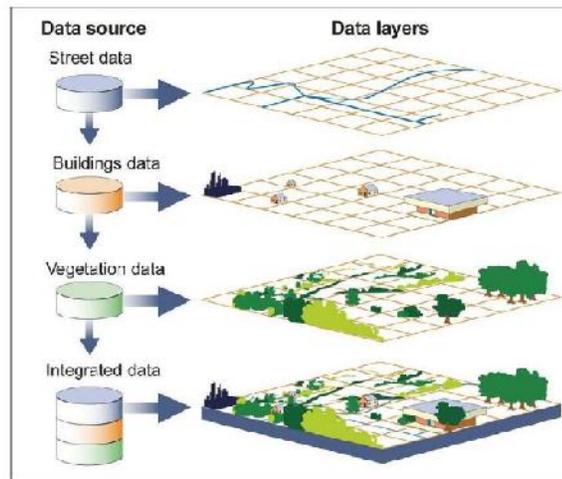


Figura 14. Representación gráfica de un SIG

Fuente: (National Geographic, 2010)

Dentro de los beneficios que brinda el uso un SIG es que, una vez ingresada la base de datos y creadas las respectivas coberturas, cada una con sus atributos, permite visualizar esta información, capa por capa, a manera de mapas de presentación. Asimismo, se obtienen notas, imágenes y bases de datos ligadas a los objetos georeferenciados (Jaramillo, 2014).

2.7.1 Base de Datos Geográfica (Geodatabase)

La geodatabase es la estructura de datos nativa para el software ArcGis y es el formato de datos principal que se utiliza para la edición y administración de datos, almacena toda la información geográfica. Cuenta con un modelo de información integral para representar y administrar información geográfica. Este modelo de información integral se implementa como una serie de tablas que almacenan clases de entidad, datasets ráster y atributos (Saavedra & Vinueza, 2013).

Además, los objetos de datos SIG avanzados agregan comportamiento SIG, reglas para administrar la integridad espacial y herramientas para trabajar con diversas relaciones espaciales de las entidades, los ráster y los atributos principales. La GDB provee herramientas topológicas para la integridad geométrica y para atributos en las capas de información.

2.7.2 Aplicación y uso de los SIG en los Modelos de Distribución de Especies

Los SIG han evolucionado durante los últimos años la forma de almacenar, analizar y disponer de grandes volúmenes de información que describen la cobertura terrestre en función de elementos bióticos y abióticos, en donde anteriormente la realización de modelos de predicción requería de la experticia en el manejo de variados software de SIG y estadística por parte de los investigadores; acompañado a esto la dificultad de tener al alcance las respectivas licencias y equipos de almacenamiento como la disponibilidad de datos espaciales de los factores ambientales a escala local, regional y mundial (Carrada & Vázquez, 2012).

Actualmente, el desarrollo de plataformas virtuales libres permite el acceso a programas computacionales de última generación que integran distintas técnicas de pre-análisis tales como, visualización de la información espacial, corrida de modelos con algoritmos de distinta naturaleza para el posterior mapeo y post-análisis de la eficiencia de los modelos con el aporte de las variables para la predicción (Espíndola, 2013).

Los programas SIG, como ArcGis, Diva-Gis, incluyen la posibilidad de modelar nichos ecológicos con base en información ambiental disponible de los sitios donde se han observado las especies (puntos de presencia). Como tal, la modelación de distribución de especies es útil para priorizar las acciones de conservación (Souza & Giovanni, 2009).

2.7.3 Uso de SIG y la vigilancia en Salud Pública

La Salud en el país está mejorando notablemente, sin embargo el uso de nuevas tecnologías para el desarrollo de estrategias en el control y mitigación ante enfermedades epidemiológicas, hace que el uso de SIG tome fuerza, como una buena alternativa en la toma de decisiones. La relación vector – enfermedad, ahora que el cambio climático se ha hecho más evidente, se hace más estrecha, por lo que al determinar la presencia de un vector, existe la posibilidad de tener la presencia de una enfermedad relacionada con el mismo (Yumiseva, 2014).

La complejidad de cada una de los procesos de transmisión de enfermedades genera también diferentes tipos de modelos a construir, sin embargo, la versatilidad de las plataformas informáticas actuales, permiten que los epidemiólogos actuales manejen y puedan desarrollar modelos espaciales (Hay, 2006).

Sin duda entonces, la utilización de los SIG en la actualidad se ha vuelto una herramienta casi indispensable para generar y manejar información de casos humanos o de reporte de presencia de vectores o reporte de condiciones especiales para el apareamiento de enfermedades (ver Figura 15) los que con adecuadas políticas de manejo y control dan un gran soporte sobre todo en tomas de decisión política referentes a Salud Pública (Yumiseva, 2014).



Figura 15. Uso de SIG en análisis de situación de salud

Fuente: (Yumiseva, 2014)

Los Sistemas de Información Geográfica (SIG) es un conjunto organizado de hardware, software y datos geográficos diseñado para capturar, almacenar, generar y manipular en todas sus formas la información geográficamente referenciada; dentro del campo de la salud permite desarrollar estrategias en el control y mitigación ante enfermedades epidemiológicas.

2.8 Biología del Vector *Aedes aegypti*



Figura 16. Vector *Aedes aegypti*

Fuente: (BBC Mundo, 2016)

El Mosquito *Aedes aegypti* es una especie diseminada por el hombre, principal vector de los virus que causan el dengue y la fiebre amarilla, así como de otras enfermedades, como la Chikungunya, la fiebre de Zika (ver Figura 16); esta especie tuvo su origen en el cinturón tropical de África donde generalmente se encuentran las especies del subgénero *Stegomyia*, fue introducido en América por medio del transporte de sus adultos, huevos, larvas o ninfas en barcos, aviones y transportes terrestres, además es el principal vector de los virus que causan el dengue, fiebre amarilla, Chikungunya y Zika (Consoli & De Oliveira, 1994).

Los seres humanos comúnmente se contagian por picaduras del mosquito (hembras) infectadas al succionar sangre de personas portadoras del virus; esto implica que son las personas, más que los mosquitos, quienes propagan rápidamente el virus ya sea durante el embarazo donde una mujer embarazada puede transmitir el virus a su feto causando microcefalia al neonato, mediante transmisión sexual y a través de transfusiones de sangre (OMS, 2015).

2.8.1 Clasificación Taxonómica

La familia Culicidae consta de 20 géneros en el Nuevo Mundo (culícidos); pertenecen al suborden Nematocera, Orden Diptera. Las características son más obvias para separar a los mosquitos adultos de los otros dípteros, por la combinación de escamas en las venas alares y en el margen posterior de las alas. La proboscis es larga, las antenas largas y filamentosas con 14 ó 15 artejos con sedas en espiral. Sin

excepción, las larvas son acuáticas, apodas, tórax en forma de bulbo más ancho que la cabeza y abdomen, cápsula cefálica completa, un solo par de estigmas funcionales dorsales en el octavo segmento abdominal (Harwood & James, 1987).

Hay tres Subfamilias: 1) Toxorhynchitinae (Megarhininae): adultos grandes con escamas de colores metálicos, proboscis marcadamente curva hacia abajo, se alimenta de néctares de flores, sus larvas son depredadoras principalmente de otros culícidos. 2) Culicinae: los palpos de las hembras miden menos del largo de la proboscis, el escutelo es trilobulado y las hembras son hematófagas. 3) Anophelinae, los palpos para los dos sexos son tan largos o casi tanto como la proboscis, el escutelo tiene forma de banda en media luna, las hembras también son hematófagas. En particular, la especie de interés pertenece a la tribu Culicini; al grupo *Aedes*; género *Aedes* Meigen, 1818; subgénero *Stegomyia* Theobald, 1901; grupo *Aedes aegypti* Linnaeus, 1762 (Bates, 1970), como se muestra en la Figura 17.



Figura 17. Familia *Aedes aegypti*

Fuente: (OMS, 2015)

En el continente africano existen tres formas. Una es la forma típica, otra *Aedes aegypti queenslandensis* y *Aedes Aegypti formosus*, este último es un mosquito selvático de color más oscuro y talla pequeña (Gómez & Rodríguez, 2013). Únicamente las dos primeras formas se encuentran en el Continente Americano; les tomó décadas adecuarse y dispersarse por el Caribe y establecerse en el Continente, lográndolo debido a que el tránsito marítimo se intensificó en esta región, desde ese entonces (Gómez & Rodríguez, 2013).

2.8.2 Ciclo de Vida

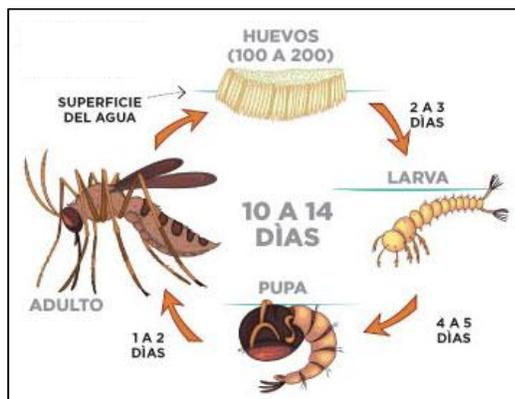


Figura 18. Ciclo de vida *Aedes aegypti*

Fuente: (MSP, 2015)

El *Aedes aegypti* mide menos de 1 cm, es de color negro y tiene manchas blancas y rayas en el cuerpo y las piernas. Es un mosquito peligroso, porque no tiene sonido. Es decir, la persona no se da cuenta de que está picado, lo que facilita la transmisión de la enfermedad, en la siguiente Tabla 3 se presenta un resumen del ciclo de vida del vector, con una breve descripción de cada fase (ver Figura 18).

Tabla 3.
Ciclo de vida *Aedes aegypti*

Fase	Descripción
Huevos	En esta fase la hembra coloca alrededor de 400 huevos en el agua, midiendo aproximadamente un milímetro de longitud, estos pueden estar solos o flotando agrupados; el embrión dentro del huevo es capaz de resistir largos períodos de desecación por meses o hasta por más de un año.
Larvas	Los huevos depositados anteriormente por la hembra se convierten en larvas, estas se desarrollan alrededor de 4 veces antes de convertirse en pupa, tarda alrededor de 2 días a 1 semana que ha dedicado a alimentarse y a crecer, ocude la muda permitiendo de esta manera el subsecuente desarrollo, se endurecen y oscurecen
Pupa	Fase que se antepone a la transformación en mosquito, la larva se transforma en pupa y se mantiene en este estado por unos 7 días. Las pupas de los mosquitos son diferentes a las de otros insectos holometabólos por presentar reacciones

Continúa →

	inmediatas a estímulos externos tales como vibraciones y cambios en la intensidad de la luz, desplazándose activamente por todo el criadero.
Mosquito Adulto	La pupa se abre y deja salir el mosquito totalmente formado, ya adulto, vive aproximadamente de 1 a 2 meses, con tamaño pequeño, de unos 5 milímetros de largo, con detalles morfológicos visibles con lupa de mano; cuerpo de color oscuro con manchas blancas en su dorso. La función más importante del adulto de <i>Aedes Aegypti</i> es la reproducción es decir la dispersión de la especie

Fuente: (MSP, 2015)

2.8.3 Hábitat y Alimentación

Depósitos de agua, ubicados en objetos o construcciones, como neumáticos, baterías viejas, recipientes de todo tipo, botellas, floreros y piletas, entre otros, le sirven al *Aedes aegypti* para establecer sus criaderos en agua limpia, con bajo tenor orgánico y de sales disueltas, mediante la puesta de huevos en la superficie del recipiente a la altura de la interface agua-aire (Carrada & Vázquez, 2012). Esta especie es más tolerante de las bajas temperaturas y mantiene una amplia variedad de criaderos, tanto en recipientes artificiales como naturales, tales características lo transforman en un vector de más difícil control e improbable eliminación, una vez que éste se radica (Carrada & Vázquez, 2012).

Las poblaciones del mosquito son más abundantes durante el verano; la temperatura y la humedad son factores críticos que afectan a los huevos y adultos (15°C - 42°C), con relación a la altitud se halla esta especie hasta 2200 msnm, sin embargo se adapta fácilmente a las condiciones climáticas donde habitan; el cambio climático ha propiciado que el mosquito *Aedes aegypti* se haya adaptado a zonas de hasta 2500 metros de altura sobre el nivel del mar y a climas fríos, según se ha reportado desde inicios de 2015 en publicaciones de revistas científicas y en estudios de universidades europeas y asiáticas, lo que implica que este vector, potencialmente, podría sobrevivir y reproducirse en ciudades con una altitud elevada (OMS, 2015).

Según Espinal (2016) Director de Enfermedades Transmisibles de la Organización Panamericana de la Salud (OPS), menciona que ciertas condiciones

vinculadas a la pobreza, hacen que las personas con escasos recursos estén más expuestas a que el mosquito se reproduzca en aguas limpias, debido que al no disponer de accesos directos al líquido vital estén obligados a almacenarlos en bidones en sus casas haciéndolos más vulnerables.

Su alimentación puede hacerla en cualquier momento de acuerdo a la disponibilidad de quien se alimenta (puede picar varias veces a las personas de una casa). Las proteínas contenidas en la sangre le son indispensables para la maduración de los huevecillos. La variación de temperatura y humedad, así como la latitud pueden hacer variar estos rangos del ciclo de vida de las cepas de mosquitos (MSP, 2015).

El Mosquito *Aedes aegypti* es una especie diseminada por el hombre, principal vector de los virus que causan el dengue, la fiebre amarilla, la Chikungunya y el virus del Zika; su ciclo de vida se divide en cuatro fases que son: Huevo, Larva, Pupa y Adulto, su hábitat principal es cualquier depósito de agua limpia que se encuentre a temperaturas que rodeen los 15°C - 42°C y una altitud máxima de 2200 msnm, aunque se adaptan fácilmente a su entorno.

CAPITULO III

METODOLOGÍA

3.1

3.1 Resumen

En el presente capítulo se aplicaron los métodos de Maxent, Regresión Logística, Lógica Difusa y MARS con el fin de determinar las zonas con mayor probabilidad de presencia del vector dentro del área de estudio (Ecuador Continental); en esta fase de estudio también se definen los procesos previos que se deben realizar a las variables de entrada como ponderación, normalización y rasterización, los software que se deben utilizar y los algoritmos adicionales que se deben tomar en cuenta para realizar cada modelamiento.

3.2 Esquema Metodológico

La metodología planteada para realizar el presente proyecto se estructuró en cuatro fases que se describen a continuación (ver Figura 19):

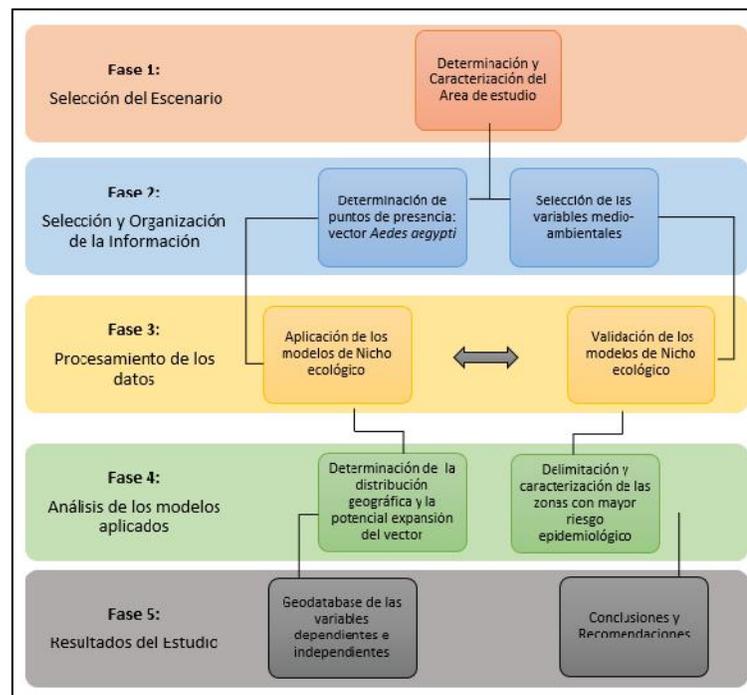


Figura 19. Esquema Metodológico

El esquema metodológico determina las directrices que se abordaron dentro del presente proyecto, el cual está dividido por fases, empezando por la selección del área de estudio hasta culminación del mismo con los resultados del modelamiento de la especie.

