



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: “GENERACIÓN DE UN MODELO DE DISTRIBUCIÓN DE LA
ESPECIE INVASORA POLISTES VERSICOLOR EN LA ISLA SANTA
CRUZ, GALÁPAGOS, MEDIANTE EL USO DE HERRAMIENTAS
GEOINFORMÁTICAS”

AUTOR: PACHACAMA VALLEJO, KENNETH RICARDO

DIRECTOR: DR. PADILLA ALMEIDA, OSWALDO VINICIO

SANGOLQUÍ - ECUADOR

ENERO 2020



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

CERTIFICACIÓN

Certifico que el trabajo de titulación, “GENERACIÓN DE UN MODELO DE DISTRIBUCIÓN DE LA ESPECIE INVASORA *POLISTES VERSICOLOR* EN LA ISLA SANTA CRUZ, GALÁPAGOS, MEDIANTE EL USO DE HERRAMIENTAS GEOINFORMÁTICAS” realizado por el señor *Pachacama Vallejo, Kenneth Ricardo*, 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 Fuerzas Armadas ESPE, por lo tanto, me permito acreditarlo y autorizar al señor *Pachacama Vallejo, Kenneth Ricardo* para que lo sustente públicamente.

Sangolquí, 19 de diciembre del 2019

Dr. Padilla Almeida, Oswaldo Vinicio



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

AUTORÍA DE RESPONSABILIDAD

Yo, *Pachacama Vallejo, Kenneth Ricardo*, con cédula de identidad N° 060393663-4, declaro que este trabajo de titulación “**GENERACIÓN DE UN MODELO DE DISTRIBUCIÓN DE LA ESPECIE INVASORA *POLISTES VERSICOLOR* EN LA ISLA SANTA CRUZ, GALÁPAGOS, MEDIANTE EL USO DE HERRAMIENTAS GEOINFORMÁTICAS**” 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, 19 de diciembre del 2019

Pachacama Vallejo, Kenneth Ricardo
C.C: 060393663-4



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**

AUTORIZACIÓN

Yo, *Pachacama Vallejo, Kenneth Ricardo*, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar en la biblioteca Virtual de la institución el presente trabajo de titulación **“GENERACIÓN DE UN MODELO DE DISTRIBUCIÓN DE LA ESPECIE INVASORA *POLISTES VERSICOLOR* EN LA ISLA SANTA CRUZ, GALÁPAGOS, MEDIANTE EL USO DE HERRAMIENTAS GEOINFORMÁTICAS”** cuyo contenido, ideas y criterios son de mi autoría y responsabilidad.

Sangolquí, 19 de diciembre del 2019

Pachacama Vallejo, Kenneth Ricardo
C.C: 060393663-4

DEDICATORIA

A mi padre Richard, mi madre Rosa y mis hermanos Vladimir y Kevin fuente de mi fortaleza y de inspiración que día a día me enseñan a crecer primero como ser humano y luego como profesional, me brindan ese cálido hogar en el cual siempre encuentro un refugio y respaldo ante cualquier adversidad. Este y todos los logros en el futuro nunca serían posibles sin su ayuda profesional, emocional y espiritual.

A toda mi familia quienes me vieron crecer, siempre estuvieron pendientes de mí en todas las etapas de vida y siempre me brindaron su apoyo incondicional, el sentir una familia unida durante toda la vida ha sentado las bases de quien soy y de todo lo que puedo llegar hacer.

A Kevin, Josué, Mateo, Pao, Katy y Dennis que unidos por la música vivimos una etapa única que siempre llevaré en el corazón y me ayudo aprender y enfrentar difíciles momentos de vida.

A todos mis amigos más cercanos de universidad con los cuales formamos una increíble amistad y permitieron que este viaje sea maravilloso, especialmente a Marilyn amiga excepcional que de manera incondicional estuvo en buenos y malos momentos.

AGRADECIMIENTO

Al Dr. Oswaldo Padilla tutor del proyecto y excelente profesional quien me guio en este trabajo y me brindo todo su ayuda en la academia.

Al Dr. Henry Herrera, Dr. Oscar Mollá investigadores de la ESPOCH que me dieron el apoyo y la oportunidad de aportar en el proyecto “ESTUDIO DE ESPECIES INVASORAS Y DIVERSIDAD DE INVERTEBRADOS TERRESTRES EN LAS ISLAS GALAPAGOS, EN EL MARCO DEL CONVENIO ESPOCH Y FUNDACION CHARLES DARWIN”.

Mi agradecimiento a los profesores de la ESPE por compartir sus conocimientos dar oportunidad y guiarme a través de estos años de estudios.

A la dirección del Parque Nacional Galápagos por brindarme las facilidades y apoyo dentro de Santa Cruz al realizar el trabajo de campo.

ÍNDICE DE CONTENIDO

CERTIFICACIÓN DEL DIRECTOR DEL TRABAJO DE TITULACIÓN.....	i
AUTORÍA DE RESPONSABILIDAD	ii
AUTORIZACIÓN	iii
DEDICATORIA	iv
AGRADECIMIENTO	v
ÍNDICE DE CONTENIDO	vi
ÍNDICE DE TABLAS	x
ÍNDICE DE FIGURAS	xi
ÍNDICE DE ANEXOS	xiv
RESUMEN	xv
ABSTRACT	xvi
CAPITULO I	1
INTRODUCCIÓN.....	1
1.0 Resumen	1
1.1 Antecedentes.....	1
1.2 Planteamiento del problema	4
1.3 Justificación e importancia	6
1.4 Objetivo general	8
1.5 Objetivos específicos	8
1.6 Metas	8
CAPITULO II.....	10
MARCO TEÓRICO	10

2.0	Resumen	10
2.1	Nicho ecológico	10
2.2	Distribución geográfica de una especie	12
2.3	Modelos de nicho ecológico y modelos de distribución de especies.....	13
2.4	Sistemas de Información Geográfica.....	14
2.4.1	Modelamiento de Distribución de Especies (MDE)	15
2.4.2	Aplicación y uso de los SIG en los modelos de distribución de especies	16
2.4.3	Técnicas de modelación de distribución de especies	17
2.4.4	Máxima Entropía (Maxent).....	17
2.4.5	Regresión Multivariada (MARS)	20
2.4.6	BIOCLIM	23
2.4.7	Regresión Logística.....	23
2.5	Validación de los modelos de distribución de especies.....	25
2.5.1	Matriz de confusión.....	25
2.5.2	Curva ROC (Receiver Operating Characteristic)	27
2.5.3	Estadístico Kappa (k)	29
2.6	Normalización de variables	29
2.6.1	Método de análisis jerárquico AHP (Thomas L. Saaty).....	30
2.7	Variables climáticas Worldclim	32
2.8	Islas Galápagos	33
2.8.1	Isla Santa Cruz	37
2.8.2	Fragilidad del ecosistema isleño de Galápagos	38
2.8.3	Especies invasoras.....	39
2.8.4	Proyectos de control y erradicación de especies invasoras	39
2.9	Descripción de la especie <i>P. versicolor</i>	41
2.10	Base Legal	43
2.11	Hipótesis	44
2.12	Variables de la investigación	44

CAPITULO III	49
METODOLOGÍA	49
3.0 Resumen	49
3.1 Metodología de desarrollo del proyecto	49
3.2 Planificación de toma de datos	50
3.2.1 Determinación de áreas con mayor probabilidad de avistamiento.....	50
3.2.2 Método de recopilación de datos.....	51
3.2.3 Aplicación del muestreo por transectos.....	52
3.3 Variables dependientes	54
3.3.1 Datos de presencia de <i>P. versicolor</i>	54
3.3.2 Datos de pseudo-ausencias de <i>P. versicolor</i>	56
3.4 Variables independientes	58
3.4.1 Variables climáticas	58
3.4.2 Altitud.....	60
3.4.3 Infraestructura	61
3.4.4 Cobertura vegetal y uso de suelo.....	62
3.5 Normalización de las variables.....	66
3.6 Modelo de Distribución de Especies	68
3.6.1 Maxent (Máxima Entropía).....	68
3.6.2 Splines de Regresión Multivariada Adaptativa (MARS)	74
3.6.3 BIOCLIM	82
3.6.4 Regresión logística	88
CAPITULO IV	92
RESULTADOS y DISCUSIONES.....	92
4.0 Resumen	92
4.1 Análisis estadísticos.....	92
4.1.1 Resultados del modelo Maxent	93

4.1.2	Resultados del modelo MARS	94
4.1.3	Resultados del modelo BIOCLIM.....	95
4.1.4	Resultados del modelo Regresión Logística	96
4.2	Análisis de la curva ROC	97
4.3	Análisis contribución de variables independientes.....	99
4.4	Análisis gráfico de los modelos de distribución de especies.....	101
CAPITULO V		107
CONCLUSIONES Y RECOMENDACIONES		107
5.0	Resumen	107
5.1	CONCLUSIONES.....	107
5.2	RECOMENDACIONES	110
REFERENCIAS BIBLIOGRÁFICAS		111
ANEXOS		118

ÍNDICE DE TABLAS

Tabla 1 Tabla de contingencia	27
Tabla 2 Descripción de las variables dependientes e independientes	45
Tabla 3 Cronograma de actividades durante el muestreo.....	54
Tabla 4 Comparación de la variable temperatura.....	60
Tabla 5 Matriz Thomas L. Saaty para la variable uso de suelo.....	63
Tabla 6 Parámetros normalizados de la variable uso de suelo.....	63
Tabla 7 Variables independientes o predictoras usadas	65
Tabla 8 Valores máximos y mínimos de cada variable.....	67
Tabla 9 Formato de los datos de presencia para Maxent.....	69
Tabla 10 Contribución de cada variable al modelo Maxent.....	72
Tabla 11 Funciones básicas resultantes del modelo MARS.....	80
Tabla 12 Coeficientes y estadísticos resultantes de las variables en el modelo Regresión Lineal.....	90
Tabla 13 Contribución de las variables independientes	100

ÍNDICE DE FIGURAS

Figura 1. Representación gráfica del nicho ecológico por sus tres variables.....	11
Figura 2. Representación gráfica de un SIG.	15
Figura 3 Ejemplo de función bisagra con un nudo en $x = 3.1$	22
Figura 4. Curva ROC (<i>Receiver Operating Characteristic</i>).	28
Figura 5: Interpretación por Swets del AUC.....	29
Figura 6. Criterios de ponderación de Thomas L. Saaty.....	31
Figura 7. Matriz de criterios de Thomas L. Saaty.....	31
Figura 8. Obtención de la evaluación Thomas L. Saaty.	32
Figura 9. Ubicación de estación Climáticas usadas por Worldclim.....	33
Figura 10. Nido de avispa <i>P. versicolor</i>	42
Figura 11. Flujograma de la metodología usada para el proyecto.	50
Figura 12. Área de estudio y planificación de muestreo.....	53
Figura 13. Puntos de muestreo sobre el área de estudio.....	56
Figura 14. Proceso para generar los puntos de pseudo-ausencias.....	57
Figura 15. Rásters de las variables climáticas usadas.	59
Figura 16. Variable altitud generada a partir del modelo digital de elevaciones.	61
Figura 17. Variable distancias a infraestructura antrópica.....	62
Figura 18. Ráster de la variable uso de suelo normalizada.....	64
Figura 19. Ejemplo de variable normalizada entre los valores de 0 a 1.....	68
Figura 20. Interfaz principal Maxent.....	70