3.3 Recopilación de Información

Como se mencionó en el capítulo anterior para el modelamiento de la distribución espacial de una especie, es necesario contar con dos tipos de información que son: las variables dependientes (datos de presencia, ausencia o pseudo-ausencia) y las variables independientes (predictoras); además de una serie de técnicas de modelado ecológico que forman parte de software libres y versión de prueba tales como Maxent, MARS, SPSS que emulan o muestran el resultado de la investigación a través de modelados estadísticos, determinando así la idoneidad del hábitat para el desarrollo de la especie; también se utilizó el software ArcGis versión 9.x que permitió recopilar, organizar, administrar, generar y analizar los insumos necesarios para cada modelo utilizado. Finalmente cabe señalar que toda la información recopilada como generada, se manejó con el Sistema de Referencia WGS84 con proyección UTM, zona 17 Sur.

3.3.1 Variables Dependientes

3.3.1.1 Datos de presencia *Aedes aegypti*

Los registros de presencia del vector *Aedes aegypti* se obtuvieron de publicaciones periódicas (Gacetas) emitidas por la Subsecretaría de Vigilancia de la Salud Pública a través de la Dirección Nacional de Vigilancia Epidemiológica especializada en Enfermedades Transmitidas por Vectores, que desde finales del año 2015 (semanas epidemiológicas 52-53) empezaron a publicarlos; los datos para el modelamiento de la distribución de la especie se tomaron hasta el reporte emitido el 14 de septiembre de 2016 (semanas epidemiológicas 1-36).

De acuerdo con el Ministerio de Salud Pública del Ecuador y como se mencionó en el capítulo anterior el Virus del Zika es transmitido generalmente por la picadura

del Vector *Aedes aegypti*, por tal motivo se georeferenció únicamente las localidades donde existen casos confirmados AUTÓCTONOS del virus (42 registros), que según la Organización Mundial de la Salud son registros de contagio epidemiológico local (OMS, 2015), es decir existe presencia del vector in situ; en la Figura 20 se resume las provincias con sus respectivos cantones en donde se encuentran los casos confirmados autóctonos e importados del Virus Zika (ZikaV) dentro del Ecuador Continental.

Prov Domic	Cantón Domic	Autóctono	Importado	Total gene..
AZUAY	CUENCA	1		1
CHIMBORAZO	CHUNCHI	1		1
EL ORO	HUAQUILLAS	1		1
	MACHALA	6		6
	PASAJE	1		1
ESMERALDAS	ATACAMES	1		1
	ESMERALDAS	106		106
	MUISNE	1		1
	QUININDE	13		13
	RIOVERDE	1		1
	SAN LORENZO	54		54
GALAPAGOS	SANTA CRUZ	2		2
GUAYAS	BALZAR	2		2
	DAULE	2		2
	GNRAL. ANTONIO E..	1		1
	GUAYAQUIL	98	1	99
	PLAYAS	1		1
LOS RIOS	BUENA FE	2		2
	QUEVEDO	3		3
	VENTANAS	10		10
	VINCES		1	1
MANABI	24 DE MAYO	1		1
	CHONE	2		2
	JARAMIJO	19		19
	JIPIJAPA	92		92
	MANTA	827	1	828
	MONTECRISTI	110		110
	PAJAN	1		1
	FEDERNALES	2		2
	PORTOVIEJO	638	1	639
	PUERTO LOPEZ	4		4
	ROCAFUERTE	3	1	4
	SAN VICENTE	3		3
	SANTA ANA	5		5
SUCRE	52		52	
	TOSAGUA	1		1
ORELLANA	LA JOYA DE LOS S..	1		1
PICHINCHA	MEJIA	1		1
	QUITO	16	15	31
SANTO DOMIN..	LA CONCORDIA	8		8
	SANTO DOMINGO	6	1	7
SUCUMBIO	LAGO AGRIO	1		1
	SHUSHUFINDI	2		2
Total general		2.102	21	2.123

Figura 20. Casos confirmados autóctonos e importados del Virus Zika

Fuente: (Subsecretaria de Vigilancia de la Salud Publica, 2016)

En la Figura 21 se muestra los puntos de presencia de Vector *Aedes aegypti* georeferenciados dentro del Ecuador continental.

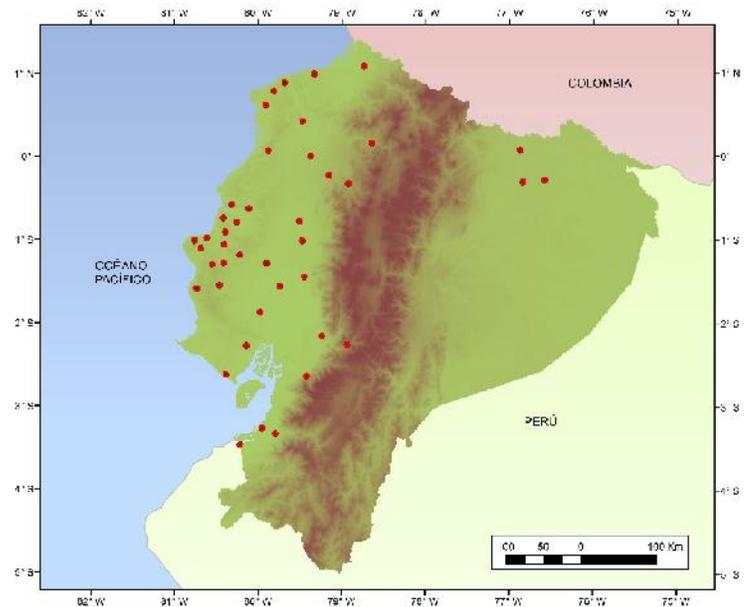


Figura 21. Presencia de Vector *Aedes aegypti* en el Ecuador Continental

Fuente: (Subsecretaría de Vigilancia de la Salud Pública, 2016)

3.3.1.2 Datos de ausencia o Pseudo-ausencia *Aedes aegypti*

Los datos de ausencia son fundamentales en los modelos de distribución de las especies, sin embargo actualmente no existen investigaciones científicas que determinen zonas no aptas para la reproducción del vector debido a que se adapta fácilmente a cualquier entorno; por tal razón se crearon pseudo-ausencias que son ausencias estimadas a partir de los datos biológicos, ecológicos e históricos del vector.

Las pseudo-ausencias para el presente estudio se generaron en función de los datos de presencia como sugieren varios autores, que manifiestan que a partir de los mismos se establezca un radio de 30 Km en el cual las condiciones medio-ambientales, topográficas y paisajísticas se estabilizan, es decir existe la probabilidad de presencia del vector dentro de esta área. Esta delimitación se realizó con la herramienta *buffer* de ArcGis, posteriormente se borró estas zonas de nuestra área de

estudio (Ecuador continental) mediante la herramienta *erase* de ArcGis, para finalmente crear puntos aleatorios que exceden en un 10% a la muestra total de presencia del vector (48 puntos pseudo-ausencia) utilizando la herramienta *create random points* de ArcGis como se muestra en la **¡Error! No se encuentra el origen de la referencia..**

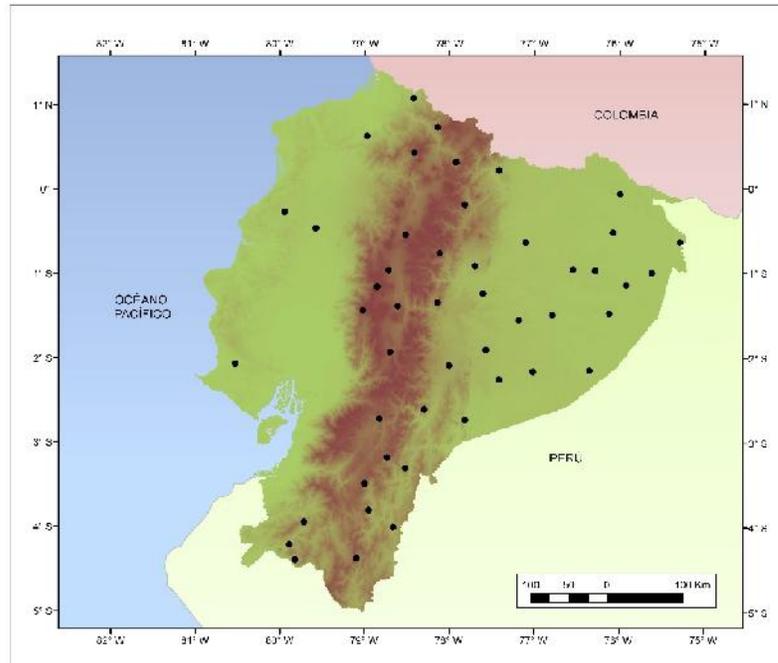


Figura 22. Pseudo-ausencia de Vector *Aedes aegypti* en el Ecuador Continental

3.3.2 Variables Independientes

3.3.2.1 Variables Bioclimáticas

Las 19 variables Bioclimáticas fueron tomadas del Sitio Web Worldclim (<http://www.worldclim.org/>), que reúne un conjunto de datos climáticos a nivel mundial con una resolución espacial de 30 segundos que es 1km² aproximadamente en la línea del Ecuador. Las capas que ofrece este portal fueron creadas a través de la interpolación de los promedios de datos climáticos mensuales, trimestrales y anuales para cada estación. Las variables que se incluyeron al modelo son: Temperatura media anual, Promedio del rango diario, Isotermalidad, Temperatura estacional, Máxima Temperatura del mes más caliente, Mínima Temperatura del mes más frío, Rango de Temperatura Anual, Promedio de temperatura del trimestre más húmedo, Promedio de temperatura del trimestre más seco, Promedio de temperatura del

trimestre más caliente, Promedio de temperatura del trimestre más frío, Precipitación anual, Precipitación del mes más húmedo, Precipitación del mes más seco, Precipitación estacional (Coeficiente de variación), Precipitación del trimestre más húmedo, Precipitación del trimestre más seco, Precipitación del trimestre más caliente y Precipitación del trimestre más frío.

Las capas están en formato .grid a nivel mundial, por tal razón se cortaron todas las variables utilizadas al tamaño del área de estudio (Ecuador Continental) con el mismo tamaño del pixel (1000x1000) metros y de igual manera el mismo número de filas y columnas (650x721) con la herramienta *Extract by Mask* de ArcGis como se puede observar en la **¡Error! No se encuentra el origen de la referencia..**

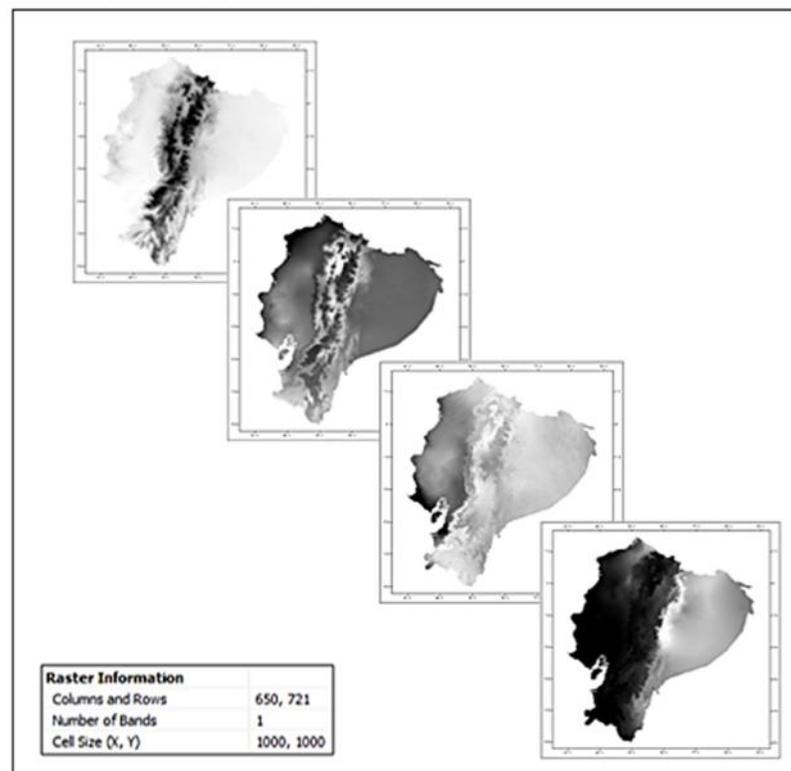


Figura 23. Variables Bioclimáticas para el modelamiento del *Aedes aegypti*

3.3.2.2 Variables Geográficas

Las variables geográficas que se utilizaron para realizar el modelamiento, forman parte de la cartografía base 1: 50000, que consta dentro del Geoportal del Sistema de Información Nacional SNI (<http://sni.gob.ec/inicio>), con excepción de la altitud; a

cada variable se le realizó un proceso previo que se describe a continuación antes de que formen parte de cada modelo:

3.3.2.2.1 Altitud

La información correspondiente a la altitud, se recopiló del Sitio Web Worldclim que a través de La Misión Topográfica Radar Shuttle (SRTM) impulsada por la NASA, elaboraron un modelo digital de la superficie terrestre a partir de la información recabada desde el espacio. De la misma forma que se hizo con las variables ambientales, se cortó y se ajustó la resolución y el tamaño del pixel al área de estudio con la herramienta Extract by Mask de ArcGis, como se puede observar en la **¡Error! No se encuentra el origen de la referencia..**

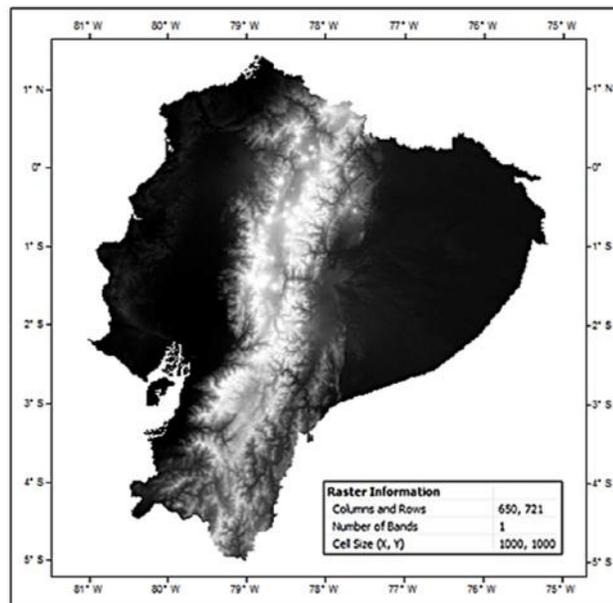


Figura 24. Variable Altitud para el modelamiento del *Aedes aegypti*

3.3.2.2.2 Áreas susceptibles a Inundaciones

La información en crudo de esta variable se encuentra en formato vector, por tal motivo se utilizó la herramienta *Euclidean distance* de ArcGis, que al rasterizarlo proporciona la distancia desde cada celda hasta el origen más cercano de las áreas susceptibles a Inundaciones (ver

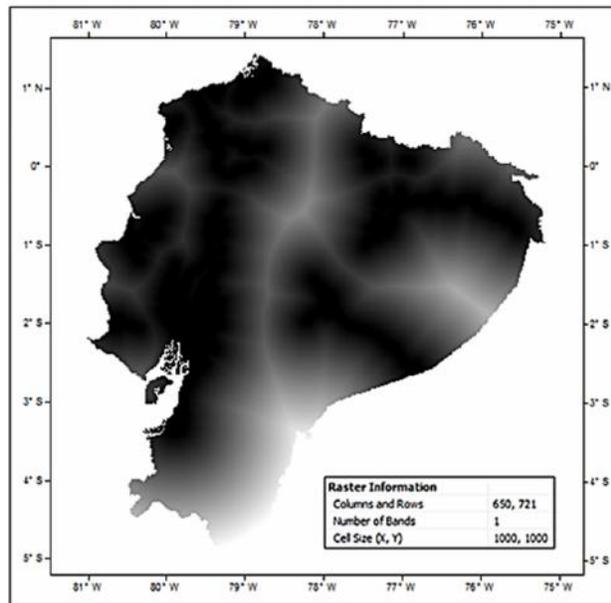


Figura 25. Variable Inundaciones para el modelamiento del *Aedes aegypti*

).