Figura 21. Diagrama del análisis Jackknife para las variables del entrenamiento. Entrenamiento sin la variable (celeste), entrenamiento solo con la variable (azul).....	73
Figura 22. Curvas de respuesta del modelamiento.....	73
Figura 23. Curva de respuesta de la variable distancias a infraestructura antrópica.....	74
Figura 24. Base de datos de entrada para MARS.....	75
Figura 25. Interfaz principal del software MARS.....	76
Figura 26. Configuración del test de MARS.....	77
Figura 27. Configuración del modelamiento MARS.	78
Figura 28. Ventana resultante de modelamiento con las curvas y estadísticos de aprendizaje....	79
Figura 29. Herramienta <i>Raster Calculator</i> : transformación de las funciones básicas a ráster. ...	81
Figura 30. Importación de variables formato ASCII a formato GRID.	83
Figura 31. Creación de Stack de variables predictoras.	83
Figura 32. Ingreso de puntos de muestreo a DIVA-GIS.....	84
Figura 33. Preparación de puntos para el entrenamiento y la prueba.	85
Figura 34. Configuración de la interfaz de BIOCLIM.....	86
Figura 35. Extracción de archivo de evaluación de los puntos de prueba.	87
Figura 36. Curva ROC resultante de modelamiento.	87
Figura 37. Ingreso de base de datos a SPSS.....	88
Figura 38. Configuración de la Regresión Logística.....	89
Figura 39. Generación de la curva ROC.	91
Figura 40. Curva ROC resultante de la Regresión Logística.	91
Figura 41. Diagrama de omisión: entrenamiento (azul); prueba (celeste); prevista (negra); fracción de pseudo-ausencias pronosticadas (rojo).	93

Figura 42. Estadísticos resultantes y curvas de aprendizaje en el modelo MARS.....	94
Figura 43. Matriz de confusión resultante del modelo MARS.	95
Figura 44. Estadístico Kappa del modelamiento BIOCLIM.....	96
Figura 45. Matriz de confusión resultante de la Regresión Logística.....	96
Figura 46. Curva ROC del modelo Maxent. curva entrenamiento (rojo), curva de entrenamiento (azul).....	97
Figura 47. Curva ROC del modelo MARS.	98
Figura 48. Curva ROC del modelo BIOCLIM.....	98
Figura 49. Curva ROC del modelo Regresión Logística.	99
Figura 50. Modelo Maxent.....	102
Figura 51. Modelo MARS.....	103
Figura 52. Modelo BIOCLIM.	104
Figura 53. Modelo Regresión Logística.....	105
Figura 54. a) Sección del ráster del modelo Maxent. b) Sección del ráster de variable distancias a infraestructura urbana.....	106

ÍNDICE DE ANEXOS

<i>Anexo 1.</i> Zonas de vegetación	119
<i>Anexo 2.</i> Estados de desarrollo de la avispa <i>P. versicolor</i>	120
<i>Anexo 3.</i> Morfología de la avispa <i>P. versicolor</i>	121
<i>Anexo 3.1.</i> Vista superior de <i>P. versicolor</i>	123
<i>Anexo 3.2.</i> Vista frontal de <i>P. versicolor</i>	123
<i>Anexo 4.</i> Muestreo de campo	124
<i>Anexo 4.1</i> Muestreo en Totuga Bay	124
<i>Anexo 4.2.</i> Muestreo en cerro Crocker	125
<i>Anexo 4.3.</i> Ejemplo de la base de datos recopilada	126
<i>Anexo 5.</i> Base de datos digital.....	126
<i>Anexo 6:</i> Mapa de distribución de <i>P. versicolor</i> (Modelo Maxent) en la isla Santa Cruz.....	127
<i>Anexo 7:</i> Mapa de distribución de <i>P. versicolor</i> (Modelo MARS) en la isla Santa Cruz.....	128
<i>Anexo 8:</i> Mapa de distribución de <i>P. versicolor</i> (Modelo BIOCLIM) en la isla Santa Cruz.....	129
<i>Anexo 9:</i> Mapa de distribución de <i>P. versicolor</i> (Modelo Regresión Logística) en la isla Santa Cruz.....	130

RESUMEN

La avispa *Polistes versicolor* Oliver 1791 (Hymenoptera: Vespidae) o comúnmente conocida como avispa del papel se encuentra en varios ecosistemas de las Islas Galápagos y está catalogada como una especie invasora que afecta a la flora y fauna endémica. Además, su comportamiento agresivo puede representar un problema para actividades turísticas y agrícolas, afectando de este modo no solo a la biodiversidad, sino también al sector productivo del archipiélago. La Dirección del Parque Nacional Galápagos (DPNG) y la Escuela Superior Politécnica de Chimborazo (ESPOCH) realizan trabajos en conjunto, con el objetivo de controlar las poblaciones de especies invasoras como *P. versicolor*. Con el fin de aportar a este esfuerzo se plantea una forma de entender la distribución espacial de esta especie mediante herramientas geoinformáticas, que permitan realizar un modelo de su distribución en las islas. Los modelos probados fueron: Maxent (máxima entropía), MARS (splines de regresión multivariada adaptativa), BIOCLIM y Regresión Logística que usan diferentes metodologías para determinar probabilísticamente zonas idóneas donde exista la especie. Los resultados de cada modelador fueron validados con estadísticos como: matriz de confusión, estadístico Kappa, AUC (Area Under the Curve) entre otros. También se realizó un análisis gráfico con cada modelo usando datos de prueba entre presencias y pseudo-ausencias elegidos al azar para realizar una comparación visual. El modelo MARS tuvo el mejor rendimiento predictivo con un AUC de 0.99 y un análisis gráfico muy satisfactorio. La variable más influyente para la distribución de la especie fue la distancia a infraestructura antrópica.

PALABRAS CLAVE:

- ***POLISTES VERSICOLOR***,
- **MAXENT**
- **MARS**

ABSTRACT

The Wasp *Polistes versicolor* Oliver 1791 (Hymenoptera: Vespidae) or often known as paper wasp is found in several ecosystems of the Galapagos Islands and is listed as an invasive species that affects endemic flora and fauna. In addition, its aggressive behavior may represent a problem for tourism and agricultural activities, thus affecting not only biodiversity, but also the productive sector of the archipelago. The Dirección del Parque Nacional Galápagos and the Escuela Superior Politécnica de Chimborazo (ESPOCH) carry out joint work, with the aim of controlling populations of invasive species such as *P. versicolor*. In order to contribute to this effort, we propose a way of understanding the spatial distribution of this species through geoinformatics tools, which can be done by a distribution model in the islands. The models tested were: Maxent (maximum entropy), MARS (adaptive multivariate regression splines), BIOCLIM and Logistic Regression that use different methodologies to determine probabilistically suitable areas where the species exists. The results of each modeler were validated with statistics such as confusion matrix, Kappa statistic, AUC (Area under the curve) among others. A graphic analysis was also performed with each model using test data between presences and pseudo-absences chosen at random to make a visual comparison. The MARS model had the best predictive performance with an AUC of 0.99 and a very satisfactory graphical analysis. The most influential variable for the distribution of the species was the distance to the anthropic infrastructure.

KEY WORDS:

- *POLISTES VERSICOLOR*
- MAXENT
- MARS

CAPITULO I

INTRODUCCIÓN

1.0 Resumen

Las Islas Galápagos son mundialmente reconocidas tanto por su atractivo turístico como por su importancia científica. Alberga especies de flora y fauna únicas en el mundo, esenciales para el entendimiento de la evolución de las especies y estudios biogeográficos. En este apartado se presenta la importancia y necesidad de cuidar el archipiélago y porque los esfuerzos de control y posible erradicación de las especies invasoras como *Polistes versicolor* Oliver 1791 (Hymenoptera: Vespidae), representan un papel fundamental en el cuidado de la biodiversidad del archipiélago.

1.1 Antecedentes

Las Islas Galápagos de origen volcánico (nunca unidas al continente) están ubicadas a 900 km de distancia de la masa continental más cercana; constituyen uno de los archipiélagos más importantes a nivel mundial para el estudio de la evolución de especies y biogeografía, ha sido denominada como uno de los “Grandes laboratorios de la evolución” por su aislamiento y características biogeográficas (Peck, 1996). El archipiélago es fuente de números estudios científicos por sus especies endémicas únicas en el mundo, que inspiraron el libro mundialmente conocido: *Sobre el origen de las especies por medio de la selección natural, o la conservación de las razas favorecidas en la lucha por la existencia 1859* (Darwin, 1980). Sin embargo, el ecosistema insular presenta una gran fragilidad ante los cambios provenientes por la colonización, debido a la destrucción de hábitats y la introducción de especies no nativas, las cuales compiten por alimento y territorio, terminando en muchos casos en la extinción de especies endémicas (Llorente, 2001).

Las especies invasoras en las Islas Galápagos representan el principal problema de amenaza a la biodiversidad terrestre ya que se estima que más de 1400 especies de plantas y animales exóticos interactúan negativamente con el ecosistema (Parque Nacional Galápagos, 2019). Hasta el 2006 se habían registrado 490 especies de insectos introducidos de los cuales las hormigas de fuego, *Wasmannia auropunctata* Roger 1863 (Hymenoptera: Formicidae) y *Solenopsis geminata* Fabricius 1804 (Hymenoptera: Formicidae); mosca parasítica, *Philornis downsi* Dodge & Aitken 1968 (Diptera: Muscidae); y avispas, *Brachygastra lecheguana* Latreille 1824 (Hymenoptera Vespidae) y *P. versicolor*; son consideradas como especies invasoras y representan un riesgo extremo para las Islas Galápagos (Parque Nacional Galápagos, 2019). Por su agresividad, fácil adaptación y alta densidad poblacional los proyectos de control y erradicación presentan una gran complejidad científica, técnica y social para las autoridades de las islas (Consejo de Gobierno del Régimen Especial de Galápagos, 2016).

En la actualidad, la generalización de los Sistemas de Información Geográfica (SIG) y el desarrollo de técnicas estadísticas aplicadas, permiten la aplicación de herramientas para el análisis de patrones espaciales de presencia y ausencia de especies; es decir, modelos de distribución de especies (MDE), que utilizan la información disponible en: bases de datos de museos, colecciones, registros de localidades de ocurrencia, publicaciones, así como las variables ambientales de dichas localidades como: precipitación, humedad, temperatura, entre otros. Y así obtener una visión más amplia del comportamiento espacial de la especie sobre un espacio geográfico determinado (Franklin, 1995).

Para realizar los modelos de distribución de especies, se puede usar cualquier clasificador estadístico que se ajuste al tipo de variable modelizada; la cual puede ser dicotómica si se tiene

valores de presencia y ausencia o continua si son de abundancia (Elith, 2006).