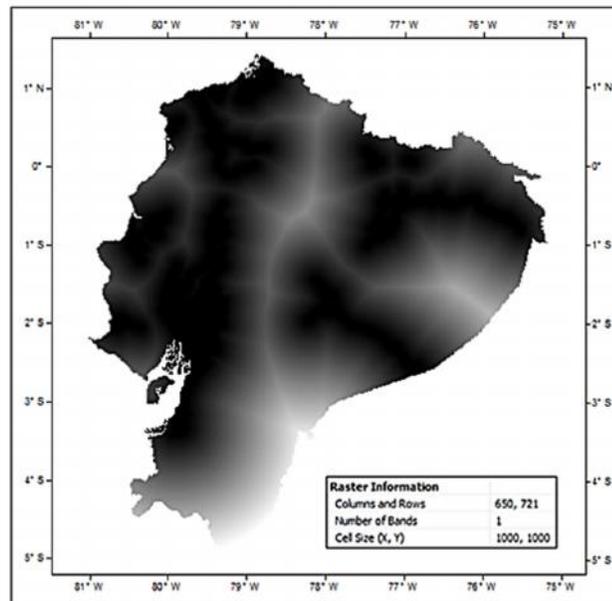


Figura 25. Variable Inundaciones para el modelamiento del *Aedes aegypti*

3.3.2.2.3 Centros Poblados

Al igual que la variable anterior (Áreas susceptibles a Inundaciones) la información en crudo se encuentra en formato vector, por tal motivo se utilizó la

herramienta *Euclidean distance* de ArcGis, que al rasterizarlo proporciona la distancia desde cada celda hasta el origen más cercano a centros poblados (ver **¡Error! No se encuentra el origen de la referencia.**).

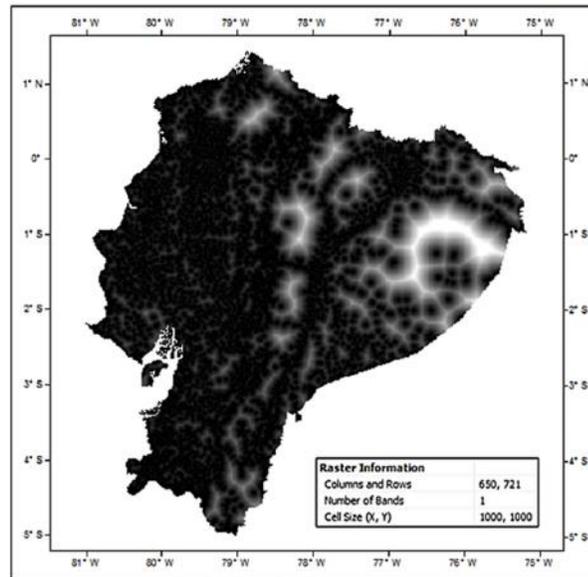


Figura 26. Variable Centros Poblados para el modelamiento del *Aedes aegypti*

3.3.2.2.4 Uso de Suelo

Para la variable de uso de suelo se utilizó el método Analítico de Saaty que se fundamenta en jerarquizar los componentes o variables mediante valores numéricos para los juicios de preferencia determinando de esta manera qué variable tiene la más alta prioridad; en la Tabla 4 se ponderó el uso de suelo a nivel nacional en función del porcentaje de presencias (Ver Figura 27), además de las características que los hacen más apto para que se vuelva hábitat del vector, posteriormente se rasterizó la información mediante la herramienta *Feature to Raster* de ArcGis (Ver **¡Error! No se encuentra el origen de la referencia.**) en función de un nuevo campo llamado “pesos” donde el rango de los valores esta entre 0 y 1.

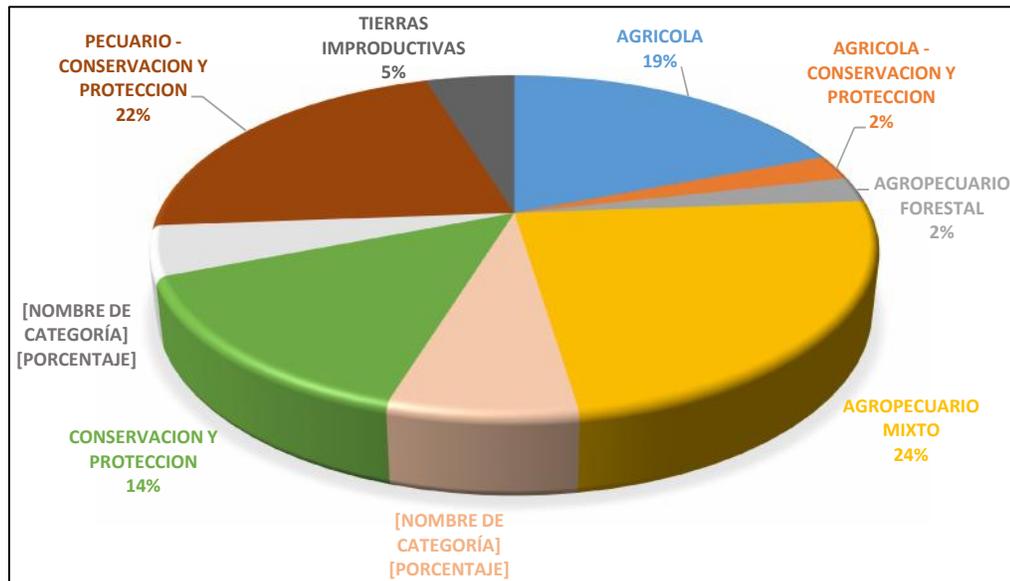


Figura 27. Presencia del *Aedes aegypti* en función del Uso de Suelo

Tabla 4.
Matriz de Saaty para Uso de Suelo

Variables	V	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	W	C	Pesos	Poderación
Cuerpo de Agua	V1	1.00	1.00	0.50	0.50	0.33	0.33	0.25	0.20	0.17	0.14	0.13	0.32	0.02	1	0.024
Forestal	V2	1.00	1.00	0.50	0.50	0.33	0.33	0.25	0.20	0.17	0.14	0.13	0.32	0.02	1	0.024
Agropecuario Forestal	V3	2.00	2.00	1.00	1.00	0.67	0.67	0.50	0.40	0.33	0.29	0.25	0.65	0.05	2	0.048
Agrícola - Conservación y Protección	V4	2.00	2.00	1.00	1.00	0.67	0.67	0.50	0.40	0.33	0.29	0.25	0.65	0.05	2	0.048
Pecuario	V5	3.00	3.00	1.50	1.50	1.00	1.00	0.75	0.60	0.50	0.43	0.38	0.97	0.07	3	0.071
Tierras Improductivas	V6	3.00	3.00	1.50	1.50	1.00	1.00	0.75	0.60	0.50	0.43	0.38	0.97	0.07	3	0.071
Antrópico	V7	4.00	4.00	2.00	2.00	1.33	1.33	1.00	0.80	0.67	0.57	0.50	1.30	0.10	4	0.095
Conservación y Protección	V8	5.00	5.00	2.50	2.50	1.67	1.67	1.25	1.00	0.83	0.71	0.63	1.62	0.12	5	0.119
Agrícola	V9	6.00	6.00	3.00	3.00	2.00	2.00	1.50	1.20	1.00	0.86	0.75	1.94	0.14	6	0.143
Pecuario - Conservación y Protección	V10	7.00	7.00	3.50	3.50	2.33	2.33	1.75	1.40	1.17	1.00	0.88	2.27	0.17	7	0.167
Agropecuario Mixto	V11	8.00	8.00	4.00	4.00	2.67	2.67	2.00	1.60	1.33	1.14	1.00	2.59	0.19	8	0.190

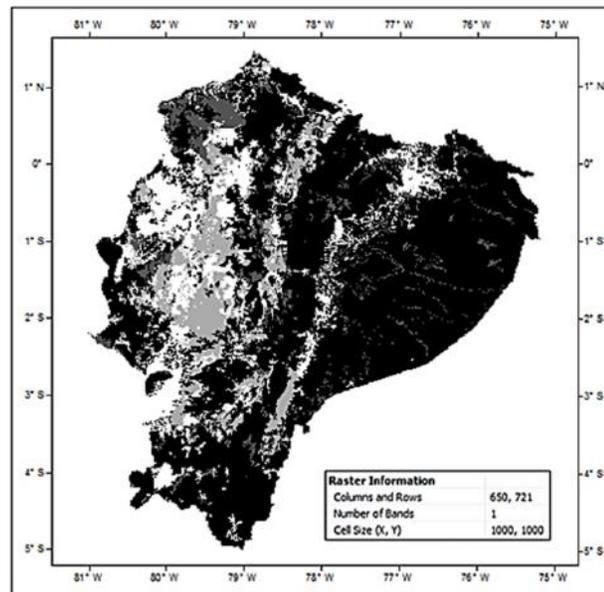


Figura 28. Variable Uso de Suelo para el modelamiento del *Aedes aegypti*

3.3.2.3 Variables Socio-Económicas

La variable Índice de Pobreza como se dijo anteriormente está vinculada a la calidad de vida en el entorno socio-económico; es decir si la calidad de vida es baja existirán viviendas sin acceso a agua potable mediante una red hídrica además de no contar con sistemas de drenaje, por lo que la población se ve obligada a almacenar agua en depósitos internos o externos acción que favorece la proliferación del vector *Aedes aegypti*. En el Geoportal del Instituto Nacional de Estadística y Censos (<http://www.ecuadorencifras.gob.ec/geoportal/>) consta esta información en formato vector, que se rasterizó mediante la herramienta de interpolación *IDW* de ArcGis. Este proceso calcula los valores de cada una de las celdas a través de una combinación ponderada linealmente en función de un campo determinado. En el presente estudio se consideró el porcentaje de índice de pobreza a través de los centroides de todas las parroquias a nivel nacional como se muestra en la **¡Error! No se encuentra el origen de la referencia..**

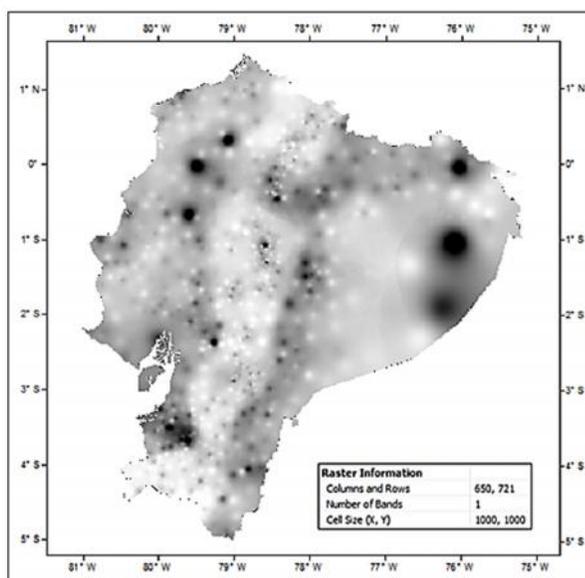


Figura 29. Variable Índice de Pobreza para el modelamiento del *Aedes aegypti*

El resumen de las variables predictoras que se utilizaron para el modelamiento de vector *Aedes aegypti* se encuentra a continuación en la Tabla 5.

Tabla 5.
Variables Predictoras para el modelamiento del *Aedes aegypti*

<i>Variables</i>	<i>Descripción</i>	<i>Tipo</i>
<i>Bio_1</i>	Temperatura media anual	Bioclimática
<i>Bio_2</i>	Promedio del rango diario (Promedio mensual (max temp - min temp))	Bioclimática
<i>Bio_3</i>	Isotermalidad (BIO2/BIO7) (* 100)	Bioclimática
<i>Bio_4</i>	Temperatura estacional (Desviación estándar *100)	Bioclimática
<i>Bio_5</i>	Max Temperatura del mes más caliente	Bioclimática
<i>Bio_6</i>	Min Temperatura del mes más frío	Bioclimática
<i>Bio_7</i>	Rango de Temperatura Anual (BIO5-BIO6)	Bioclimática
<i>Bio_8</i>	Promedio de temperatura del trimestre más húmedo	Bioclimática
<i>Bio_9</i>	Promedio de temperatura del trimestre más seco	Bioclimática
<i>Bio_10</i>	Promedio de temperatura del trimestre más caliente	Bioclimática
<i>Bio_11</i>	Promedio de temperatura del trimestre más frío	Bioclimática
<i>Bio_12</i>	Precipitación anual	Bioclimática

Continúa →

Bio_13	Precipitación del mes más húmedo	Bioclimática
Bio_14	Precipitación del mes más seco	Bioclimática
Bio_15	Precipitación estacional (Coeficiente de variación)	Bioclimática
Bio_16	Precipitación del trimestre más húmedo	Bioclimática
Bio_17	Precipitación del trimestre más seco	Bioclimática
Bio_18	Precipitación del trimestre más caliente	Bioclimática
Bio_19	Precipitación del trimestre más frío	Bioclimática
Altitud	Altura Referida al Nivel Medio del Mar	Geográfica
Inundaciones	Distancia a Áreas Susceptibles a Inundaciones	Geográfica
Uso de suelo	Estado Actual de la Distribución de la vegetación	Geográfica
Población	Distancia a Centros Poblados	Geográfica
Pobreza	Zonas con mayor Índice de Pobreza	Socio-económica

Cabe señalar que no todas las variables que se describieron anteriormente aportan de la misma manera a cada modelo, incluso se puede prescindir de cualquiera de ellas debido al estadístico que realiza cada método de modelamiento.

El modelamiento de las especies requiere como insumos de entrada variables dependientes (datos de presencia, ausencia o pseudo-ausencia) y las variables independientes (predictoras) dentro de un mismo Sistema Geográfico de Referencia; cierta información fue recopilada de instituciones tales como; MSP, SNI, IGM, WORLDCLIM, etc., y otros insumos fueron generados a través de herramientas geo-informáticas.

3.4 Normalización de las variables

Como se mencionó en el capítulo anterior, la mayoría de modelos de distribución de las especies requieren la homologación de los valores tanto de las variables dependientes como independientes, por tal razón, se adimensionó y referenció la información dentro de una misma escala; es decir en un rango de [0,1] donde se considera que existe una mayor probabilidad de presencia del vector cuando los valores son más cercanos a 1.

Dentro del área de estudio se determinaron los valores máximos y mínimos con presencia del vector de cada una de las variables, mediante la herramienta *Extract Multi Values to Points* de ArcGis con lo que se pudieron establecer las condiciones extremas para la sobrevivencia del vector y a su vez excluir las zonas con ausencia del mismo, como se evidencia en la Tabla 6. Mediante el uso de la herramienta “Reclassify” de ArcGis, se asignó valor “0” los rangos excluidos del raster que tiene los datos de la variable en particular.

Tabla 6.
Valores máximos y mínimos de las Variables Independientes

Variables	General		Presencia	
	Mínimo	Máximo	Mínimo	Máximo
Bio_1	-61	261	170	258
Bio_2	45	152	60	111
Bio_3	47	95	65	88
Bio_4	72	1743	170	1449
Bio_5	-6	331	237	321
Bio_6	-112	220	106	217
Bio_7	75	172	79	139
Bio_8	-60	269	170	267
Bio_9	-66	266	168	254
Bio_10	-59	270	171	269
Bio_11	-66	258	167	254
Bio_12	104	5092	202	3499
Bio_13	36	729	59	553
Bio_14	0	310	1	218
Bio_15	9	189	15	140
Bio_16	84	2012	155	1483
Bio_17	0	1047	3	747
Bio_18	10	1995	125	1483
Bio_19	0	1519	6	1010
Altitud	-17	6169	3	1550
Índice de Pobreza	0	97	39	85
Distancia a Centros Poblados	0	43705	0	3344
Distancia a Áreas de Inundaciones	0	199517	0	61320

Una vez obtenidos los valores de máximos y mínimos con presencia de vector se utiliza la fórmula (21) para normalizar citada en el capítulo anterior para cada una de las variables:

$$X_n = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad (21)$$

Donde X_n es la variable normalizada, X_i es la variable, X_{min} y X_{max} representan los valores mínimos y máximos respectivamente (ver **¡Error! No se encuentra el origen de la referencia.**). Se debe acotar que con la variable uso de suelo no se realizó este proceso, debido a que anteriormente fue normalizada mediante la aplicación del método de Saaty.