La aplicación de los SIG ha sido principalmente con grupos de mamíferos y plantas, y en menor caso con algunos grupos de invertebrados (Sánchez V. , 2001). Sin embargo, la dominancia numérica, la función dentro de los ecosistemas y el potencial no aprovechado; hacen que los invertebrados sean de suma importancia para la dinámica y biodiversidad de los ecosistemas (Guzmán, 2016).

En un estudio realizado por el Ministerio de Agricultura, Alimentación y Medio Ambiente del Gobierno de España (Ministerio, 2014), donde se realizó un modelo de riesgo de invasión de la avispa *Vespa velutina* Lepeletier 1936 (Hymenoptera: Vespidae), se contó únicamente con los datos de presencia proporcionado por localidades y se seleccionaron cuatro factores ambientales potencialmente relacionados con la distribución de la especie: temperatura media anual, precipitación media anual, usos del suelo, distancia a ríos. En este trabajo se recomienda tomar los modelos resultantes con prudencia, ya que no se conoce con certeza los datos de ausencia de la especie, las zonas de baja probabilidad no se deben entender como zonas con probabilidad cero. También la variable distancia a ríos casi no obtiene ganancia, así que la variable por sí misma no es útil para estimar la distribución y es la variable precipitación la más importante para la determinación de la distribución espacial.

En base a este trabajo, se puede deducir que es muy importante definir una metodología clara para la aplicación del modelo de distribución de especies donde se determine las variables independientes más importantes y de mejor accesibilidad, así como las variables dependientes provenientes de la recopilación de datos de presencias y ausencias, pues así se podría optimizar recursos y lograr altas precisiones en el modelo de distribución final.

En este documento se pretende generar una herramienta que ayude a estimar la distribución de la especie en el archipiélago lo que será de gran utilidad para su control, basándose en la determinación de la técnica de modelación con mejor rendimiento y partiendo de un muestro de campo de la especie en la isla Santa Cruz. Así también la técnica determinada podrá ser replicada en otras islas y con otras especies.

1.2 Planteamiento del problema

Según el Ministerio del Ambiente de Ecuador en su proyecto para la erradicación y control de especies invasoras, existe evidencia que indica que ciertas especies de animales introducidos a las Islas Galápagos han provocado extinciones (Ministerio del Ambiente, 2016). Por ejemplo, la rata negra es la causante de la extinción de las ratas nativas en las islas Santa Cruz, Baltra y San Cristóbal. Esta misma rata, casi llevó a la extinción a la población de tortugas gigantes *Chelonoidis Ehippium* Gunther 1896 (Testudines: Testunidae) de la isla Pinzón, debido a que impedía su reproducción natural (Ministerio del Ambiente, 2016). En cuanto a invertebrados introducidos, en Galápagos se han registrado 543 especies, de los cuales 492 son insectos y de estos se considera a 55 como potencialmente invasores con características para causar impactos severos sobre las islas y 8 de esos son de alto riesgo (Consejo de Gobierno del Régimen Especial de Galápagos, 2016). Aunque las especies invasoras son consideradas como una gran amenaza para la biodiversidad única de las Islas Galápagos, la verdad es que, poco se conoce acerca del impacto generado por los invertebrados introducidos y aunque no todas estas especies suponen el mismo nivel de riesgo, todas tienen algún efecto como la depredación o competencia por alimentos y espacio de vida, creando algún tipo de alteración al ecosistema natural. Además, muchas de estas especies sirven como hospederos o portadores de parásitos y enfermedades que también podrían tener efectos

devastadores sobre el frágil ecosistema de las islas (Ministerio del Ambiente, 2016).

Una de estas especies invasoras y objeto de estudio para el presente proyecto es la avispa *P. versicolor* que fue reportada por primera vez en 1988 en la isla Floreana (Abedrabbo, 1991). Desde entonces se la encuentra en varios lugares del archipiélago (Albelo & Causton, 1999). En una publicación hecha sobre el estado de artrópodos terrestres en Galápagos, se menciona que las principales presas de esta avispa serían las larvas de Lepidoptera (elemento principal en la dieta de algunas aves endémicas) y los escarabajos carábidos endémicos *Calosoma granatense* Géhin 1885 (Coleoptera: Carabidae) (Heraty & Stewart, 1998). Estudios han demostrado que *P. versicolor* compite con las aves endémicas de las Islas Galápagos por alimento, por ejemplo, en algunas islas se ha observado que la biomasa de esta avispa es más alta que la biomasa de los pinzones (Parent, 2000). Además, también afecta a las prácticas agrícolas en las temporadas más altas de presencias, ya que los agricultores pueden sentirse amenazados por la dolorosa picadura de la avispa al realizar sus actividades productivas, por lo cual las actividades económicas y de producción pueden verse afectadas (McGlynn, 1999).

En las islas la avispa *P. versicolor* enfrenta enemigos naturales parásitos como: *Erechtias minúscula* Walsingham 1897 (Lepidoptera: Tinaeidae) y una especie aun no identificada del género *Holocera sp.* perteneciente a la subfamilia Holocerinae, familia Blastobasidae, así como una amplia gama de enemigos naturales predadores, entre los cuales están: aves silvestres, aves de corral, reptiles, mamíferos y otros insectos (Lasso, 1997). Otro depredador identificado es *Taygete sphecophilia* Meyrick 1936 (Lepidoptera: Autostichidae) que probablemente fue introducido después de la llegada de *P. versicolor* (Parent, 2000).

En la actualidad la Escuela Superior Politécnica de Chimborazo (ESPOCH) junto a la Dirección

del Parque Nacional Galápagos (DPNG) tienen en marcha varios estudios para realizar la búsqueda de enemigos naturales y evaluar la eficacia de estos frente a dicha especie invasora y así disminuir el impacto negativo que ésta produce en las islas. Con la misma finalidad dicha universidad, también trabaja junto a la Estación Científica Charles Darwin, en el estudio de atractivos (Escuela Superior Politécnica de Chimborazo, 2019).

Con el fin de apoyar a dicho proyecto se va a realizar el modelamiento de la distribución de la especie *P. versicolor* a partir del muestreo in situ de presencia, ausencia y abundancia en la isla Santa Cruz.

1.3 Justificación e importancia

La introducción de especies, voluntaria o involuntariamente fuera de su área natural de distribución, representa la segunda causa de pérdida de biodiversidad tanto en flora como en fauna endémica de una zona (Meffe, 1997). Estas especies invasoras normalmente poseen una gran capacidad de reproducción en el nuevo entorno, escasa dedicación a las crías y amplia valencia ecológica (Pleguezuelos, 2002). Por lo tanto, el control de las especies invasoras es uno de los desafíos más grandes para la conservación de especies silvestres endémicas en las Islas Galápagos.

La introducción de *P. versicolor* en las Islas Galápagos ocurrió aparentemente en 1988 durante un envío de bananos y otras frutas desde el Ecuador continental a la isla Floreana (Abedrabbo, 1991). En el continente y en cantidades normales las avispas no son consideradas animales plaga; sin embargo, siendo una especie introducida en las remotas Islas Galápagos las cosas pueden ser diferentes ya que, al no tener un depredador natural que regule su población, ésta afecta a la sensible dinámica natural de las islas (Baert, 1995).

Los sistemas geoespaciales son herramientas para el manejo y análisis de información que tienen el componente geoespacial, es decir, sus componentes son geolocalizados mediante un sistema de referencia específico y son diseñados para ser utilizados como aplicaciones de escritorio, su aplicación permitirá optimizar recursos al crear metodologías de predicción de distribución de dichas especies que analicen el territorio y ambiente donde la especie se desarrolla, siendo menos necesario una búsqueda de la especie invasora por todo el territorio, lo cual conllevaría muchos recursos económicos y tiempo para realizar un diagnóstico completo de una zona determinada (Franklin, 1995)

Además, este proyecto se relaciona con la prioridad de conservación promulgada por diferentes normas legales no solo nacionales sino internacionales, que fueron determinadas por la importancia que tienen las Islas Galápagos para los estudios científicos sobre la vida, tomando como ejemplo: los fines de conservación de la Dirección del Parque Nacional Galápagos y la Estación Científica Charles Darwin (Foundation, 2019).

Este proyecto de investigación es de gran relevancia, debido a que es una herramienta que servirá de apoyo para el control de especies invasoras en las Islas Galápagos.

1.4 Objetivo general

Realizar un modelo de distribución espacial de la especie invasora *P. versicolor* en la isla Santa Cruz, Galápagos, mediante el modelamiento de variables espaciales y uso de herramientas geoinformáticas.

1.5 Objetivos específicos

- Generar una geodatabase con información normalizada y estructurada de las variables dependientes e independientes que tendrán efecto sobre el modelamiento.
- Planificar el muestreo de campo en la isla Santa Cruz, mediante el análisis de los patrones de distribución donde haya más posibilidades de encontrar a la especie.
- Determinar presencia y ausencia de la especie invasora sobre la isla Santa Cruz en base a la planificación realizada para generar una base de datos espaciales.
- Realizar el modelo predictivo más adecuado y validación de éste según la base de datos espaciales y todas las variables definidas.

1.6 Metas

- Un modelo de predicción de la especie invasora *P. versicolor* sobre la isla Santa Cruz.
- Una metodología de muestreo de la especie invasora *P. versicolor* sobre la isla Santa Cruz.
- Una tabla con todas las variables espaciales que influyen en el modelamiento.
- Una base de datos del muestreo de la especie sobre la isla Santa Cruz.
- Un cuadro comparativo con las diferencias encontradas entre los modelos de distribución usados para el modelamiento de la especie invasora.

- Una tabla resumen con los resultados de la validación del modelo de predicción generado.

CAPITULO II

MARCO TEÓRICO

2.0 Resumen

Se realiza el fundamento teórico para entender los modelos de nicho ecológico (MNE) y los modelos de distribución de especie (MDE), así como los métodos usados para generarlos. Los modelos Maxent, MARS, BIOCLIM y Regresión Logística se fundamentan en procesos estadísticos y cartográficos, de este modo generan las áreas probabilísticamente más aptas para que se presente un fenómeno específico, en este caso la presencia de la especie. La validación de este proceso se realiza a través del rendimiento de predicción, que es medido a través de estadísticos resultantes como: matrices de confusión, estadístico Kappa, curva ROC entre otros. Para tener una visión mucho más amplia del fenómeno se hace una recopilación bibliográfica acerca de las Islas Galápagos y la especie *P. versicolor*.

2.1 Nicho ecológico

Para el estudio del espacio desde un punto de vista biogeográfico y ecológico, se necesita un enfoque claramente cuantitativo; el cual integra una serie de subdisciplinas que ayudan a la identificación de patrones y procesos subyacentes (Ezcurra & Ruggiero, 2003). De este modo se trata de describir cuales son las interacciones y dinámicas que caracterizan y permiten la evolución de un ecosistema sobre un espacio geográfico determinado o lo que puede ser denominado como “nicho ecológico”, dicho término ha tenido a través del tiempo varias interpretaciones que intentan definir toda una serie de variables ambientales y su relación con los seres vivos sobre un área específica (Illoldi & Tanla, 2008).

En la actualidad el concepto más utilizado por la mayoría de los biólogos fue dado por Hutchinson (1957), quien define al nicho como “*La suma de todos los factores ambientales que actúan sobre un organismo; el nicho es una región sobre un espacio multidimensional*”; de tal modo que, si se considera p factores ambientales, entonces al espacio se lo puede medir como p -dimensional, siendo este un espacio ecológico. Partiendo de lo dicho, un nicho ecológico es el conjunto de características, variables ambientales tanto bióticas como abióticas e interacciones entre especies, que describen las condiciones precisas que necesita un organismo para sobrevivir (López, 2007).

En el espacio ecológico se pueden diferenciar regiones ambientales en donde las interacciones serán diferentes, estas serán conocidas como tipos de nicho ecológico: El nicho fundamental es definido como las condiciones bióticas y abióticas en donde el organismo es capaz de sobrevivir, el nicho potencial es la interacción del nicho fundamental con el espacio ecológico disponible sobre el planeta Tierra y por último el nicho realizado que es definido como la parte del espacio ecológico donde existe la especie (López, 2007). En la Figura 1 se puede ver la representación gráfica de la interacción entre los 3 tipos de nicho.

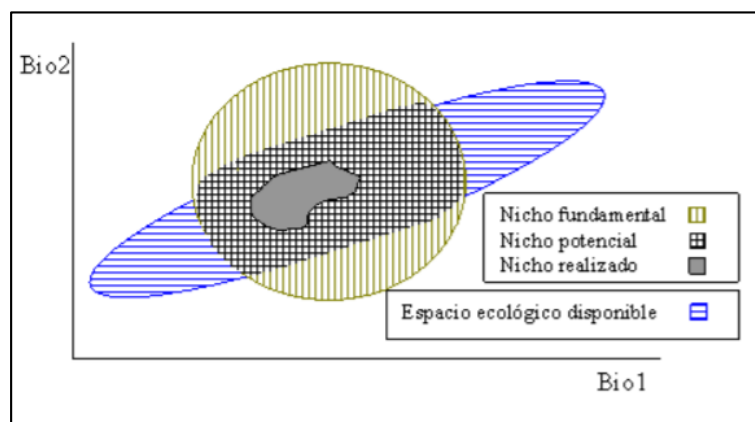


Figura 1. Representación gráfica del nicho ecológico por sus tres variables.

Fuente: (López, 2007).

2.2 Distribución geográfica de una especie

La distribución de una especie en su entorno depende de la gran variedad de factores existentes dentro de un espacio geográfico determinado, que van desde un punto de vista macro como los ecosistemas, hasta una visión mucho más precisa como las diferentes derivas genéticas; estos factores hacen que la naturaleza sea biodiversa en todos sus aspectos. En varios estudios se ha podido demostrar que cada uno de estos factores es vital para el desarrollo de una especie, a tal punto que el cambio causado en una especie tiene efectos casi inevitables sobre otras (Nakamura & Soberon, 2007).

Al espacio geográfico o región donde la vida de la especie es viable, se la conoce como distribución geográfica y se la puede clasificar en dos tipos: distribución realizada que es el espacio donde la especie existe actualmente y la distribución potencial donde las condiciones de vida son favorables pero no necesariamente se encuentra allí; este comportamiento es debido a diferentes factores como: evolución, accesibilidad, competencia, entre otras (Felicísimo & Muñoz, 2011). El conocimiento de la distribución de una especie sobre su territorio permite desarrollar estrategias de conservación como la formulación de inventarios o la predicción de especies invasoras (López, 2007).

La distribución de una especie puede ser interpretada de forma ligera mediante los patrones de dispersión o patrones de distribución, que se refieren a como los individuos de una población se distribuyen en el espacio en un momento determinado. Así los individuos de dicha población pueden estar espaciados de manera más o menos uniforme, dispersos sin patrón predecible, o formado por grupo. Estos patrones de distribución se conocen como uniforme, aleatorio y agrupado, respectivamente (Academy, 2019).

2.3 Modelos de nicho ecológico y modelos de distribución de especies

Los modelos de nicho ecológico (MNE) y modelos de distribución de especies (MDE) son usados indistintamente ya que conceptualmente pueden ser muy parecidos. Mientras los modelos de nicho ecológico son principalmente usados para relacionar los valores ambientales con datos de presencia, ausencia y abundancia de la especie, los modelos de distribución de especies son el resultante del análisis del MNE, donde se determina la veracidad de la distribución en función de la realidad del espacio geográfico y no solo las variables ambientales (Graham et al., 2004).

Partiendo de esta base teórica se realizan los modelos de distribución de especie cuya importancia radica en resolver preguntas de ecología aplicada, conservación y biogeografía; esta herramienta ayuda a describir el comportamiento y la forma en la que una especie determinada se expande sobre el territorio, así como las interacciones que dicha especie tiene o puede tener con otras especies y factores abióticos (lluvia, humedad, temperatura, entre otras), por este motivo su aplicación en proyectos para la conservación de fauna endémica y el control de especies invasoras es de mucha ayuda para la toma de decisiones, con una visión más amplia de la especie sobre el territorio (Guisan & Thuiller, 2005).

Una observación muy importante acerca de estos dos modelos es que; el modelo de nicho ecológico no necesariamente representa la distribución de una especie, de tal modo que, puede estar sobre representando el área de distribución real, el modelo de nicho ecológico encuentra todas las áreas donde hay las mismas condiciones ecológicas sobre las cuales se tomaron los datos de presencia, sin embargo, estas áreas pueden estar en lugares donde históricamente se conoce que la especie no se reproduce. Las razones pueden estar relacionadas directamente con el taxón más que con el proceso del modelamiento; razones como barreras geográficas, quizás su extinción reciente,

falta de muestreo o incluso la competencia con otra especie, hacen que la distribución real de una especie en el modelo pueda no ser exacta y deba ser tomada más como una hipótesis que como una representación de la realidad (Maciel & Manríquez, 2015). Es de mucha importancia que a los resultados finales se le realice un análisis por parte de expertos del taxón que ajusten el modelo en caso de ser necesario mediante la edición del modelo final en base a un criterio experto (Iloldi & Tanla, 2008).

2.4 Sistemas de Información Geográfica

A lo largo del tiempo varios autores han contribuido a la definición del término Sistemas de Información Geográfica (SIG), cada uno en base a la temporalidad y necesidades que se presentaban en sus momentos. Una apreciación clásica es la de Tomlin (1990), quien decía que un SIG es un elemento para el análisis, presentación e interpretación de cualquier hecho relativo a la superficie terrestre. El mismo autor explica que es una definición muy amplia y de una forma más específica la definía como: *“Es un conjunto de software y hardware diseñado específicamente para la adquisición, mantenimiento y uso de datos cartográficos”*.

Por otra parte, pero en un ambiente similar, se define un SIG como un sistema de información diseñado para trabajar con datos referenciados mediante coordenadas espaciales o geográficas. Es decir, un SIG es tanto un sistema de bases de datos que tienen capacidades específicas para trabajar con datos georreferenciados, como un conjunto de operaciones que permiten el tratamiento de estos (Olaya, 2014).

Un SIG es considerado como un “mapa de orden superior” que plantea una forma más avanzada de hacer todo aquello relacionado a los mapas (Star & Estes, 1990). Pero hoy en día estas definiciones son muy simplistas ya que, en el contexto actual, mapa y SIG no son conceptos

equiparables. Debe entenderse a un SIG como un aparato complejo que abarca una serie de elementos acoplados, los cuales desempeñan una función particular para el manejo y representación de datos, estos elementos son: los datos, los procesos, la visualización, la tecnología y el factor organizativo (Lara & Simeón, 2006). En la Figura 2 se puede ver una representación gráfica de un SIG, el cual está compuesto de varias capas de información que serán copiladas para su análisis.

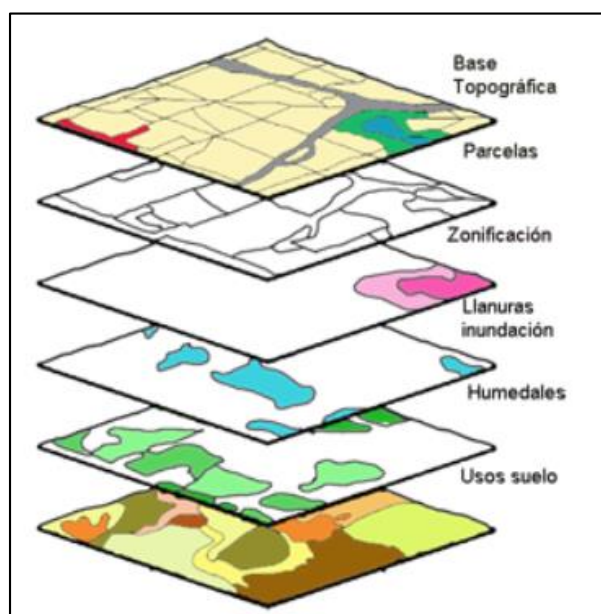


Figura 2. Representación gráfica de un SIG.

Fuente: (Salamanca, 2019).

Entonces por lo ya dicho, SIG es un sistema que integra tecnología informática, personas e información geográfica y cuya principal función es capturar, analizar, almacenar, editar y representar datos georreferenciados (Olaya, 2014).

2.4.1 Modelamiento de Distribución de Especies (MDE)

Un modelo es una representación parcial de la realidad que refleja algunas de sus propiedades, por lo tanto, los modelos son simplificaciones debidas tanto a la necesidad de reducir la

complejidad del objeto real, como al desconocimiento de muchas de sus propiedades. Los modelos de distribución de especies se convierten en representaciones cartográficas de la idoneidad de un espacio para la presencia de una especie en función de las variables empleadas para generar dicha representación. La idoneidad no es más que la relación matemática o estadística entre la distribución real conocida y un conjunto de variables independientes que se usan como indicadores. Estas variables suelen ser geológicas, topográficas o climáticas, y se espera que, con algunas de ellas, individualmente o en combinación, se puedan definir los factores ambientales que delimiten las condiciones favorables para la presencia de la especie (Zimmermann & Guisan, 2000).

Los modelos de distribución de especies son herramientas muy prácticas para la evaluación, gestión y conservación de especies cuando los datos de campo son puntuales y escasos. En base de un análisis de datos de cobertura, clima, tipo de suelo, geomorfología y observaciones de campo se puede construir una distribución de la idoneidad de un sitio para que sea hábitat de una especie (Gidahatari, 2017).

2.4.2 Aplicación y uso de los SIG en los modelos de distribución de especies

Con el avance de los SIG, una de las tendencias principales es la integración de una serie de aplicaciones que se enriquecen con los conceptos y funcionalidades que ofrecen los Sistemas de Información Geográfica. El panorama que encuentra hoy en día un usuario de SIG es sumamente complejo debido a que el sistema toma elementos de otras aplicaciones unificando todo en una sola plataforma que se ha ido especializando en distintos ámbitos, siendo una base genérica sobre la cual se construyen herramientas de análisis geográfico adaptadas a distintos fines (Olaya, 2014).