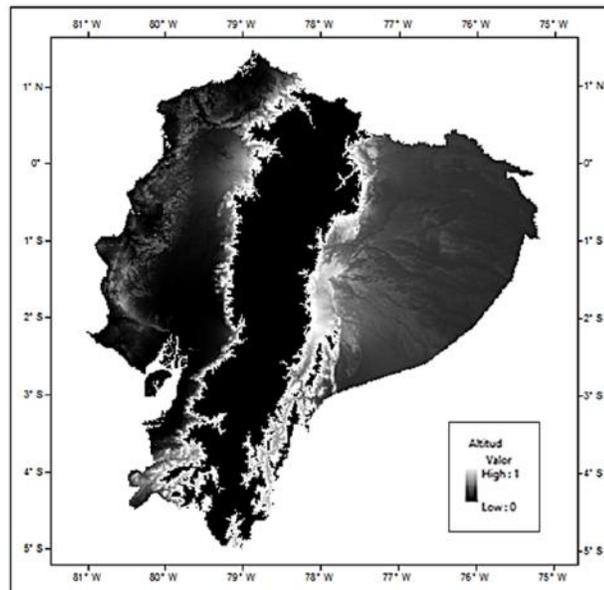


Figura 30. Reclasificación de valores de las Variables

El proceso de Normalización tiene como objetivo principal adimensionar y definir dentro de una sola escala [0,1] el valor de las variables, adicionalmente se ajustó a cada una de ellas en relación a las condiciones extremas para la sobrevivencia del vector, con el fin realizar un sesgo previo al modelamiento y minimizar posibles errores.

3.5 Modelos de distribución de las especies

En la presente investigación se aplicaron varios modelos estadísticos con la finalidad de determinar cuál de ellos es el más óptimo para predecir la distribución espacial del mosquito *Aedes aegypti* vector del virus Zika. Los modelos empleados se detallan a continuación:

3.5.1 Maxent (Máxima Entropía)

Para la aplicación del modelo de máxima entropía o Maxent se utilizó el software libre con el mismo nombre MaxEnt en su versión 3.3.3k desarrollado por Phillips, Dudik & Schapire y disponible de forma gratuita en la página web: <https://www.cs.princeton.edu/schapire/maxent>; como se mencionó anteriormente este estadístico es discriminante, ese decir necesita datos de presencia y de ausencia, sin embargo Maxent proporciona sus propias ausencias, denominadas “background” o muestra de fondo. Adicionalmente esta plataforma virtual no requiere que las variables predictoras estén normalizadas debido a que este proceso estadístico lo realiza internamente el programa.

3.5.1.1 Aplicación del Modelo

Previamente para que los datos de entrada sean compatibles con el programa Maxent, se debe generar un archivo delimitado por comas (formato .CSV) que contenga los datos de presencia del vector en tres columnas únicamente (el nombre de la especie y las coordenadas de ubicación en cada punto), tal como se muestra en la Figura 31.

```

ESPECIES,X,Y
Aedes_Aegypti,674635,10110900
Aedes_Aegypti,646638,10102624
Aedes_Aegypti,632831,10086205
Aedes_Aegypti,621150,10068681
Aedes_Aegypti,655904,10023254
Aedes_Aegypti,841318,10003193
Aedes_Aegypti,750453,9992684
Aedes_Aegypti,745310,9953994
Aedes_Aegypti,704736,9974825
Aedes_Aegypti,565290,9923670
Aedes_Aegypti,681085,10000077

```

Figura 31. Datos de presencia en formato .CSV

Adicionalmente todas las variables predictoras se deben transformar a formato ASCII mediante la herramienta *Raster to ASCII* de ArcGis. Maxent es un programa amigable con el usuario que cuenta con una ventana principal en donde se ingresan las variables compatibles con él, en la sección *Samples* se ingresan los datos de presencia del vector y las variables predictoras en la sección *Environmental layers*, como se indica en la Figura 32.

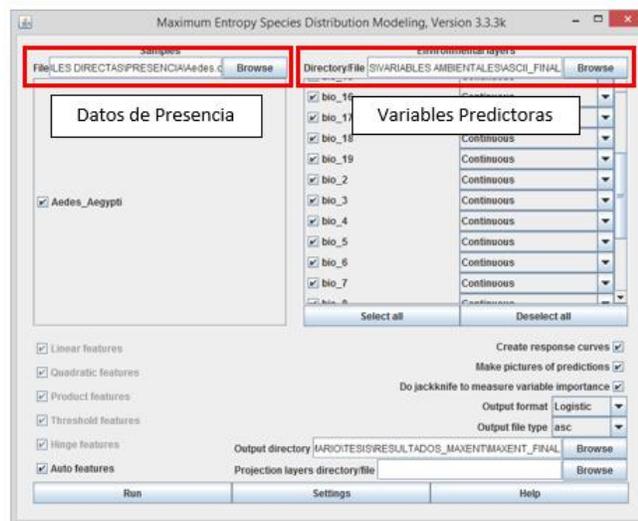


Figura 32. Interface principal Maxent

Fuente: (Maxent, 2016)

Se determinó a continuación si las variables ambientales que se van a ejecutar en el modelo son categóricas o continuas. Para el presente estudio se establecieron como continuas. En la opción de “Settings” se configura el porcentaje de puntos aleatorios que va a necesitar el programa (*Random test percentage*) para validar el modelo. Varios autores recomiendan que sea el 10% de muestra total, el resto de las opciones no es necesario modificarlas (default) debido a que se utilizan para estudios ecológicos más específicos (ver Figura 33).

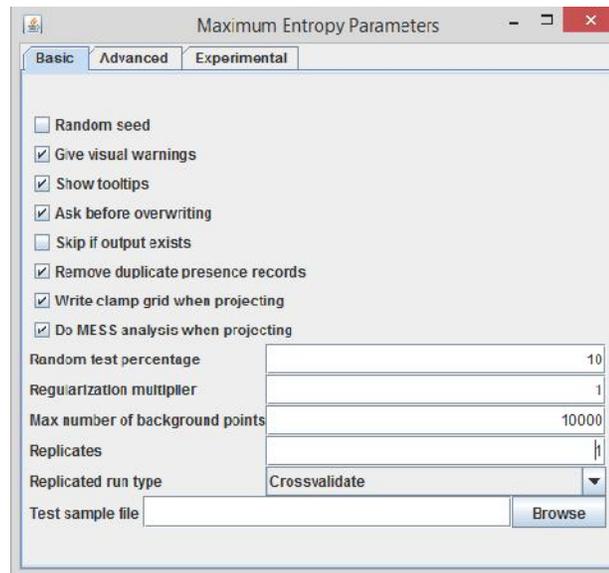


Figura 33. Configuración Maxent

Fuente: (Maxent, 2016)

Finalmente, se registra la dirección en donde se va a guardar la información debido a que Maxent además de generar archivos en formato .csv y .asc, proporciona un archivo principal con formato .html que contiene el resumen de todo lo que analizó el modelo en función de las variables ingresadas, que a continuación se describen.

3.5.1.1.1 Análisis curva ROC

Como se mencionó anteriormente, el modelo fue configurado para que se utilice el 10% de las muestras para usarlos como prueba simple dentro de un umbral de predicción binaria, definiendo de esta manera las condiciones adecuadas por encima de dicho umbral y las inadecuadas por debajo. La primera gráfica muestra una comparación entre las curvas ROC tanto de los datos de entrenamiento como los de prueba. La característica fundamental de esta gráfica es el Área bajo la curva es evaluar los datos de entrenamiento (ver Figura 34).

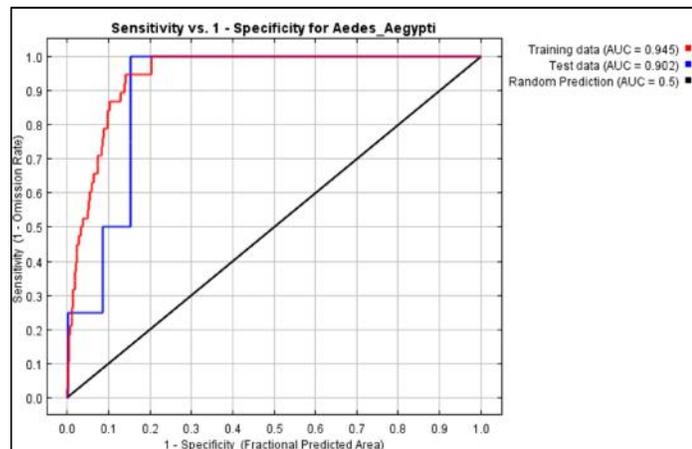


Figura 34. Curva ROC (Receiver Operating Characteristic analysis)

Fuente: (Maxent, 2016)

La gráfica determinó valores de AUC de 0,945 y 0,902 para los datos de entrenamiento y prueba respectivamente, estableciendo de esta manera que los estadísticos que realizaron el modelo son satisfactorios, donde la curva de color rojo (entrenamiento) representa el ajuste del modelo a los datos de muestreo. La curva de color azul (test) indica el grado de ajuste del modelo a los datos de test, y supone el test real del poder predictivo del modelo y la línea de color negro representa la predicción esperada si el modelo no fuese mejor que “al azar”.

3.5.1.1.2 Análisis de contribución de las variables

En el proceso de entrenamiento, Maxent permite distinguir de manera general cuales son las variables predictoras que contribuyen en mayor medida al modelo mediante la aplicación de una serie de algoritmos (ver Figura 35).

Variable	Percent contribution	Permutation importance
poblados	30.4	23
altitud	25.3	47
bio_15	16.6	0
inundaciones	6.2	4.9
bio_12	4.5	0
pobreza	4.2	5
bio_4	2.3	2.2
bio_18	2	0
bio_10	1.8	0.2
bio_19	1.5	0.3
use_soil	1.2	0.6
bio_17	1.1	0
bio_5	0.7	4.1
bio_13	0.6	5.4
bio_14	0.5	0.1
bio_9	0.3	3.8
bio_16	0.2	0
bio_6	0.2	2.9
bio_2	0.1	0.1
bio_8	0.1	0
bio_3	0.1	0.3
bio_1	0	0
bio_7	0	0
bio_11	0	0

Figura 35. Contribución de las variables predictoras

Fuente: (Maxent, 2016)

Las variables que contribuyen más al modelo son: Distancia a centros y la Altitud, mientras que las variables que aportan de manera nulo o insignificante son: Temperatura media anual, Rango de Temperatura Anual y Promedio de temperatura del trimestre más frío. El test de significancia o también llamado *jackknife*, es una prueba estadística que realiza Maxent para establecer la ganancia que aporta cada variable al modelo en conjunto y con respecto a las demás, este test es aplicado sobre los datos de entrenamiento, de prueba y con respecto al AUC.

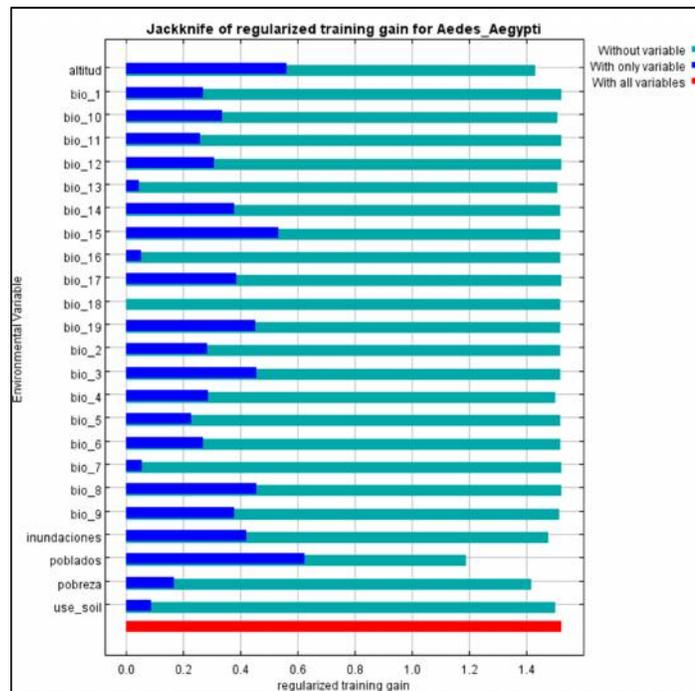


Figura 36. Test Jackknife en función del AUC

Fuente: (Maxent, 2016)

En la Figura 36 se refleja en color azul la contribución de cada variable predictora por sí sola, mientras que en color verde la contribución sin la variable en mención. Finalmente la barra de color roja muestra la contribución de todas las variables en conjunto en función al AUC. La altitud y la distancia a centros poblados en el test son las que más aportan, al contrario de la Precipitación del trimestre más caliente que no contribuye al modelo. Comúnmente este test se interpreta para decidir que variables deben ir en el modelo y cuáles no, en función de su grado de importancia, sin embargo se ha demostrado que este concepto es erróneo debido a que el comportamiento de todo el modelo sin la variable mencionada es relativamente similar comparada con las variables.

Maxent también brinda la oportunidad de efectuar un análisis individual de cada variable mediante las curvas de respuesta (ver Figura 37), las mismas que representan como cada variable influye sobre el modelo y la correlación que existe entre ellas.

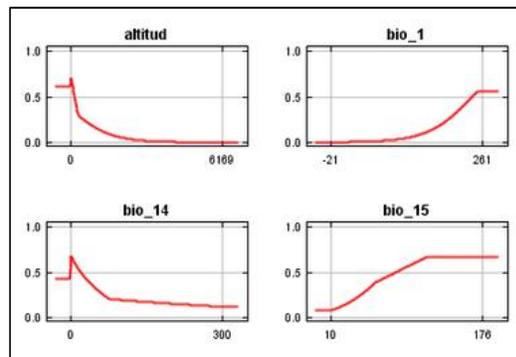


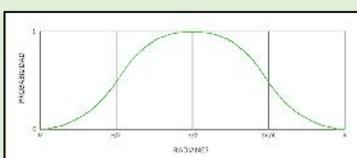
Figura 37. Curvas de respuesta

Fuente: (IBM SPSS, 2016)

3.5.2 Lógica Difusa

La principal aplicación de Lógica difusa como se mencionó anteriormente es representar valores cuantitativos (valores numéricos entre 0 y 1) a través de insumos lingüísticos es decir cualitativos mediante el manejo de dominios que no están dentro del ámbito de la lógica clásica. Para la Lógica Fuzzy las funciones que se aplican son el Seno y el Coseno debido a que, el rango en el que se trabaja es entre 0 y 1. La metodología que maneja la Lógica difusa consiste en determinar la interacción de cada variable que forma parte del modelo con la probabilidad de la presencia de la especie dentro de tres posibles escenarios o casos que se describieron en la Tabla 2 del capítulo anterior. A continuación se analiza cómo reaccionan las variables predictoras con respecto a la probabilidad de presencia de la especie, a fin de determinar cuál de los casos planteados en la metodología Difusa será aplicado. Este análisis se indica en la Tabla 7.