En la actualidad la generalización y facilidad de acceso a los SIG, así como el desarrollo de técnicas estadísticas aplicadas, han permitido que el análisis de patrones espaciales de presencia y

ausencia de especies tenga una mayor aceptación; por ejemplo, es el caso de los modelos de distribución de especie, que gracias a la determinación de nichos ecológicos que propician la habitabilidad para una especie, se pueda determinar geográficamente zonas donde existen mejores condiciones y probabilidades para su desarrollo, como también es posible que el modelo delimite zonas potenciales muy alejadas geográficamente de las actuales, la probabilidad de encontrarlas en principio no serán altas aunque las condiciones ambientales fueran favorables (Rubén & Felicísimo, 2011).

2.4.3 Técnicas de modelación de distribución de especies

Según Elith (2006), los métodos disponibles se pueden clasificar en tres grupos: (1) Técnicas discriminantes donde se utilizan datos de presencia y ausencia para construir el clasificador, donde entran métodos como Maxent (máxima entropía) o MARS (Splines de regresión multivariada adaptativa). (2) Técnicas descriptivas, técnicas que solo requieren datos de presencia, fueron las primeras usadas en modelización como el método deductivo donde entra la opinión de un experto, programas como BIOCLIM, BIOMAP o métodos de distancias matemáticas, como DOMAIN que utiliza la distancia Gower. (3) Técnicas mixtas donde se emplean varias reglas, algunas de ellas descriptivas y otras discriminantes, a la vez genera sus propias pseudo-ausencias; aquí destacan Desktop-GARP y OM-GARP. A parte de esta clasificación, también se puede mencionar métodos estadísticos como la Regresión Logística la cual puede ser usada cuando la variable de respuesta es politómica (admite varias categorías de respuesta) pero es especialmente usada cuando la variable de respuesta es dicotómica (dos posibles respuestas) (Fernández, 2011).

2.4.4 Máxima Entropía (Maxent)

Para realizar la mejor aproximación a una distribución desconocida el principio planteado por

Jaynes (1957) dice que, la mejor manera será garantizar que la aproximación cumpla con cualquier restricción que se pueda conocer de la distribución desconocida y que sujeto a esto, la aproximación a la distribución debería tener la máxima entropía.

El proceso requiere que la media de cada característica esté cerca del promedio empírico sobre los datos de presencia, las restricciones son expresadas en términos de funciones simples de las variables ambientales, llamadas características. Un ejemplo de esto puede ser la característica precipitación anual, donde la restricción dice que la media anual de precipitación predicha por el modelo debería estar cerca de la precipitación promedio observada (Phillips, 2019). Dado que el conjunto de restricciones generalmente no se especifica en el modelo, entre todas las distribuciones de probabilidad que satisfacen las restricciones se elige la de máxima entropía ya que es la menos restringida (Phillips, 2007).

El algoritmo de Maxent está fundamentado en este concepto para predecir a base de información incompleta; el programa inicia con una distribución uniforme de probabilidades e iterativamente va alternando los pesos de las características para maximizar la similitud de los datos de ocurrencia. El algoritmo está diseñado para converger a una distribución probabilística óptima, como no aleatoriza, entonces los resultados son determinísticos. Los resultados de Maxent para cada celda son acumulativos, representando el porcentaje de valor de probabilidad para la celda en cuestión y todas aquellas con valores iguales o menores de probabilidad. Una celda con valor de 1 es la más adecuada para existir un nicho ecológico, mientras que una celda con valor de 0 será la menos adecuada (Illoldi & Tanla, 2008).

2.4.4.1 Descripción matemática de Maxent

Para la estimación de máxima densidad de la entropía, la distribución de la especie está representada como una probabilidad de distribución π sobre un conjunto x de lugares en el área de estudio. Así π evaluaciones para cada sitio x de tal forma que se asigna una probabilidad no-negativa $\pi(x)$ para cada punto x y estas probabilidades sumen 1 es decir el 100%. Se produce un modelo de distribución con respecto al conjunto de restricciones derivadas de los datos de entrada. La entropía de la distribución π está definida como:

$$H(\pi) = - \sum_{x \in X} \pi(x) \ln \pi(x) \quad (1)$$

Donde \ln es el logaritmo natural.

Otra forma de interpretar a la entropía es como la medida de cuanta “elección” está involucrada en la selección de un evento, por lo tanto, una distribución con más entropía involucra más elecciones. Por esta razón es de suma importancia para el principio de máxima entropía no colocar restricciones infundadas (Shannon, 1948). Desde una perspectiva de “machine learning” se asume que se tiene un conjunto de funciones conocidas como “características” (en este caso variables ambientales) dadas por f_1, \dots, f_n en X de este modo a cada característica f_j se le asigna un valor real $f_j(x)$ para cada punto de x en X . De este modo la expectativa de la característica bajo la probabilidad π está definida como $\sum_{x \in X} \pi(x) f_j(x)$ y denotada por $\pi[f_j]$ la cual podrá ser aproximada usando un conjunto de muestras x_1, \dots, x_m por tanto la media empírica de f_j es $\frac{1}{m} \sum_{i=1}^m f_j(x_i)$ la cual se denota como $\tilde{\pi}[f_j]$ donde $\tilde{\pi}$ es la distribución uniforme de una muestra de puntos. Haciendo uso de la dualidad convexa y considerando toda probabilidad de distribución se define (Della Pietra & Lafferty, 1997):

$$q_{\lambda}(x) = \frac{e^{\lambda * f(x)}}{Z_{\lambda}} \quad (2)$$

Donde λ es un vector coeficiente de valor real o característica de peso, f es el vector en todas la características o variables ambientales y Z_{λ} es una constante de normalización que hace que $q_{\lambda}(x)$ sume 1, a esta distribución se la conoce como distribución de Gibbs. Y para minimizar la probabilidad de puntos negativos.

$$\ln Z_{\lambda} - \frac{1}{m} \sum_{i=1}^m \lambda * f(x_i) \quad (3)$$

Para que Maxent proponga un entrenamiento de ajuste de datos las medias bajo $\tilde{\pi}$ solo deberían restringirse para estar cerca de sus valores empíricos (Dubik, 2004), por lo tanto

$$|\tilde{\pi}[f_j] - \pi[f_j]| \leq \beta_j, \quad \text{para cada } f_j \quad (4)$$

Esto también cambia la caracterización dual dando como resultado l -regularizaciones de este modo la distribución Maxent es una distribución de Gibbs que minimiza.

$$\tilde{\pi}[-\ln(q_{\lambda})] + \sum_j \beta_j |\lambda_j| \quad (5)$$

Para cada constante β_j , esto hace que Maxent fuerce l -regularizaciones o iteraciones y se enfoque en la característica (variable ambiental) más importante y las iteraciones hacen que los modelos tiendan a pocos valores de cero para λ_j (Schapire, 2005).

2.4.5 Regresión Multivariada (MARS)

El modelo splines de regresión multivariada adaptativa (MARS), fue presentado por Friedman y tiene como finalidad principal la predicción de valores en base a un conjunto de variables independientes como pueden ser las variables climatológicas. El modelo propone un procedimiento

para una regresión adaptativa no paramétrica de tal forma que no se haga ninguna suposición entre la relación de las variables dependiente e independientes, en lugar de esto construye una relación de conjuntos de coeficientes y funciones base que provienen de los datos de la relación, esto hace que, MARS sea muy adecuado para problemas con dimensiones grandes en los insumos de entrada y para el análisis de más de dos variables (Friedman, 1991).

MARS es un método que usa modelación no paramétrica que extiende el modelo lineal incorporando no linealidades e iteraciones de variables, divide el espacio de las variables productoras en diferentes subregiones. El proceso MARS genera puntos de corte para las diferentes variables, dichos puntos serán identificados a través de las funciones basales o funciones base, las cuales indicarán el inicio y el término de una región. En cada subregión en la que se ha dividido al espacio, se ajusta una función base de una variable que será lineal, en el resultado o modelo final se mostrará una composición constituida por una combinación de las funciones base generadas. Las funciones de corte están definidas por el algoritmo *forward/backward stepwise* por etapas. En este algoritmo primero se genera un modelo sobreestimado con un gran número de funciones base, y en la siguiente etapa se procede a eliminar los nodos y casos que menos contribuyen al ajuste global. El algoritmo requiere de un número máximo de funciones fijadas por el usuario (Vanegas, 2017).

2.4.5.1 Descripción matemática MARS

La ecuación general del modelo MARS se describe de la siguiente manera:

$$f(x) = \sum_{m=1}^M a_m B_m(x) \quad (6)$$

Donde a_m es el coeficiente de expansión el cual después de un ajuste tiene la mejor afinación a

los datos, B_m es la función base de la forma:

$$B_m(x) = I[x \in R_m] \quad (7)$$

Donde I es una función que indica que toma el valor de 1 si el argumento es verdad y 0 si no lo es, R_m son las subregiones o cortes desde las cuales hay una sola función base que no sea cero para cada punto x , B_m puede tomar la forma de función bisagra la cual tiene la forma de $\max(0, x - \text{constante})$ o $\max(0, \text{constante} - x)$, el algoritmo MARS selecciona de forma automática la variables y valores para cada nudo de las funciones bisagra. La forma $\max(0, x - c)$ o $\max(0, c - x)$ significa que se toma la parte positiva del condicional $x - c$ o $c - x$ (Friedman, 1991). Como se puede ver un ejemplo en la Figura 3 donde MARS genera dos funciones bisagra $\max(0, 3.1 - x)$ correspondiente a la línea entrecortada y $\max(0, x - 3.1)$ correspondiente a la línea continua, las dos con nudo en $x = 3.1$.

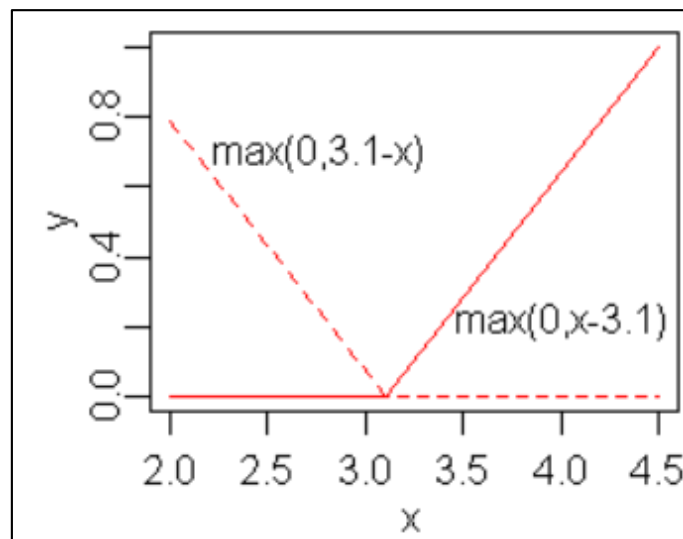


Figura 3. Ejemplo de función bisagra con un nudo en $x = 3.1$.

Fuente: (Milborrow, 2008).

2.4.6 BIOCLIM

BIOCLIM es uno de los primeros predictores bioclimáticos y uno de los métodos más fáciles de implementar, fue desarrollado por: Nix desarrollador del concepto, Busby desarrollador del código y Hutchinson desarrollador de la interpolación en su primera versión en 1986 e implementado por Houlder (Nix, 1986; Busby, 1991; Hutchinson M. , 1995). Este método usa las correlaciones de interpolación para estimar como la media de variables como: temperatura, precipitación, radiación solar, entre otros, afectan a los puntos de ubicación de la especie de interés (Booth, 2018). Con este método llamado de envoltura, BIOCLIM también genera parámetros bioclimáticos a partir de las superficies climáticas suministradas en ubicaciones conocidas de hábitat de la especie de planta o animal. Los parámetros son usados para crear un perfil bioclimático. Estos parámetros también generan una lista de ubicaciones de hábitat para la especie en forma de cuadrícula regular. El perfil climático más la cuadrícula regular de parámetros bioclimáticos son el principal insumo para predecir las ubicaciones de las especies (Hutchison, 2000).

BIOCLIM usa modelamiento correlativo que relaciona las variables climáticas con cada punto de longitud, latitud y elevación si están disponibles. El algoritmo básico usa el intervalo climático en el que se encuentran los puntos de presencia en relación con las variables predictoras y realiza el análisis teniendo tres suposiciones: 1) La distribución de las especies está determinada por el clima. 2) La distribución estadística de las variables es normal. 3) Todas las variables con intervalos restringidos influyen a la especie (Illoldi & Tanla, 2008).

2.4.7 Regresión Logística

La Regresión Logística es un procedimiento cuantitativo muy útil cuando la variable de

respuesta forma parte de un grupo o conjunto finito (Bioestadística, 2019). Esta metodología de modelamiento forma parte de un conjunto de métodos estadísticos y es la variante que corresponde, cuando varias variables de diferentes factores contribuyen a la ocurrencia de un evento simple. De modo general la Regresión Logística es adecuada cuando la respuesta es politómica (admite varias categorías de respuesta), pero es especialmente útil cuando solo existen dos posibilidades de respuesta, la respuesta dicotómica (Fernández, 2011).

Esta metodología no deja de ser un caso particular del análisis discriminante en el que la variable dependiente tiene dos categorías y parte de supuestos menos restrictivos que permiten introducir variables categóricas. La Regresión Logística permite cuantificar la relación que existe entre las variables dependientes e independientes, es decir, conocer la importancia de cada variable independiente al procedimiento. También permite clasificar individuos dentro de las categorías (presente/ausente) según la probabilidad que tenga la variable dependiente de pertenecer a cada una de ellas, en función de las variables independientes (Berlanga & Vilá, 2014).

2.4.7.1 Descripción matemática de la Regresión Logística

El modelo logístico matemáticamente se entiende de la siguiente manera: sea y una variable dependiente binaria de valores posibles 0 y 1 y sea x un conjunto de n variables independientes (x_1, x_2, \dots, x_n) observadas con el fin de predecir el valor de y .

$$P\left[y = \frac{1}{x_1, x_2, \dots, x_n}\right] \rightarrow P\left[y = \frac{0}{x_1, x_2, \dots, x_n}\right] = 1 - P\left[y = \frac{1}{x_1, x_2, \dots, x_n}\right]$$

Para esto se construye el modelo

$$P\left[y = \frac{1}{x_1, x_2, \dots, x_n}\right] = p(x_1, x_2, \dots, x_n; \beta)$$

donde:

$p(x_1, x_2, \dots, x_n; \beta)$ es una función de enlace que depende de un vector de parámetros β (Fernández, 2011):

$$\beta = (\beta_1, \beta_2, \dots, \beta_n) \quad (8)$$

2.4.7.2 Modelo de regresión logístico binario

Este modelo establece la siguiente relación entre la probabilidad de que ocurra el suceso, dado que el individuo presenta los valores ($X = x_1, X = x_2, \dots, X = x_n$):

$$P \left[y = \frac{1}{x_1, x_2, \dots, x_n} \right] = \frac{1}{1 + e^{(-\beta_0 - \beta_1 x_1 - \beta_2 x_2 - \dots - \beta_n x_n)}} \quad (9)$$

El objetivo es hallar los β_n que más se ajusten a la expresión funcional (Fernández, 2011).

2.5 Validación de los modelos de distribución de especies

Validar un modelo de distribución de especies consiste en contrastar las predicciones, en función de las presencias o variables dependientes con los datos considerados independientes, esto con el objetivo de obtener una medida de su significancia en relación con la hipótesis nula y para medir el desempeño del modelador respecto a la variable dependiente (capacidad de clasificar correctamente los datos), para esto se usan algunos estadísticos como los mencionados a continuación (Conabio, 2019).

2.5.1 Matriz de confusión

Es una tabla de contingencia usada como herramienta estadística, para el análisis de observaciones emparejadas y ofrece una visión completa de la distribución de acuerdos y errores

entre las clases. El contenido de esta tabla es un conjunto de valores que mide el grado de semejanza existente entre un conjunto de datos de control y un conjunto de datos de referencia, en los cuales se ha establecido una clasificación, formando así una matriz de dimensión $M \times M$ donde M denota el número de clases consideradas (López F. , 2015).

En los modelos de distribución de especies, estas clases son binarias (presencias y ausencias) dando una matriz de confusión de 2×2 como se ve en la Tabla 1. En esta tabla se puede ver las frecuencias de cada uno de los cuatro tipos de posibles pronósticos: (a) verdadero positivo o presencias correctamente identificadas por el entrenamiento; (b) falso positivo o presencias mal identificadas por el entrenamiento; (c) Falso negativo o ausencias mal identificadas por el entrenamiento; (d) verdadero negativo o ausencias correctamente identificadas por el entrenamiento (Conabio, 2019).

En el caso de falso positivo, éste provoca una sobre predicción denominada error por comisión y en el caso de falsos negativos provoca una pobre predicción denominada error por omisión. Entre estos dos tipos de errores el peor es el error por omisión, que no predice lugares donde si existe presencia lo cual puede ser crucial para entender la distribución de una especie (Conabio, 2019).

Tabla 1
Tabla de contingencia

		DATOS DE PRUEBA	
		Especie presente	Especie ausente
		+	-
DATOS DE ENTRENAMIENTO	Especie predicha +	Verdadero positivo (a)	Falso positivo (b)
	Especie no predicha -	Falso negativo (c)	Verdadero negativo (d)

2.5.2 Curva ROC (Receiver Operating Characteristic)

Es una prueba dicotómica y es uno de los estadísticos más usados para determinar la confiabilidad y validar un modelo de distribución de especies, su objetivo es clasificar a los individuos de una población en dos grupos; uno que presente un evento y otro que no, describe la identificación correcta de presencias (verdaderos positivos) o sensibilidad (S), contra la tasa de falsas alarmas (falsos positivos) o especificidad (E) es decir, $y = f(x)$ (Benavides, 2019) donde:

$$ROC(c) = \begin{cases} y = S(c) \\ x = 1 - E(c) \end{cases}$$

Sirve para realizar una comparación entre lo que se esperaría por selección al azar y la predicción de un modelador, mientras mayor sea el AUC (Area Under Curve) el modelamiento más se alejará del azar (Benavides, 2019). La Figura 4 presenta un ejemplo de curva ROC con un AUC de 0.68 donde se determina que el entrenamiento se acerca mucho a lo que se podría esperar por selección al azar ya que la curva ROC (línea en negro) se acerca a la línea diagonal, la cual

representa al azar, por tanto el rendimiento del modelamiento es poco satisfactorio (Benavides, 2019).

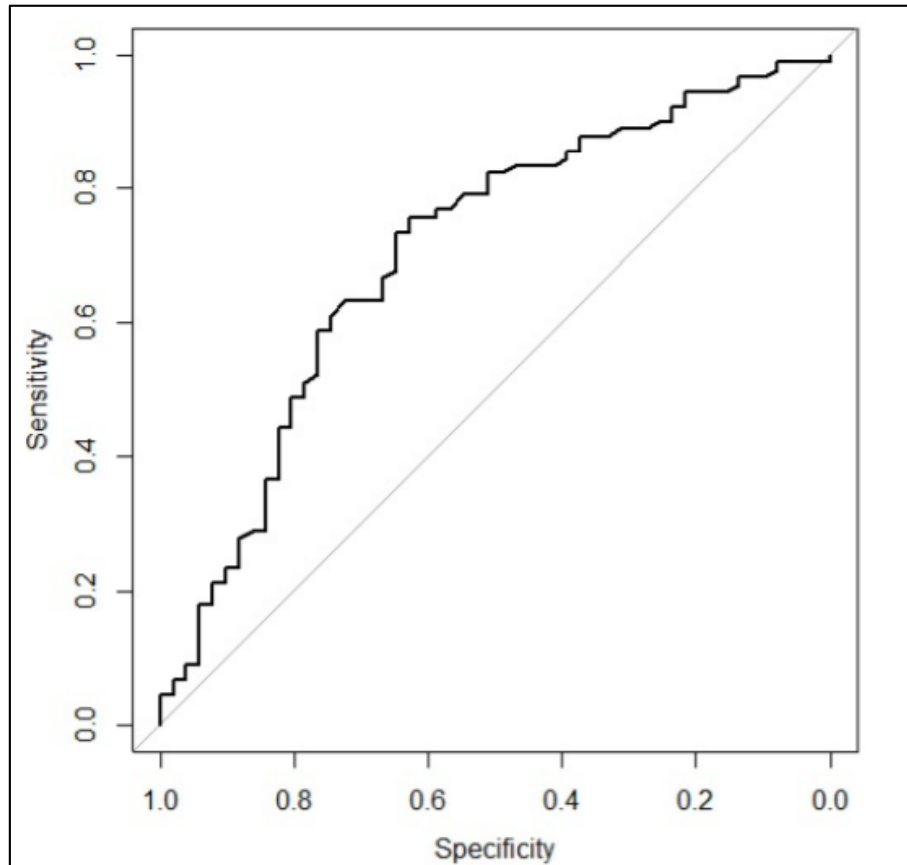


Figura 4. Curva ROC (*Receiver Operating Characteristic*). Eje y: sensibilidad. Eje x: especificidad.

El cálculo de AUC se lo realiza mediante una integral simple de la siguiente manera

$$AUC = \int_0^1 ROC(c)dc$$

El rango de valores va desde 0.5 hasta 1, si es menor no tiene capacidad discriminante y si es mayor los dos grupos están perfectamente diferenciados por la prueba. En una clasificación echa por (Swets, 1988) se muestra la interpretación de estos valores (Figura 5).

Baja exactitud [0.5 a 0.7]

Útiles para algunos casos [0.7 a 0.9]

Exactitud alta [0.9 a 1]

Figura 5: Interpretación por Swets del AUC.