Tabla 7.
Análisis de las Variables

Variables	Descripción	Caso Fuzzy
Bioclimáticas	Las condiciones extremas meteorológicas, generalmente pueden deteriorar el hábitat de la especie, por tal motivo las condiciones no extremas representan una condición apropiada para la vida.	<p style="text-align: center;">Caso 1</p> 

Continúa →

Altitud	El <i>Aedes aegypti</i> tiene su hábitat dentro de un rango de elevación, al salir de este rango, ya sea por mayor o menor elevación, la especie mengua o desaparece.	Caso 1	
Índice de Pobreza	Ciertas condiciones vinculadas a la pobreza, hacen que las personas con escasos recursos estén más expuestas al mosquito y a las enfermedades que trasmite el mismo.	Caso 3	
Distancia a Centros Poblados	Varios estudios han demostrado que el <i>Aedes aegypti</i> tiene un rango de vuelo de aproximadamente 3500 metros a centros poblados, fuera de él la especie desaparece debido a que la hembra se alimenta de sangre humana.	Caso 1	
Distancia a Áreas de Inundaciones	Depósitos naturales de agua formados por inundaciones favorecen la proliferación del vector <i>Aedes aegypti</i> , debido a que ahí las hembras depositan sus huevos.	Caso 2	

Nota: Para la variable uso de suelo no fue necesario determinar el caso correspondiente de Fuzzy debido a que previamente se realizó este proceso mediante el método de Saaty.

3.5.2.1 Aplicación del Modelo

Una vez desarrollado el proceso de normalización que se realizó previamente y ya identificados los correspondientes escenarios (casos) de lógica difusa, se procedió a transformar el valor de las variables a radianes mediante la siguiente fórmula:

$$R = X_n * V \quad (23)$$

Donde R es el valor de la variable en radianes, X_n es la variable normalizada y V es el valor de 3.14159265359 ó 1.57079632679 según el rango correspondiente a cada escenario.

Para analizar la probabilidad de ocurrencia de cada variable es necesario utilizar las siguientes fórmulas:

$$P = \text{Sen}(R) \wedge P = \text{Cos}(R) \quad (24)$$

Las funciones trigonométricas que se utilizaron fueron el seno y el coseno debido a que luego de realizar el análisis de escenarios planteados por Fuzzy se determinó que el modelo cuenta con los tres casos.

Finalmente se realiza el promedio de las probabilidades de todas las variables mediante la herramienta *Raster Calculator* de ArcGis aplicando la fórmula que se describe a continuación:

$$Y = (P_{\text{Bio}_1} + P_{\text{Bio}_2} + P_{\text{Bio}_3} + P_{\text{Bio}_4} + P_{\text{Bio}_5} + P_{\text{Bio}_6} + P_{\text{Bio}_7} + P_{\text{Bio}_8} + P_{\text{Bio}_9} + P_{\text{Bio}_{10}} + P_{\text{Bio}_{11}} + P_{\text{Bio}_{12}} + P_{\text{Bio}_{13}} + P_{\text{Bio}_{14}} + P_{\text{Bio}_{15}} + P_{\text{Bio}_{16}} + P_{\text{Bio}_{17}} + P_{\text{Bio}_{18}} + P_{\text{Bio}_{19}} + P_{\text{altitud}} + P_{\text{pobreza}} + P_{\text{inundaciones}} + P_{\text{poblados}} + \text{Uso_suelo}) / 24$$

Donde

Y es la probabilidad del modelo

3.5.3 Regresión Logística

Para la aplicación del método de Regresión Logística se utilizó el Software estadístico SPSS con sus siglas en inglés (Statistical Package for the Social Sciences), que se emplea muy a menudo para realizar procesos analíticos con el fin de dirigir investigaciones y para la toma de mejores decisiones; dentro de este paquete estadístico existe el elemento de regresión logística binaria necesario para calcular la constante y los coeficientes que mejor se ajusten a la expresión funcional de las variables. El método de regresión logística es discriminante por tal motivo es necesario contar con los insumos necesarios tales como: las variables predictoras, datos de presencia y pseudo-ausencia que ya fueron generados previamente.

3.5.3.1 Aplicación del Modelo

Para realizar los análisis estadísticos correspondientes en el programa SPSS 23 (versión de prueba) es necesario generar una matriz con todos los valores de las variables predictoras ya normalizadas en función de los puntos de ausencia y presencia. Esta información se obtiene mediante la herramienta *Extract Multi Values to Points* de ArcGis como se muestra en la Figura 38.

	pre_pres_1	bia_6a	bia_7a	bia_6b	bia_4a	ure_cdi1	bia_13a	bia_14a	bia_15a	bia_16a
1	1,00000	,354355	,283333	,845228	,136223	,137	,252227	,524282	,124000	,521133
2	,00000	,747728	,263000	,647619	,391478	,119	,797571	,903228	,040000	,774849
3	1,00000	,961939	,633000	,797714	,175137	,190	,356761	,216690	,296000	,807077
4	1,00000	,927920	,033000	,617967	,210591	,107	,248300	,069124	,570000	,261935
5	1,00000	,801932	,053000	,635228	,259187	,197	,311311	,081988	,190000	,318645
6	,00000	,381391	,753000	,023910	,310164	,119	,163368	,061988	,240000	,132537
7	,00000	,752793	,869567	,692476	,199275	,119	,627530	,603912	,216000	,663614
8	1,00000	,1300300	,003000	,702361	,306076	,107	,417064	,302777	,300000	,476677
9	,00000	,000000	,663000	,000000	,000000	,103	,210528	,073311	,256000	,213102
10	1,00000	,801932	,553333	,139819	,396203	,197	,020348	,223921	,418000	,892223
11	,00000	,000000	,100000	,000000	,332819	,190	,066590	,175116	,020000	,246697
12	,00000	,675576	,003000	,633962	,377619	,119	,486130	,781122	,024000	,760279
13	1,00000	,140312	,467000	,178671	,350610	,098	,108000	,161996	,304000	,512691
14	1,00000	,152739	,819967	,139819	,200000	,006	,867896	,601782	,021000	,191810
15	1,00000	,310311	,369967	,114286	,390000	,130	,451002	,052166	,534000	,530873
16	1,00000	,801932	,633333	,821429	,452697	,190	,100000	,173607	,644000	,130000
17	,00000	,946347	,115567	,603046	,242777	,119	,666190	,763667	,076000	,516975
18	,00000	,000000	,753000	,000000	,198775	,119	,103900	,082166	,070000	,162911

Figura 38. Matriz de las variables en SPSS

Fuente: (IBM SPSS, 2016)

El tipo de Regresión que se escogió para el presente estudio fue logística binaria, debido a que los valores de los insumos son dicotómicos es decir se encuentran dentro del rango [0,1] ajustándose de manera óptima este estadístico al modelo. Seguidamente se configuran cuáles son las variables dependientes e independientes (covariables) como se muestra en la Figura 39.

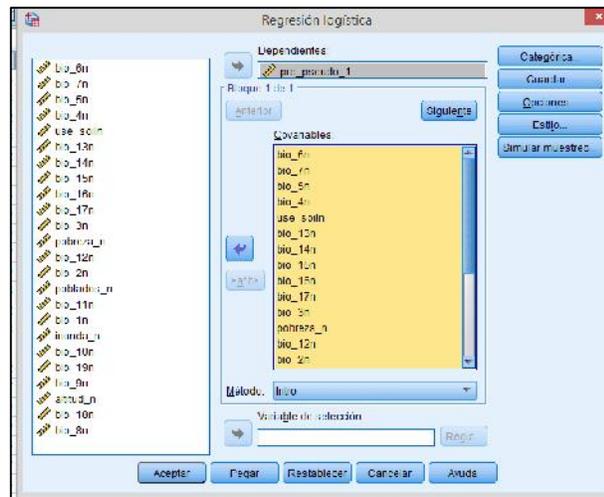


Figura 39. Configuración de Variables

Fuente: (IBM SPSS, 2016)

De la ejecución del programa se obtuvieron varios estadísticos que evalúan la confiabilidad de los resultados como: tablas de clasificación de las variables, pruebas de ómnibus, matriz de correlación y principalmente la tabla de coeficientes necesaria para generar el modelo $(\beta_0, \beta_1, \beta_2 \dots \beta_P)$, como se indica en la Figura 40.

	bio_6n	bio_7n	bio_5n	bio_4n	use_soiln	bio_13n
	2,312	1,062	-10,328	,530	1,357	-7,843
	bio_14n	bio_15n	bio_16n	bio_17n	bio_3n	pobreza_n
Constante	,919	4,253	7,521	1,017	,359	-,195
-4,089	bio_12n	bio_2n	poblados_n	bio_11n	bio_1n	inunda_n
	1,746	,281	,366	-,877	-,583	-,533
	bio_10n	bio_19n	bio_9n	altitud_n	bio_18n	bio_8n
	-1,205	-1,471	,489	,122	-1,393	11,341

Figura 40. Tabla coeficientes de las variables

De este modo, los valores de las constantes son multiplicadas por cada variable con la ayuda de la herramienta *Raster Calculator* de ArcGis, siguiendo la fórmula de regresión logística.

$$Y = \frac{1}{1 + (\text{Exp}(-1 * (\text{Constante} + \beta_1 * \text{Bio}_1 + \beta_2 * \text{Bio}_2 + \beta_3 * \text{Bio}_3 \dots + \beta_P * \text{Altitud})))}$$

Donde

Y es la probabilidad del modelo

3.5.3.2 Splines de Regresión Multivariada Adaptativa (MARS)

El programa MARS con sus siglas en inglés (Multivariate Adaptive Regression Splines) fue creado por la Universidad de Salford Inglaterra que adicional a este elemento cuenta con otros productos estadísticos tales como CART, TreeNet y Random Forests, todos enfocados en la elaboración de modelos predictivos y descriptivos para el análisis de bases de datos de cualquier tamaño y de diferente complejidad.

Como se mencionó en el capítulo anterior el método MARS plantea un análisis completo de las variables en función a la importancia que tiene cada una de ellas para la predicción del evento, ajustándose el modelo no solo a una curva predictora sino que más bien la divide en zonas (funciones base) mediante nodos o los también llamados puntos de inflexión, lo que mejora los resultados. Al igual que los modelos anteriores los insumos necesarios para aplicar el modelo son las variables predictoras, datos de presencia y pseudo-ausencia.

3.5.3.3 Aplicación del Modelo

Para la aplicación del programa MARS (versión de prueba) se requirió que el archivo de entrada se encuentre en formato delimitado por comas (.CSV) y adicionalmente contenga una matriz con todos los valores de las variables predictoras ya normalizadas en función de los puntos de ausencia y presencia que se obtuvieron a través de la herramienta *Extract Multi Values to Points* de ArcGis como se muestra en la Figura 41.

pre_pseudo	use_soulin	bic_3n	bic_4n	bic_5n	bic_5n	bic_7n	pcorre_n	bic_2n	pcblado_n	bic_1n	inmuda_n	bic_10n	bic_11n
1	0.167	0.956522	0.136525	0.815238	0.951955	0.253333	0.714681	0.476588	0.392168	1	0.02133	0.928571	1
0	0.119	1	0.051475	0.517619	0.747748	0.25	0	0.40595	0	0.727273	0.474551	0.983673	0.735032
1	0.19	0.869566	0.170137	0.780714	0.981982	0.15	0.829787	0.291118	0.277296	0.968636	0.01118	0.938776	0.988506
1	0.167	0.652144	0.218021	0.542857	0.927026	0.05	0.740426	0.117647	0.520102	0.886364	0.062145	0.846039	0.873563
1	0.167	0.608696	0.263457	0.595238	0.891892	0.05	0.574468	0.096019	1	0.840909	0.153072	0.806122	0.816092
0	0.119	0.956522	0.010154	0.02381	0.081081	0.75	0	0.960764	0	0.068182	0	0.061224	0.060966
0	0.119	0.869565	0.199375	0.890476	0.797793	0.355557	0.87734	0.505804	0	0.818182	0.378958	0.77551	0.804598
1	0.167	0.434783	0.226395	0.702381	1	0	0.659574	0	0	0.943182	0.088175	0.908163	0.908046
0	0.143	0.956522	0	0	0	0.55	0	0	0.520102	0	0.697825	0	0
1	0.167	0.565217	0.3352	0.797619	0.891892	0.333333	0.829787	0.352541	0.551592	0.931818	0.038307	0.908163	0.890552
0	0.19	0.956522	0.012858	0	0	1	0.93617	0	0.554592	0	0.81045	0	0
0	0.219	0	0.057858	0.530052	0.675676	0.5	0.574468	0.764706	0	0.738636	0.05051	0.574468	0.735632
1	0.048	1	0.05854	0.178571	0.747347	0.451111	0.778723	0.705667	0.712163	0.329545	0.235127	0.116327	0.133333
1	0.095	1	0.700538	0.797619	0.797793	0.515557	0.638298	0.77549	0.349163	0.875	0.047775	0.826531	0.83506
1	0.19	0.565217	0.366938	0.714296	0.810811	0.355557	0.787234	0.372549	0.520102	0.829545	0.716581	0.816527	0.793183
1	0.19	0.565217	0.452597	0.821429	0.801802	0.533333	0.170213	0.525412	0.520102	0.875	0.015118	0.877551	0.827586
0	0.119	0.913043	0.212377	0.869048	0.846847	0.515557	0	0.686275	0	0.943182	0.667572	0.877551	0.896552
0	0.219	1	0.155576	0	0	0.75	0.71168	0.980552	0	0	0.77801	0	0
1	0.19	0.608696	0.380756	0.833333	0.688045	0.453333	0.297872	0.505804	0.277296	0.683818	0.231429	0.893878	0.650172
0	0.19	0.652144	0.188585	0.761905	0.765786	0.155557	0.851064	0.5402	0.321817	0.829545	0.616038	0.826327	0.793183
1	0.143	0.869565	0.216575	0.833333	0.801802	0.55	0.723404	0.705667	0.777957	0.877727	0.151722	0.876735	0.867066

Figura 41. Matriz de las variables en MARS

Una vez obtenidos los insumos de entrada necesarios, se definieron cuáles son las variables dependientes e independientes, además del tipo de estadístico que se va a utilizar y el porcentaje de datos que se tomara como muestra para validar el modelo (10%), como se indica en la Figura 42.

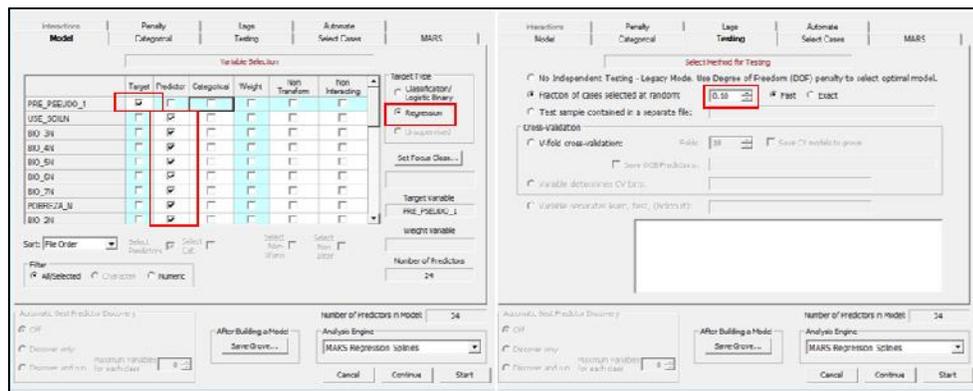


Figura 42. Configuración Programa MARS

Fuente: (MARS, 2016)

Además de la configuración anterior se estableció el número de “Funciones Base” que según varios autores recomiendan que sea dada de acuerdo al tamaño de la muestra. Para el presente estudio será de 15 funciones, el resto de opciones se mantienen por defecto como se indica en la Figura 43.