Autor: (Swets, 1988).

2.5.3 Estadístico Kappa (k)

Es una medida de concordancia que determina el efecto del azar en la proporción de concordancia, permitiendo estimar la exactitud del modelo, toma en cuenta las proporciones de predicciones correctas que también se podría tener de manera aleatoria y se calcula usando los valores de la matriz de confusión de la siguiente forma

$$Kappa = \frac{P_0 - P_e}{1 - P_e} \quad (10)$$

Donde P_0 es la proporción de concordancia observada, P_e es la proporción de concordancia esperada por azar y $1 - P_e$ representa el acuerdo o concordancia máxima posible no debida al azar. De este modo Kappa corrige el acuerdo por azar por lo cual la proporción por acuerdo sobrepasa al acuerdo por azar, entonces si este valor se acerca a 1, la concordancia es perfecta, es decir, 100% de precisión y no existe azar, de otro modo si Kappa es 0 existe un total desacuerdo por lo que la proporción observada será igual a la proporción por azar (Cortez, 2010).

2.6 Normalización de variables

La normalización es un proceso mediante el cual se transforman o reescalan varios grupos de datos de diferente origen con la ayuda de un criterio, normalmente estadístico, esto con el fin de tener una misma escala o unidad de medida que permita realizar operaciones entre los diferentes

grupos de datos, dando como resultado grupos de datos homogéneos (Cubero, 2019).

2.6.1 Método de análisis jerárquico AHP (Thomas L. Saaty)

Este método fue desarrollado por Thomas L. Saaty en 1998 y está diseñado para resolver problemas con criterio múltiple. El método requiere que, quien tome las decisiones haga evaluaciones subjetivas respecto algún criterio en específico y asigne un valor numérico de importancia de criterio. El resultado del AHP es una jerarquización con valores numéricos que muestran las prioridades para cada una de las alternativas de decisión (Hurtado, 2019).

El método permite de manera eficiente y gráfica administrar la información respecto de un objetivo, descomponerla y estudiarla por partes, logrando visualizar los efectos de cambios en los diferentes niveles al momento de asignar un criterio, según la FAO (2000) éste se fundamenta en:

- Estructuración de modelo jerárquico (identificación de metas y criterios)
- Priorización de los elementos del modelo
- Comparaciones binarias entre los elementos
- Evaluación de los elementos mediante la asignación de pesos
- Ranking de las alternativas de acuerdo con los pesos dados
- Síntesis
- Análisis de sensibilidad

Las comparaciones pareadas y la asignación de valores por criterio subjetivo usan una escala subyacente con valores de 1 a 9 para calificar las preferencias relativas de los dos elementos. En

la Figura 6 se muestra los criterios usados por Thomas L. Saaty.

Planteamiento verbal de la preferencia	Calificación Numérica
Extremadamente preferible	9
Entre muy fuertemente y extremadamente preferible	8
Muy fuertemente preferible	7
Entre fuertemente y muy fuertemente preferible	6
Fuertemente preferible	5
Entre moderadamente y fuertemente preferible	4
Moderadamente preferible	3
Entre igualmente y moderadamente preferible	2
Igualmente preferible	1

Figura 6. Criterios de ponderación de Thomas L. Saaty.

Fuente: (Hurtado, 2019).

La matriz resultante está dada por $N \times M$ donde N son las alternativas y M son los criterios, si $N = M$ entonces se asigna el valor de 1 de la siguiente manera (Figura 7).

	Criterio 1	Criterio 2	...	Criterio m
Alternativa 1	P_{11}	P_{12}	...	P_{1m}
Alternativa 2	P_{21}	P_{22}	...	P_{2m}
...
Alternativa n	P_{n1}	P_{n2}	...	P_{nm}

Figura 7. Matriz de criterios de Thomas L. Saaty.

Fuente: (Hurtado, 2019).

La normalización de esta matriz a través de los pesos ayuda a encontrar la ponderación numérica asignada a cada alternativa de la siguiente manera (Figura 8).

$$\begin{pmatrix} 1 & w_1/w_2 & \dots & w_1/w_n \\ w_2/w_1 & 1 & \dots & w_2/w_n \\ \vdots & \vdots & \ddots & \vdots \\ w_n/w_1 & w_n/w_2 & \dots & 1 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} nw_1 \\ nw_2 \\ \vdots \\ nw_n \end{pmatrix} = n \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}$$

Figura 8. Obtención de la evaluación Thomas L. Saaty.

Fuente: (Hurtado, 2019).

Donde W_n es el valor numérico por cada alternativa usada, a partir de estos datos se puede normalizar una variable para escalar en un rango de 0 a 1 donde 0 es ninguna importancia y 1 es la mayor importancia indiscutible (Hurtado, 2019).

2.7 Variables climáticas Worldclim

Worldclim proporciona información climática en formato ráster de capas climáticas globales, estas capas se generan mediante la interpolación de datos climatológicos recogidos de estaciones meteorológicas de todo el mundo, obteniendo información con una resolución de hasta 30 arcos de segundo lo cual es aproximadamente 1 Km^2 en su mejor resolución. Las variables disponibles son precipitación y temperatura mensual, máxima y mínima, así como 19 variables bioclimáticas derivadas de estas (Worldclim, 2019). La interpolación de datos se genera usando la siguiente información:

- Bases de datos recopiladas en la Red Global de Climatología Histórica (GHCN), Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO), World Meteorological Organization (OMN), Centro Internacional de Agricultura Tropical (CIAT) y la Red Regional de Datos Hidrometeorológicos para América del sur, América Central y el Caribe (R-HYdronet) así como una serie de redes menores para Australia, Nueva Zelanda, Europa,

Ecuador, Perú y Bolivia, entre otros.

- Base de datos de elevación de Shuttle Radar Topography Mission (SRTM) resolución de 30 arcos de segundo
- El software ANUSPLIN el cual permite interpolar datos ruidosos múltiples utilizando splines de suavizado de placa delgada. Se usa la latitud, longitud y la elevación como variables independientes.

A continuación, se presenta un mapa proveniente de Worldclim donde se muestra la ubicación espacial de las estaciones climáticas de las cuales se obtuvo datos para la interpolación. Como se puede ver en la Figura 9 los datos de las estaciones climáticas del Ecuador continental e insular están incluidas.

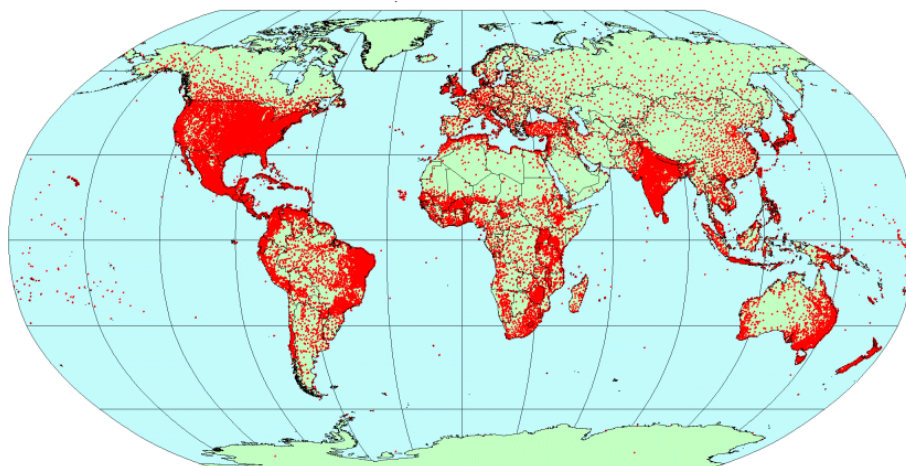


Figura 9. Ubicación de estaciones climáticas usadas por Worldclim (puntos rojos).

Fuente: (Worldclim, 2019).

2.8 Islas Galápagos

El archipiélago de Galápagos se ubica en el océano Pacífico a la altura de la línea ecuatorial y tiene su centro geográfico a $0^{\circ} 32.22'S$ y $90^{\circ} 31.26'O$. Está compuesto de 234 unidades terrestres

emergidas (islas, islotes y rocas) con una superficie terrestre total de 8010 km², aunque esta cifra puede variar por el carácter dinámico de los procesos geológicos que aún se mantienen en el archipiélago. El 93.2% de la superficie terrestre total se concentra en las islas Isabela, Santa Cruz Fernandina, Santiago y San Cristóbal, siendo Isabela la mayor con un tamaño del 58.7% de la superficie total (Galápagos, 2016). El origen de todas las islas del archipiélago Galápagos es volcánico debido a un levantamiento de la superficie marina ya que se ha confirmado que las islas no tienen ningún nexo con el continente. Las islas se levantan sobre una plataforma basáltica formada por lava submarina de diferente modo que la plataforma continental, esta teoría es la más aceptada entre la comunidad científica y puede ser verificada en la mayoría de las islas por la presencia de piroclastos acumulados alrededor de los conos volcánicos, el vulcanismo ha sido acumulativo en varias erupciones sucesivas y esta actividad no ha terminado hasta la actualidad (Martillo et al., 2011).

El clima de estas islas está determinado por las corrientes marinas y los movimientos del Frente Intertropical sobre el Pacífico oriental, así se determinan las dos estaciones principales una seca y la otra más húmeda, las cuales tienen un efecto muy importante en el comportamiento de la vegetación y la fauna (Grenier, 2007). Las estaciones están balanceadas en aproximadamente 6 meses:

- Estación caliente/húmeda: esta temporada va desde diciembre a mayo y la temperatura del aire es cálido y húmedo y va desde 27 a 31°C, los cielos están comúnmente despejados con lluvias fuertes ocasionales. Los vientos fuertes vienen desde el este y por lo general la corriente marina es más calmada (Foundation, 2019).
- Estación fría/ seca: esta temporada va desde junio hasta noviembre y el aire es más frío

alrededor de 24 a 16°C, los cielos pueden estar más nublados y las precipitaciones en las partes bajas son muy escasas, mientras que en las partes altas comúnmente existe una llovizna llamada “garúa”. Durante esta temporada los vientos vienen principalmente del sureste y el mar es más turbio. La temperatura del agua marina es más fría ya que la corriente de Humboldt predomina (Galapagos, 2019).

La biodiversidad puede variar de isla a isla y a escala global no se considera alta. Sin embargo, el endemismo y el estado de conservación en la mayor parte del archipiélago son altos. Como ejemplos más representativos, están las 14 especies de pinzones de Darwin, las tortugas gigantes, las *Opuntia* endémicas, entre otras (Galápagos, 2016). Los ecosistemas marinos de las Islas Galápagos están ecológicamente menos aislados que los terrestres por lo cual el nivel de endemismo es menor, pero son uno de los sistemas oceánicos más complejos y únicos en el mundo, ya que son un sistema interdependiente que presenta varias especies de fauna y flora de zonas tropicales y frías. En general el endemismo se presenta en mayor porcentaje en especies de reptiles, mamíferos terrestres, aves terrestres, aves y mamíferos marinos (Galápagos, 2016).