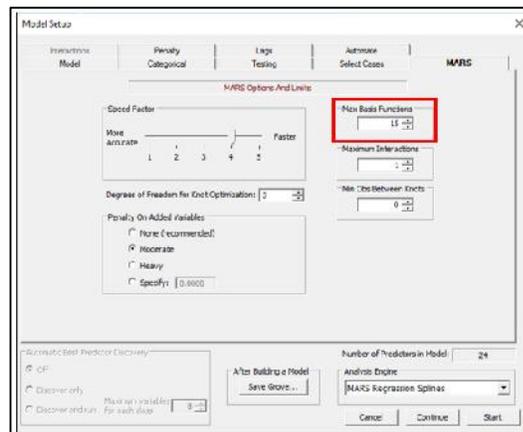


Figura 43. Definición de Funciones Base
Fuente: (MARS, 2016)

MARS genera un proceso interno de iteraciones llamado *forward* sobre las funciones bases definidas por las variables predictoras originales. Posteriormente el programa descartará las que menos se ajustan al modelo mediante otro proceso llamado *backward*, que convierte la variable X , original en una nueva variables definida como: $\max(0, X-c)$ ó $\max(c-X)$, donde c es el umbral establecido por los nodos, como se muestra a continuación.

$$BF2 = \max(0, 0.416 - [\text{bio}_{15n}])$$

$$BF3 = \max(0, [\text{inunda}_n] - 0.015118)$$

$$BF4 = \max(0, 0.015118 - [\text{inunda}_n])$$

$$BF5 = \max(0, [\text{pobreza}_n] - 0.723404)$$

$$BF6 = \max(0, 0.723404 - [\text{pobreza}_n])$$

$$BF7 = \max(0, [\text{poblados}_n] - 0.392163)$$

$$BF8 = \max(0, 0.392163 - [\text{poblados}_n])$$

$$BF9 = \max(0, [\text{bio}_{8n}] + 5.96046e-008)$$

$$BF11 = \max(0, 0.72619 - [\text{bio}_{5n}])$$

$$BF12 = \max(0, [\text{bio}_{12n}] - 0.863209)$$

$$BF14 = \max(0, [\text{bio}_{2n}] - 0.764706)$$

Finalmente MARS crea una ecuación predictora final que está definida por las funciones bases generadas, que a su vez son multiplicadas por los coeficientes que más se ajustan al modelo, mediante el uso de la herramienta *Raster Calculator* de ArcGis.

$$Y = -0.6306 * BF1 - 1.29024 * BF2 - 0.26154 * BF3 - 18.1487 * BF4 - 1.46987 * BF5 - 0.283358 * BF6 - 0.480713 * BF7 - 0.583341 * BF8 + 2.07942 * BF9 + 2.0564 * BF11 + 3.74046 * BF12 + 1.33308 * BF14$$

Donde

Y es la probabilidad del modelo

Se aplicaron los métodos de Maxent, Regresión Logística, Lógica Difusa y MARS con el fin de determinar las zonas con mayor probabilidad de presencia del vector dentro del área de estudio (Ecuador Continental); en esta fase del estudio se detalla cuáles deben ser los insumos necesarios, el software y el proceso que se debe tomar en cuenta para realizar cada modelamiento.

3.6 Desviación Estándar y Ajuste de los Modelos

3.6.1 Desviación Estándar de los modelos

La desviación estándar es un conjunto de datos o una medida de dispersión que indica cuanto pueden alejarse los valores obtenidos con respecto al promedio (media), es decir la importancia que tiene este estadístico se basa en la probabilidad de que un evento suceda o no. Los valores de las desviaciones estándar de los modelos aplicados se describen en la Tabla 8 además de sus valores máximos, mínimos y medias aritméticas.

Tabla 8.
Desviación Estándar de los Modelos

Modelos	Mínimo	Máximo	Media	Desviación Estándar
Maxent	0	0.970063	0.104460	0.173263
Lógica Difusa	0	0.844170	0.104500	0.221607
Regresión Logística	0	0.999995	0.347007	0.302110

Mars	0	1	0.361043	0.105743
-------------	---	---	----------	----------

3.6.2 Ajuste de los modelos

El ajuste se realizó sobre los modelos finales previamente normalizados dentro de un rango de [0,1], en función a la desviación estándar; aplicando la siguiente fórmula:

$$N = \text{Valor Medido} - \text{Valor Calculado} \quad (25)$$

Donde:

N es el valor ajuste, el Valor Medido es el valor máximo al que podría llegar los modelos, es decir 1 (probabilidad de presencia), y el Valor Calculado es el valor de la desviación estándar de los promedios de probabilidades de los diferentes modelos.

La Tabla 9 muestra las desviaciones estándar de los modelos y el ajuste que se realizó a cada uno de ellos.

Tabla 9.
Ajuste de los Modelos

Modelos	Maxent	Lógica Difusa	Regresión Logística	MARS
Media	0.107683	0.123790	0.347000	0.361043
Desviación Estándar	0.178610	0.262514	0.347000	0.105743
Ajuste	0.821390	0.737486	0.653000	0.894257

Cada uno de los métodos aplicados determina ciertos parámetros estadísticos de los productos resultantes (raster), como son: la desviación estándar, valores máximos, mínimos y valores de la media, que son determinantes en la fase de evaluación de los modelos.

CAPITULO IV

RESULTADOS Y DISCUSIONES

4.1

4.1 Resumen

En el presente capítulo se describen los resultados estadísticos que arrojaron cada uno de los modelos, siendo Maxent el método con mayor rendimiento predictivo con un valor de $AUC=0,998$. Seguidamente el Modelo de MARS con un valor de $AUC=0.996$ y finalmente con un valor similar los modelos de Lógica Difusa y Regresión logística de $AUC=0.986$, con un ajuste aceptable solo superado por el método de MARS y además visualmente describe las zonas con mayor probabilidad del vector asemejándose a la distribución real de la especie en función a los datos de presencia y las variables predictoras; tras el análisis anterior se determinó que, en gran parte del territorio de la región litoral probabilísticamente existe presencia del vector especialmente en las provincias de El Oro y Manabí con valores de 23.22% y 17.96% respectivamente siendo las más vulnerables ante un posible riesgo epidemiológico.

4.2 Evaluación de los modelos

Para seleccionar el modelo con un mayor rendimiento predictivo de la distribución espacial del vector *Aedes aegypti* dentro del Ecuador continental, se realizaron varios análisis y comparaciones tanto estadísticas como gráficas de los cuatro modelos aplicados.

4.2.1 Análisis del ajuste de los modelos

Mediante el análisis del ajuste de los modelos previamente normalizados (ver Tabla 10) se observó que el método de MARS con un ajuste de 0.89425 es el que más se acerca al valor de uno, es decir existe una baja dispersión de sus datos con respecto a la media lo cual significa que la predicción es eficaz; seguidamente se encuentran los métodos de Maxent y Lógica Difusa con valores ajustados dentro de un rango aun aceptable de 0.821390 y 0.737486 respectivamente; por último se

encuentra el método de regresión logística con el valor ajustado de 0.653000 lo que describe un bajo rendimiento predictivo.

Tabla 10.
Análisis del Ajuste de los Modelos

Modelo	Maxent	Lógica Difusa	Regresión Logística	MARS
Desviación Estándar General	0.178610	0.262514	0.347000	0.105743
Desviación Estándar Normalizada	0.173263	0.221607	0.302110	0.105743
Diferencia	0.005347	0.040907	0.044890	0
Ajuste	0.821390	0.737486	0.653000	0.894257

4.2.2 Análisis de la Curva AUC de los modelos

La capacidad predictiva de los modelos se la validó mediante el análisis estadístico AUC (Area Under the Curve) a partir de la curva ROC (Receiver Operating Characteristics), cuya función principal es calcular la sensibilidad y especificidad de los valores de las ocurrencias de la especie al intersecar las presencias con las capas (raster) de cada uno de los modelos obtenidos. La curva ROC se fundamenta en la unión de distintos puntos de corte, correspondientes en el eje Y a la “sensibilidad” y en el eje X a la “1-especificidad”. Los dos ejes contienen valores entre 0 y 1 (0% a 100%); el intervalo de confianza que se utilizó para el análisis fue del 95%.

- **Modelo Maxent**

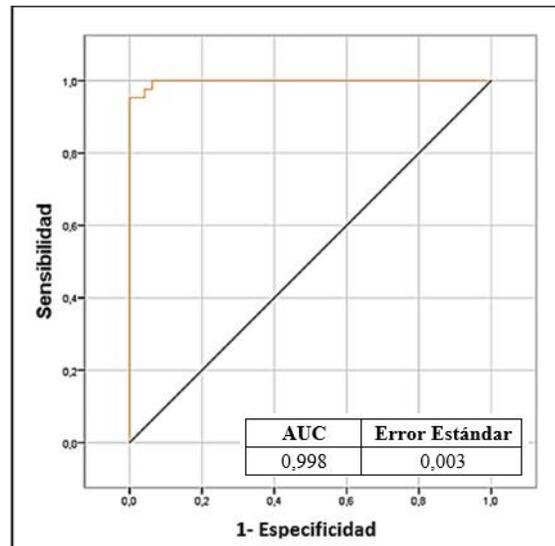


Figura 44 Análisis de la Curva ROC - Modelo Maxent

- Modelo Lógica Difusa

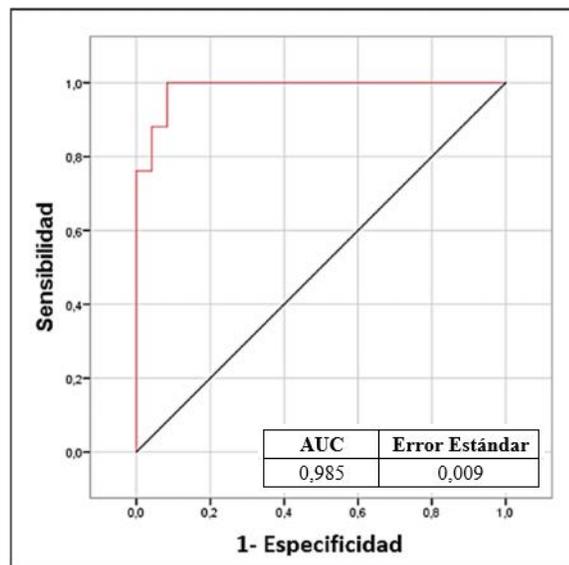


Figura 45 Análisis de la Curva ROC - Modelo Lógica Difusa

- **Modelo Regresión Logística**

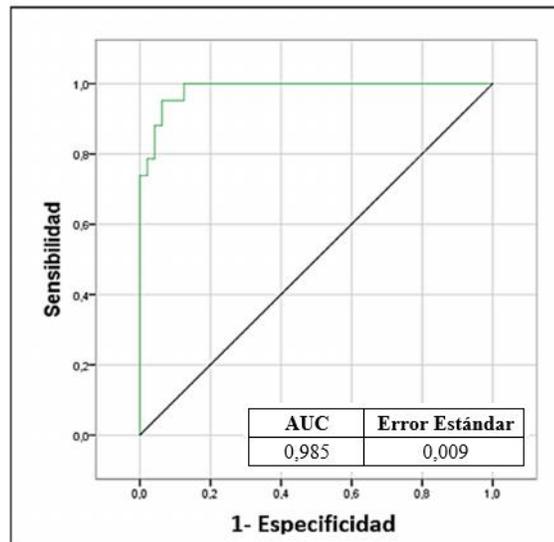


Figura 46. Análisis de la Curva ROC - Modelo Regresión Logística

- **Modelo MARS**

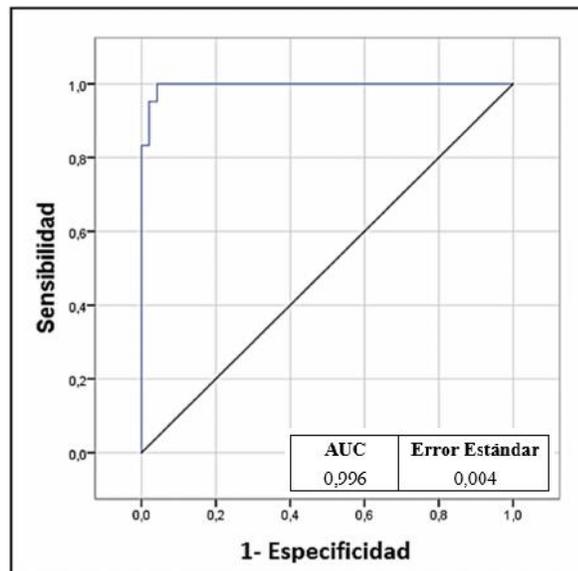


Figura 47. Análisis de la Curva ROC - Modelo MARS

Las gráficas de la curvas ROC se superpusieron en la Figura 48, mostrando resultados predictivos satisfactorios $> 0,9$ el Área Bajo la Curva (AUC) en todos los

modelos, lo que significa que dentro del entorno de la matriz de confusión existen predicciones con verdaderas Presencias y verdaderas Ausencias.

Sin embargo el modelo que más se acerca a 1 es el Modelo de Maxent (ver Figura 44) con un valor de AUC=0,998. Seguidamente el Modelo de MARS (ver Figura 47) con un valor de AUC= 0.996 y finalmente con un valor similar los modelos de Lógica difusa (ver Figura 45) y Regresión logística (ver Figura 46) de AUC= 0.986.

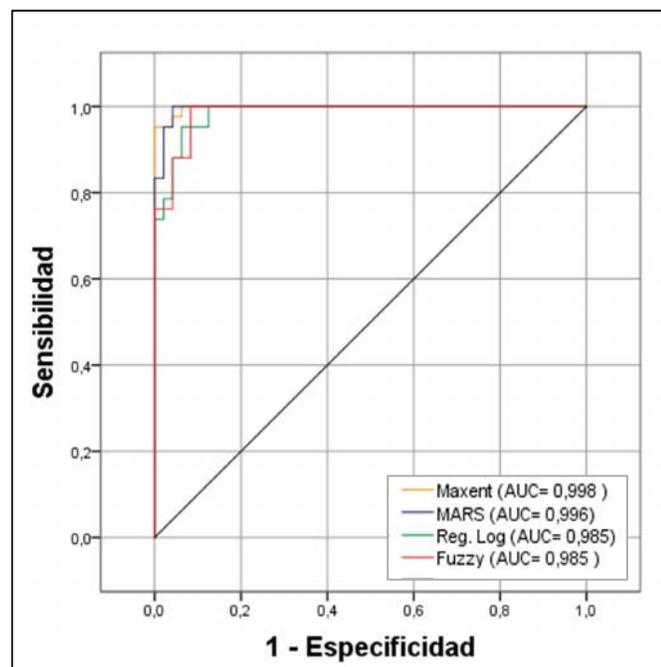


Figura 48. Análisis de la Curva ROC - General

4.2.3 Análisis Gráfico de los modelos

Como resultado de la comparación de los modelos utilizados para la predicción de la distribución espacial del vector *Aedes aegypti*, se evidenció que tanto los modelos de Maxent (ver Figura 49) como los de Lógica Difusa (ver Figura 50) se ajustan a la distribución real del vector, debido a que las zonas con mayor probabilidad predicha por estos modelos se encuentran cercanos a los puntos de presencia. Adicionalmente estas zonas deben cumplir estas zonas con parámetros biológicos primordiales para la supervivencia del vector tales como bajas altitudes no mayores a 1600 metros sobre el nivel del mar y distancias a centros poblados que no

excedan los 4000 metros ya que las hembras necesitan disponer de suficiente sangre como fuente de proteínas para multiplicarse (para la producción de huevos).

Los modelos de Regresión Logística (ver Figura 51) y MARS (ver Figura 52) predicen una probabilidad alta de presencia del vector en casi toda la región litoral sin discriminar zonas en las que podrían existir ausencias del mismo debido a factores climáticos, topográficos, etc.

- **Modelo Maxent**

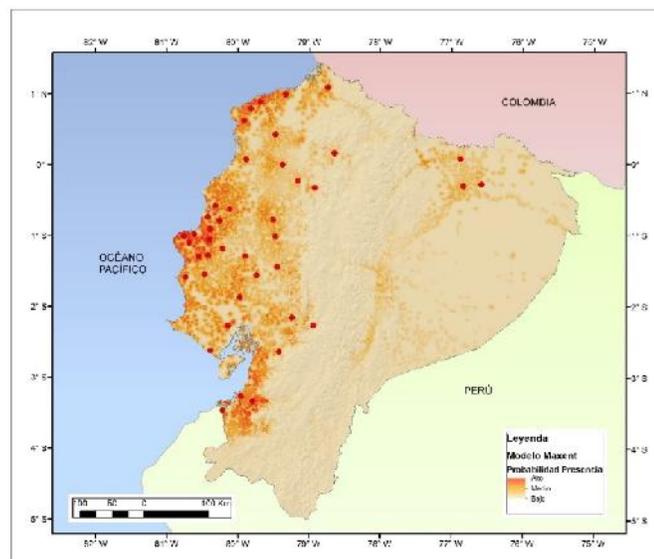


Figura 49. Distribución del *Aedes aegypti* (Modelo Maxent)

- **Modelo Lógica Difusa**

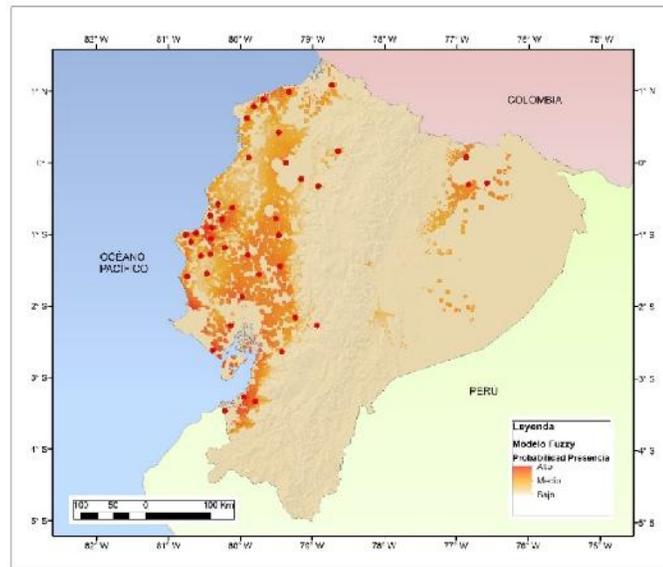


Figura 50. Distribución del *Aedes aegypti* (Modelo Lógica Difusa)

- **Modelo Regresión Logística**

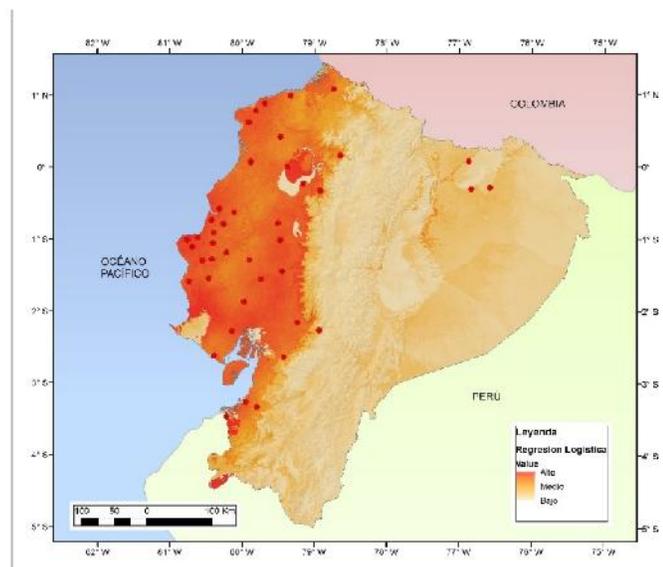


Figura 51. Distribución del *Aedes aegypti* (Modelo Regresión Logística)

- **Modelo MARS**

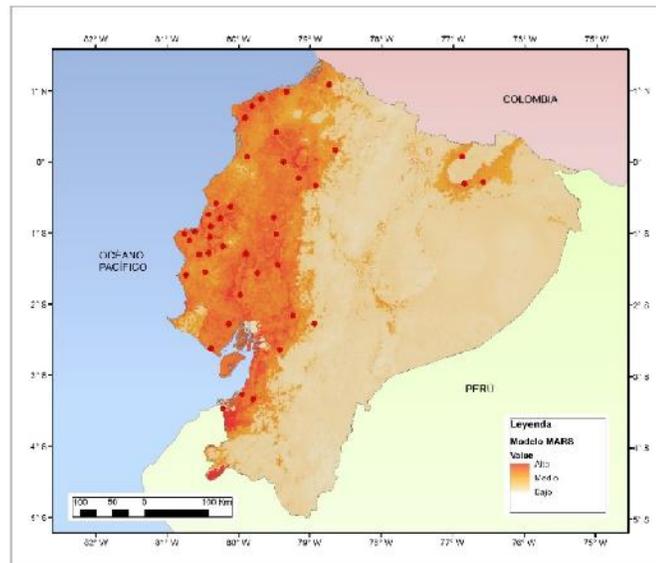


Figura 52. Distribución del *Aedes aegypti* (Modelo MARS)

4.2.4 Selección del Mejor Modelo

Después de haber realizado varios análisis y comparaciones tanto estadísticas como gráficas de los cuatro modelos aplicados en la presente investigación, el modelo Maxent cuenta con un mejor rendimiento predictivo de la distribución espacial del vector *Aedes aegypti*, ya que representa un desempeño satisfactorio en el análisis de curva ROC, con un valor AUC=0.998 (ver Figura 44), y con un ajuste de la desviación estándar 0.821390 (ver Tabla 10), solo por debajo del modelo de MARS. Además, visualmente describe las zonas con mayor probabilidad del vector, las cuales se asemejan convenientemente a la realidad en función a los datos de presencia y las variables predictoras.

Maxent fue seleccionado como el modelo con mejor rendimiento predictivo tras ciertos análisis estadísticos como el ajuste de la desviación estándar, los resultados de la curva AUC y comparaciones gráficas con los otros modelos aplicados.

4.3 Delimitación de zonas idóneas para la presencia del vector

La estimación de las zonas con mayor probabilidad de presencia del vector *Aedes aegypti* se determinó mediante el análisis de su frecuencia dentro del área de estudio, dando como resultado el histograma de la Figura 53. En él se observa que el menor número de frecuencias está en un rango de (0-30%) con apenas 4 presencias, una frecuencia intermedia con un valor de 18 presencias en un rango de (30-60%) y la mitad de los datos de presencias con un valor de 22 en el rango de (60-100%).

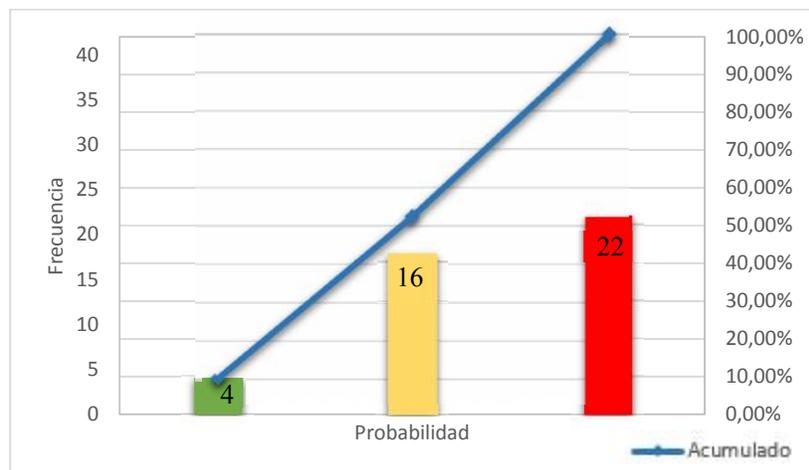


Figura 53. Histograma de Probabilidades

La Figura 54 muestra gráficamente el rango de probabilidades que se determinó anteriormente, lo cual evidencia claramente que la región litoral es la más apta para que exista la presencia del vector.

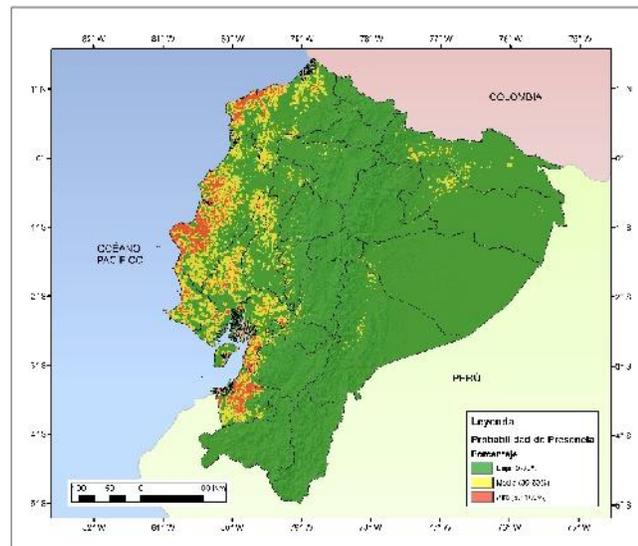


Figura 54. Probabilidades de Presencia del vector *Aedes aegypti*

De acuerdo con el análisis de frecuencia del vector dentro del área de estudio que se realizó anteriormente se estimó que las zonas con mayor probabilidad de presencia del vector se encuentran dentro del rango (60-100%) que es proporcional al riesgo epidemiológico. Posteriormente se reclasificó únicamente las zonas con probabilidad alta de presencia del vector, como se observa en la Figura 55.

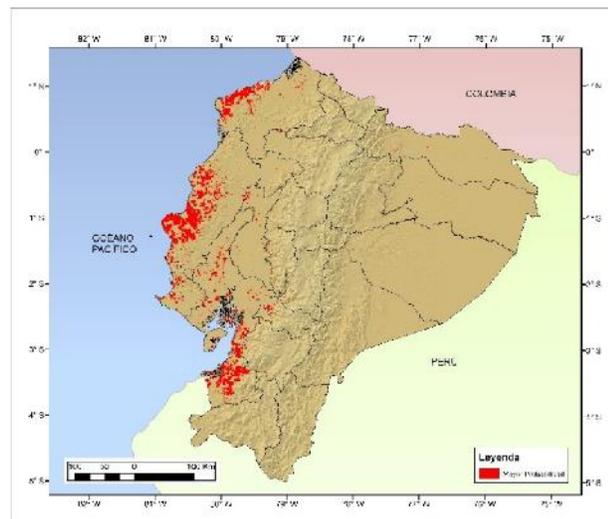


Figura 55. Zonas idóneas para la Presencia del vector *Aedes aegypti*

Las zonas idóneas para la presencia del vector *Aedes aegypti* cubren un área de 7806 Km², es decir el 3,15% del territorio del Ecuador Continental. Dicha área se

encuentra repartida en 16 provincias (ver Tabla 11) y a su vez en 111 cantones. Principalmente existe una presencia alta en la región Costa en localidades importantes tales como Guayaquil, Machala, Babahoyo, Portoviejo, Salinas, etc., una presencia baja en la región Amazónica en los cantones Lago agrio, La joya de los Sachas, Mera, etc. y ausencia casi en su totalidad en la región Sierra (ver Anexo 1).

Tabla 11.
Provincias con mayor riesgo Epidemiológico

Provincia	Área km ²	Porcentaje (%)
El Oro	1334	23.22
Manabí	3402	17.96
Santa Elena	331	8.94
Esmeraldas	1301	8.06
Guayas	1103	7.43
Los Ríos	112	1.46
Azuay	102	1.22
Bolívar	31	0.79
Cañar	15	0.47
Santo Domingo de los Tsáchilas	8	0.23
Sucumbíos	37	0.21
Chimborazo	12	0.18
Pichincha	4	0.04
Orellana	6	Continúa →
Morona Santiago	5	
Pastaza	3	0.01
Cotopaxi	0	0
Imbabura	0	0
Carchi	0	0
Loja	0	0
Napo	0	0
Tungurahua	0	0
Zamora Chinchipe	0	0

Se determinó que la probabilidad de presencia del vector *Aedes aegypti* existe en 111 cantones pertenecientes a 16 provincias siendo El Oro y Manabí con valores de 23.22% y 17.96% respectivamente los más vulnerables ante un posible riesgo epidemiológico.

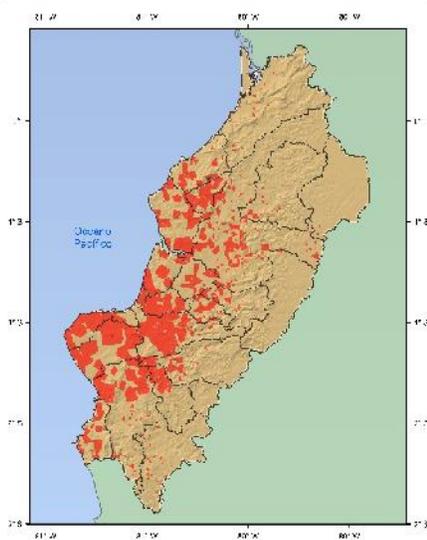
4.4 Caracterización de las zonas con mayor riesgo epidemiológico

Existen varias zonas a lo largo de la costa ecuatoriana que pueden tomarse para realizar la caracterización de riesgo epidemiológico, en ellas, la probabilidad de presencia del vector es muy alta; sin embargo se consideraron las provincias de Manabí (ver Tabla 12) y El Oro (ver Tabla 13) para realizar dicho estudio, debido a que más del 15% de sus áreas territoriales están expuestas a una posible distribución espacial del vector.

La provincia de Manabí cuenta con 22 cantones, de los cuales 21 determinan una posible presencia del vector. Principalmente en las localidades de Rocafuerte, Montecristi y Manta se evidencia una distribución espacial del vector en casi todo su territorio, mientras que la provincia de El Oro cuenta con una expansión territorial en menor escala a comparación de Manabí, con 14 cantones de los cuales todos muestran presencia del vector siendo Pasaje y Machala los más afectados (ver Anexo 1).

Tabla 12.
Caracterización de la Provincia de Manabí

Provincia:	Manabí
Capital:	Portoviejo
Área:	18.400 km ²
Nº Habitantes:	1'369.780
Nº Cantones:	22
Precipitación:	500 a 1000 mm al año
Altitud promedio:	6 a 350 m.s.n.m
Índice de Pobreza:	39,8 %
Temp. Promedio:	25 °C



Clima: Subtropical seco a Tropical húmedo (SNI, 2011).

Cantones con Probabilidad del Vector

Jaramijó (29.87%), Junín (25.94%), Rocafuerte (67.59%), Manta (74.86%), Tosagua (32.27%), Puerto López (27.70%), 24 De Mayo (20.20%), Jama (19.09%), Sucre (47.97%), San Vicente (36.40%), Montecristi (62.31%), Portoviejo (48.86%), Santa Ana (9.67%), Pichincha (0.35%), Paján (1.50%), El Carmen (0.13%), Flavio Alfaro (1.87%), Jipijapa (25.09%), Pedernales (0.98%), Chone (11.10%) y Bolívar (5.76%)

Geografía de la Provincia

Manabí tiene aproximadamente 350 km de costa marítima, con importantes accidentes geográficos como la bahía de Cojimíes, punta Jama, etc. Forma parte de la cordillera de Chongón-Colonche y la Cordillera de Balzar, por donde cruza el río Chone y río Portoviejo que son los únicos con cauce profundo; esto determina que ciertas zonas de la provincia estén predispuestas a inundaciones en las temporadas invernales de mayor precipitación (SNI, 2011).

Uso de suelo

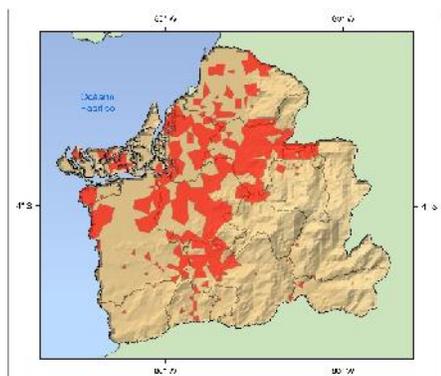
Existe un predominio de pastos cultivados 51,3% que representan poco más de la mitad de la superficie provincial utilizada. Los montes y bosques 21,5% y los cultivos permanentes 13, 2%) sumados a las zonas de pastizales dejan ver la existencia de zonas protegidas así como zonas aptas para la ganadería (INEC, 2012).

Asentamientos Humanos

Los más relevantes son las ciudades de Portoviejo y Manta, no obstante, poseen índices altos de déficit en servicios básicos residenciales (agua, eliminación de aguas servidas, suministro eléctrico) con el 80% y 61,20% respectivamente (SNI, 2011).

Tabla 13.
Caracterización de la Provincia de El Oro

Provincia:	El Oro
-------------------	--------



Capital: Machala

Área: 5767 km²

N° Habitantes: 648,316

N° Cantones: 14

Precipitación: 200 a 1.500 mm al año

Altitud: 6 a 3580 msnm

Índice de Pobreza: 23,4 %

Temp. Promedio: 25 °C

Clima: Seco Costanero,
Tropical Sabana,
Tropical Mozón
(SNI, 2014).

Cantones con Probabilidad del Vector

Huaquillas (40.15%), Balsas (0.99%), Marcabelí (8.02%), Atahualpa (7.17 %), Portovelo (1.10%), Las Lajas (0.51%), Machala (55.39%), Chilla (13.89%), Pasaje (57.11%), El Guabo (23.31%), Piñas (19.55%), Zaruma (0.70%), Arenillas (21.00%) y Santa Rosa (43.23%)

Geografía de la Provincia

Esta provincia está dividida en dos áreas: hacia el noroeste, piedemonte que baja hacia el Golfo de Guayaquil se encuentran las llanuras, donde se cultiva banano; y el sureste zona montañosa que está atravesada por la Cordillera Occidental de los Andes, y la temperatura va descendiendo de acuerdo a la altura (SNI, 2014).

Uso de suelo

La superficie de suelo que se ocupa en la actividad agropecuaria es de 457.025 ha, está distribuida de la siguiente manera: 2,17% de cultivos transitorios; 53,56% pastos (cultivables y naturales); 12.36% bosques; 13.45% otros usos, descanso y páramos. También existen yacimientos para la explotación minera tales como el oro, feldespatos y materiales pétreos que se utilizan en la construcción (INEC, 2012).

Asentamientos Humanos

Según el último censo, se determinó que el 41% de la población de la provincia se concentra en el cantón Machala. Además, que de la población total de la provincia, el 79% se encuentra concentrado en la zona urbana y el 21% en la zona rural (SNI, 2014).

Estas provincias cuentan con una marcada problemática ambiental y social donde la contaminación y los altos índices demográficos y de pobreza son los más

representativos. Además, por ubicación geográfica son más vulnerables a desastres naturales como tsunamis, inundaciones y sismos.

Las provincias de Manabí y El Oro por su situación geográfica, climática y además de contar con una marcada problemática ambiental y social donde la contaminación y los altos índices demográficos, establecen sitios favorables con mayor probabilidad de presencia del vector *Aedes aegypti*.

CAPITULO V

CONCLUSIONES Y RECOMENDACIONES

5.1

5.1 Resumen

Dentro del presente estudio se aplicó varios métodos para el modelamiento de especies, con el fin de determinar las zonas con mayor probabilidad de presencia del vector *Aedes aegypti*, las cuales son referentes para la planificación en el control y prevención ante un posible riesgo epidemiológico en el país.

5.2 Conclusiones

- Maxent se estableció como el modelo con mayor rendimiento predictivo tras varias validaciones estadísticas (ver Figura 48), además visualmente describe las zonas con mayor probabilidad de presencia del vector (ver Figura 49), que adicionalmente se asemejan a la distribución real del vector (ver Figura 21) en función de los datos de presencia y las variables predictoras.
- La Metodología de Lógica Difusa o Fuzzy determina predicciones bastante aceptables con un rendimiento predictivo escasamente inferior al método de Maxent (ver Tabla 10) el cual construye su modelo en base a la generación de *background* por la necesidad que requiere su algoritmo, provocando algunas veces un sobre ajuste del modelo. A diferencia de Fuzzy que realiza sus predicciones en función a las variables predictoras, datos de presencia y de pseudo-ausencia, y el cual mejoría si se incluirá al modelo datos de ausencia muestreados en campo.
- Los métodos de regresión logística y Splines de Regresión Multivariada Adaptativa (MARS) registraron valores satisfactorios en los análisis estadísticos (ver Tabla 10), sin embargo gráficamente registran una alta probabilidad del vector en casi toda la región litoral sin discriminar zonas en las que podrían existir ausencias del mismo (ver Figura 51 y Figura 52), debido a factores climáticos, topográficos, etc.

- Las zonas con mayor probabilidad de presencia del vector *Aedes aegypti* cubren de manera general un área de 7806 Km² repartidas en 16 provincias dentro de 111 cantones (ver Anexo 1). Principalmente existe una presencia alta en la región litoral en localidades importantes como Guayaquil, Machala, Babahoyo, Portoviejo, Salinas, etc.; una presencia baja en la región Amazónica en los cantones Lago agrio, La joya de los Sachas, Mera, etc. y ausencia casi en su totalidad en la región Sierra.

Las provincias de Manabí (ver Tabla 12) y Los Ríos (ver Tabla 13

-) por sus características geográficas, climáticas e incluso socio-económicas tales como altitud, temperatura, precipitación y entornos, asociadas a las condiciones de vida humana, establecen sitios favorables con mayor probabilidad de presencia del vector *Aedes aegypti*.
- El modelo de distribución espacial del *Aedes Aegypti* permitió la caracterización de las zonas con mayor probabilidad de presencia del vector (ver Figura 55); de esta manera se puede definir la dinámica ecológica de transmisión de ZIKA, la evaluación del impacto epidemiológico-económico y las estrategias de intervención que se deben tomar ante un posible riesgo epidemiológico en el país.

5.3 Recomendaciones

- Es determinante contar con registros de ocurrencia de la especie obtenidos en base a investigaciones en campo y debidamente georreferenciadas además de variables predictoras actualizadas para de esa forma obtener una predicción más precisa de las zonas con mayor probabilidad de presencia de la especie.
- Para mejorar la calidad de los resultados es indispensable realizar una evaluación previa de la aportación así como la correlación que tienen cada una de las variables dentro del modelo para incluir o prescindir de una de ellas.

- Es pertinente realizar una comprobación en campo para verificar la fiabilidad del modelo, debido a que la mayoría de los insumos aplicados fueron generados a través de información indirecta como lo son los datos de presencias y pseudo-ausencias.
- Debido a que esta especie se adapta fácilmente y coloniza nuevos nichos ecológicos es necesario incluir a la presente investigación estudios que evalúen índices entomológicos, tasa de picadura a humanos, estimación de índices de infección para de esta manera mejorar la predicción de las áreas con mayor riesgo epidemiológico.

BIBLIOGRAFÍA

- Alvarado, & Serrano. (2013). Ecological niche models in phylogeographic studies: applications, advances and precautions. *Mol Ecol Resour*, 14 (2): 233-248.
- Avilés, M., & Lizzete, H. (2009). *Análisis y Modelamiento de Susceptibilidad a deslizamientos mediante un SIG*. Sangolquí: Tesis Universidad de las Fuerzas Armadas ESPE.
- Barba, & Romero. (1987). Panorámica actual de la decisión multicriterio discreta. *Investigaciones Económicas*, 279-308.
- Barba-Romero, S., & Pomerol, J. (1997). *Decisiones Multicriterio. Fundamentos Teóricos y Utilización Práctica*. Alcalá de Henares, España: Servicio de Publicaciones.
- Bates, M. (1970). The natural history of mosquitoes. *Gloucester, Mass. USA*, 378.
- Carrada, T., & Vázquez, L. (2012). La ecología del dengue y el *Aedes aegypti*. *ercera parte. Salud Pública*, 297-311.
- Carranza, D. (2008). Aplicación de la Lógica Difusa para la ubicación de especies faunísticas y florísticas, y su comparación con otros métodos geoestadísticos. *Tesis de Grado Universidad de las Fuerzas Armadas Espe*.
- CONAVE, C. N. (2015). *Aviso Epidemiológico*. Mexico: SEMAR.
- Consoli, R., & De Oliveira, R. (1994). *Principais mosquitos de importância sanitária no Brasil*. Río de Janeiro: Fiocruz.
- Cruz, A., & Camargo, B. (2012). Glosario de términos en Parasitología y Ciencias Afines. *Instituto de Biología UNAM*, 162-165.
- Espinal, M. (22 de Julio de 2016). Pobreza y cambio climático aumentan riesgo de dengue y zika, dice OPS. (R. L. Vanguardia, Entrevistador)
- Espíndola, M. (2013). *Modelamiento del Hábitat Potencial Actual y Futuro del Zamarrillo Pechinegro *Eriocnemis nigrivestis* (Bourcier & Mulsant, 1852) en la Sierra Norte del Ecuador*. Sangolquí: Universidad de las Fuerzas Armadas ESPE.
- Franklin, J. (2009). Mapping Species Distributions: Spatial Inference and Prediction. *Cambridge*.
- Friedman, J. (1991). Multivariate Adaptive Regression Splines (with discussion). *The Annals of Statistics*, 1-141.

- Graham, C. (2012). Modelos de distribución de las especies y el desafío de pronosticar distribuciones futuras. *Cambio Climático y la Biodiversidad de los Andes Tropicales*, 147-186.
- Gómez, D., & Rodríguez, M. (2013). El Dengue y la Fiebre Hemorrágica del Dengue: Clínica y Epidemiología. *PAC Asociación Mexicana de Infectología y Microbiología Clínica.*, 7-32.
- Gómez, M., & Barredo, J. (2005). *Sistemas de Información Geográfica y la Evaluación Multicriterio*. Madrid: RA-MA.
- Guisan, A., & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecol Let*, 8(9): 993-1009.
- Hall, L. (1997). The habitat concept and a plea for standard terminology. *Wildlife Soc. B*, 173-182.
- Harwood, M., & James, M. (1987). Entomología Médica y Veterinaria. *Limusa*, 201-203.
- Hay, S. (2006). Global environmental data for mapping infectious disease distribution. *Advance in Parasitology*, 33-77.
- Hutchinson, G. (1957). Concluding remarks, Cold Spring Harbor. *Symposium of Quantitative Biology*, 415-427.
- INAMHI. (2015). *Anuario Meteorológico*. Quito: Instituto Nacional de Meteorología e Hidrología.
- INEC. (2012). *Encuesta de Superficie y Producción Agropecuaria Continua 2012*. Quito: Instituto Nacional de Estadística y Censos.
- Jaramillo, C. (2014). *Determinación De La Probabilidad De Ocurrencia De Deslizamientos Aplicando Lógica Fuzzy Vs Mora Vahrson, En El Área Monjas – Itchimbia – Panecillo, Del Distrito Metropolitano De Quito, (DMQ) Y Simulación En 3D*. Sangolquí: Tesis Universidad de las Fuerzas Armadas ESPE.
- Marquez, H. (1999). *Métodos matemáticos de evaluación de factores de riesgo para el patrimonio arqueológico: una aplicación GIS del método de jerarquías analíticas de T.L. Saaty*. Sevilla: SPAL8.
- Martínez, N. (2012). El nicho ecológico: útil concepto aún en debate. *Cienciorama*, 5-10.
- Miller, J. (2010). Species Distribution Modeling. *Geography Compss*, No.46, 490-509.

- Molero, E., & Rodriguez, A. (2007). Escenarios de Aptitud y Modelización Cartográfica del Crecimiento Urbano Mediante Técnicas de Evaluación Multicriterio. *GeoFocus*.
- Mora, P. (2014). *1er Congreso Nacional de Geomática*. Obtenido de Mediciones Científicas e Industriales C.A.: <http://www.mecinca.net/Presentaciones/DronesFotoAutoNube.pdf>
- Morales, J. (21 de septiembre de 2015). *Equilibrio Ecológico*. Obtenido de Hábitat y Nicho Ecológico: <http://equilibrioecologicofisica.blogspot.com>
- Moreno, J. (2001). *Proceso Analítico Jerárquico (AHP): Fundamentos, Metodología y Aplicaciones*. Valencia: Universidad de Valencia.
- MSP. (2015). *Alerta Epidemiológica: Ante La Posibilidad De Introducción Del Virus Zika En Ecuador*. Quito: Ministerio de Salud Pública.
- Muñoz, J., & Felicísimo, A. (2012). Modelos de distribución de especies: Una revisión sintética. *Universidad de Castilla-La Mancha, España*, 4-8.
- OMS, O. M. (2015). *Reporte Mundial 2014 sobre el Paludismo*. Geneva - Suiza: Servicio de Producción de Documentos de la OMS.
- Ortiz, F. (2002). Normas y Conormas de la Lógica Difusa. *Universidad Nacional de Cajamarca*.
- Padilla, O. (2006). Fundamento teórico para la modelización de variables ambientales mediante operadores difusos. *Revista Geoespacial No. 4*.
- Pahari, K., & Murai, S. (1999). Modelling for prediction of global deforestation based on the growth of human population. *ISPRS Journal of Photogrammetry & Remote Sensing*, 54: 317-324.
- Pearson, R. (2008). Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography* (27), 285-298.
- Peterson, M. (2003). Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful. *Global Ecology and Biogeography*, 12(5): 361-371.
- Phillips, S., & Dudík, M. (2004). A maximum entropy approach to species distribution modeling. *Proceedings of the Twenty-First International Conference on Machine Learning*, 655-662.
- PNBV. (2013). Plan Nacional para el Buen Vivir 2013-2017. *Secretaría Nacional de Planificación y Desarrollo*, 147.

- Rotela. (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.
- Saaty, T. (1980). *Fundamentals of Decision Making and Priority Theory*. McGrawHill.
- Saaty, T. (1980). *The Analytic Hierarchy Process*. New York: McGraw-Hill.
- Saavedra, A., & Vinuesa, I. (2013). *Implementación de un plan piloto para la Gestión de riesgos para el fortalecimiento de los miembros de los comandos operacionales, en la jurisdicción de la ciudad de Esmeraldas*. Sangolquí: Universidad de las Fuerzas Armadas ESPE.
- Sainz de la Maza, M. (2000). Delimitación de Hábitat Potencial del Urogallo (Tetrao Urogallus) mediante Lógica Borrosa (Fuzzy Logic) en el Parque Natural del Cadí-Moixeró (Pirineo oriental). *SIG y Modelización de Fauna Protegida*.
- Sastre, P. (2010). *Sistemas de Información Geográfica: Técnicas básicas para estudios de biodiversidad*. Madrid: Instituto Geológico y Minero de España.
- Sillero, N., Barbosa, M., & Martínez, F. (2010). Los modelos de nicho ecológico en la herpetología ibérica: pasado, presente y futuro. *CICGE, Centro de Investigação em Ciências Geo-Espaciais. Universidade do Porto. Faculdade de Ciências.*, 2-5.
- SNI. (2011). *Plan de Ordenamiento Territorial del Gobierno Provincial de Manabí*. Quito: Sistema Nacional de Información.
- SNI. (2014). *Plan de Ordenamiento Territorial del Gobierno Provincial de El Oro*. Quito: Sistema Nacional de Información.
- Soberón, J., & Peterson, A. (2005). Interpretation of models of fundamental ecological niches and species distributional areas. *Biodiversity Informatics* , 1-10.
- Soberón, N., & Hensz, A. (2012). Ecological Modelling. *Elsevier*, 10-14.
- Soberón, J., & Nakamura, M. (2009). Niches and distributional areas: Concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences*, 106(2): 19644-19650.
- Souza, M., & Giovanni, D. (2009). OpenModeller: a generic approach to species potential distribution modelling. *Geoinformatica.*, 15-23.

- Stockwell, D., & Peterson, A. (2002). Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, 148: 1-13.
- Villar, B. (2014). *Modelos bayesianos para la distribución de especies con registros de solo presencias*. México: Maestría en Ciencias Agrícolas.
- Yumiseva, C. (2014). *Modelamiento Predictivo de Distribución de Rhodnius Ecuadoriensis, Vector de la enfermedad de Chagas en Ecuador*. Quito: Tesis de Grado San Francisco de Quito.