Las islas fueron descubiertas en 1535 por el obispo Fray Thomas de Berlanga de Panamá, quien intentaba realizar un viaje desde Panamá hasta Perú. En un comienzo fueron conocidas como “Las Encantadas” y fue usada como base de piratas y puerto del correo español hacia América del sur. Jackson (1993) menciona que las islas fueron nombradas Islas Galápagos por Abraham Ortelius en 1574, debido a que, el caparazón de las tortugas gigantes de las islas tenía un parecido con la silla de montar española llamada Galápagos. Como las tortugas podían sobrevivir durante meses sin agua ni comida fueron llevadas por miles, para suministros de carne en las embarcaciones. Charles Darwin visitó las islas en 1835 y fue esta visita la que haría que lo recordaran como uno de los

mejores científicos de todos los tiempos (Jackson, 1993).

En la actualidad las Islas Galápagos deben ser entendidas como un “socio-ecosistema” que implica que los ecosistemas y la sociedad deben concebirse y administrarse como un todo, por lo tanto, los procesos deben vincular a los sistemas tanto humanos como naturales, para que la capacidad de generación de bienes y servicios vaya en concordancia con la recuperación o no empeoramiento de los ecosistemas. Esto conlleva un gran esfuerzo tomando en cuenta que el 96.7% de la superficie terrestre de la provincia y el 100% de la zona marina dentro de una franja de 40 millas náuticas (medidas a partir de la línea base que se forman al unir los puntos más exteriores del archipiélago), se encuentran en alguna categoría de conservación (Galápagos, 2016).

Ecuador reconoce la importancia científica de las islas en 1934 declarando la legislación para la protección de la flora y fauna. Con los inicios de la segunda guerra mundial en 1942, Estados Unidos crea una base militar en la isla Baltra y controla cualquier tipo de ventaja en el canal de Panamá; adquiriendo de esta manera una importancia estratégica militar (Fitter & Hosking, 2016). En el año de 1959 las islas fueron declaradas Parque Nacional, se organizó la Fundación Internacional Charles Darwin para las Islas Galápagos y se fundó la Estación Científica Charles Darwin en la isla Santa Cruz, abriendo las puertas a científicos, periodistas, fotógrafos y misiones de todo el mundo. El 18 de febrero de 1973 el archipiélago es declarado una provincia de la República del Ecuador (Jackson, 1993).

2.8.1 Isla Santa Cruz

Santa Cruz, cabecera cantonal Puerto Ayora está ubicada en la parte central del archipiélago a $90^{\circ}10'15''\text{O}$ y $0^{\circ}22'21''\text{S}$, es la segunda isla más grande después de la isla Isabela con una extensión de 986 km^2 y una altitud máxima de 864 msnm. Además, es la más habitada y económicamente activa del archipiélago de las Galápagos. Es un volcán durmiente cuya última erupción se estima que ocurrió hace un millón y medio de años. Como testimonio de su pasado volcánico existen dos enormes cavidades llamadas "Los Gemelos" que se formaron a partir de una cámara de magma. La isla se caracteriza por su paisaje geológico y la variedad de su vegetación, distinguiéndose 7 zonas vegetales. La isla está llena de encantos naturales, hermosas playas, especies animales únicas, vegetación exuberante, cráteres y túneles de lava. Actualmente la población de la isla es de aproximadamente 15.701 habitantes (Cruz, 2019).

2.8.1.1 Zonificación vegetal

Las zonas de vegetación natural reconocidas son: (1) Zona litoral tolerante a la salinidad; aquí prevalecen matorrales y árboles pequeños de los cuales sobresalen los manglares. (2) Zona árida de cactus y *microphyllous* (Euphorbiales: Euphorbiaceae); sobresalen plantas espinosas e impenetrables, en esta zona ocurre el mayor porcentaje de endemismo correspondiente al 58%. (3) Zona mixta de transición; Bosques formados por árboles de mayor tamaño, la mayoría de estos puede perder las hojas durante época de mayor sequía. (4) Zona forestal siempre verde húmeda; prevalece la *Scalesia pedunculata* Hook.f (Asterales: Asteraceae), también llamada lechoso, esta zona ha sido la más alterada por ser la más propicia para la agricultura. (5) Zona arbustiva. (6) Pampa. (7) Zona alta árida. Aparte de estas también están las zonas no naturales como: (8) Asentamientos humanos y (9) Actividades agrícolas (ver Anexo 1). Todas dependen del gradiente

de elevación y la exposición del viento por lo tanto su distribución no es estricta. La zonificación es similar en islas grandes como San Cristóbal, Floreana, Santa Cruz, Isabela y Fernandina, las islas pequeñas no tienen esta zonificación (Peck, 2006).

2.8.1.2 Zonas ecológicas

Mientras que la vegetación natural divide a las islas en 7 zonas, las zonas ecológicas lo hacen en 3 zonas básicas. La primera es la zona costera caracterizada por animales y plantas las cuales dependen mucho de la proximidad del océano. La segunda es la zona árida y de transición, la cual se caracteriza por largos periodos de sequía y falta de humedad, los animales y plantas que se encuentran aquí pueden sobrevivir largos periodos sin agua como por ejemplo las tortugas. Y la tercera es la zona de tierras altas húmedas, donde puede haber periodos de sequía, pero por lo general las plantas y animales de esta zona están adaptados a una abundancia relativa de agua (Fitter, 2016).

2.8.2 Fragilidad del ecosistema isleño de Galápagos

Los ecosistemas insulares, al carecer de las presiones selectivas como el pastoreo por herbívoros o la presencia de depredadores, son altamente vulnerables a las invasiones biológicas. Debido a esto, los mamíferos introducidos en las islas en los dos últimos siglos han tenido una colonización altamente exitosa (Carrión, 2016). Conforme se incrementaron las actividades humanas, el número de especies de animales introducidos fue también creciendo, por ejemplo, cuando el científico inglés Charles Darwin visitó las islas, ya reportó en la isla Santiago la presencia de cerdos (Carrión, 2016). Los efectos de las especies introducidas son difíciles de predecir y en la mayoría de los casos, solo se las puede detectar una vez que se han establecido, es decir, cuando ha sido demasiado tarde para tratar de resolver el problema en su etapa inicial, cuando intenta adaptarse a las nuevas

condiciones de vida y los números poblacionales son bajos (Ministerio del Ambiente, 2016).

2.8.3 Especies invasoras

Las especies invasoras son animales, plantas u otros organismos que se desarrollan fuera de su área de distribución natural, en hábitats que no le son propios o con una abundancia inusual, produciendo alteraciones en la riqueza y diversidad de los ecosistemas, pueden provocar graves daños al convertirse en competidoras de las especies autóctonas (Bonilla, 2006). En las Islas Galápagos existen más de 1400 especies de plantas y animales exóticos que no son endémicas de la zona (Parque Nacional Galápagos, 2019). La erradicación de especies potencialmente invasoras es una estrategia clave para cualquier intento de mitigar impactos contra la biodiversidad de las Islas Galápagos (Estación Científica Charles Darwin, 2006). Por ello se implementó el Sistema de Inspección y Cuarentena para Galápagos (SICGAL), la idea es detectar organismos antes del ingreso hacia las islas, a través de controles en el puerto de Guayaquil y en los aeropuertos de Quito, Guayaquil y Manta. A manera de segunda barrera, se registra el equipaje de mano de todos los pasajeros al ingreso de los aeropuertos de Galápagos y de forma aleatoria la carga que ingresa por barco y avión. Como una tercera barrera se ha desarrollado una estrategia de detección temprana de nuevas introducciones, a través de monitoreos permanentes en las islas pobladas, en donde existe el mayor riesgo de introducción debido al movimiento permanente de bienes y servicios, que demanda el desarrollo de las actividades productivas. El objetivo de esta actividad es interceptar a tiempo las posibles introducciones de nuevas especies, de manera especial los invertebrados, que son los más difíciles de detectar (Ministerio del Ambiente, 2016).

2.8.4 Proyectos de control y erradicación de especies invasoras

En este punto se debe mencionar que las Islas Galápagos han tenido tres grandes proyectos de

control y erradicación de especies invasoras: Proyecto GEF para el manejo y control de las especies, que inicio actividades en 2003 y finalizo en el 2011, su objetivo principal fue la erradicación de chivos; Proyecto Pinzón, cuyo principal resultado fue la erradicación total de roedores de las islas Rábida y Pinzón; Control biológico de la escama algodonosa, mediante la utilización del agente de control biológico, *Rodolia cardinalis* Mulsatn 1850 (Coleoptera: Coccinellidae). Este proyecto ha resultado en la recuperación de plantas endémicas amenazadas (Ministerio del Ambiente, 2016).

- **Especie introducida**

El control de especies introducidas es fundamental para el mantenimiento de los ecosistemas ya que, al eliminar los espacios de estas especies introducidas, se permite el crecimiento y reproducción de las especies endémicas. Para el 2017 algunas especies de animales vertebrados e invertebrados fueron controlados y erradicados, disminuyendo su incremento con métodos de control y erradicación, también a las acciones de investigación que se ejecutan para conocer la ecología de estas especies y a futuro implantar nuevos métodos de control considerando que estas especies tienen una amplia distribución (Dirección del Parque Nacional Galápagos, 2017). Entre las especies introducidas más comunes están (Fabián & Lincango, 2006):

- Avispa del papel (*Polistes versicolor*)
- Avispa negra (*Brachygastra lecheguana*)
- Mosca de la fruta (*Anastrepha fraterculus*)
- Hormiga colorada (*Wasmannia auropunctata*)
- Hormiga negra (*Solenopsis geminata*)

2.9 Descripción de la especie *P. versicolor*

La especie que será estudiada en este proyecto será la avispa *P. versicolor* por esto se realiza una revisión bibliográfica de esta especie en particular. Las colonias de avispas *P. versicolor* se pueden encontrar en diferentes tipos de sustratos como hojas, ramas, raíces, piedras y también en nidos abandonados de otras especies de avispas sociales. Los nidos se construyen con material vegetal, se mastican y se mezclan con la secreción de las glándulas salivales, y el pedúnculo es resinoso (Figura 10). En áreas antrópicas, se ha observado la presencia de nidos que utilizan varias estructuras como sustratos de anidación (techos, vigas y cables entre otros) (Lasso, 1997). Sin embargo, el comportamiento y la biología de esta especie en el ambiente antrópico son poco comprendidos (Brito et al., 2015).

Las colonias pueden ser iniciadas por una sola hembra fertilizada (Haplometrosis), o por dos o más hembras (Pleometrosis) (Lasso, 1997). El desarrollo de *P. versicolor* es holometábolo (comprende los estados de huevo, larva, pupa y adulto) (ver Anexo 2), con cinco estados larvales (Brito et al., 2015). Al igual que otras avispas sociales, *P. versicolor* exhibe la distribución del trabajo, la cooperación entre los jóvenes y las generaciones que se superponen; estas características contribuyen a mantener la cohesión de la colonia. El factor principal que influye en la cohesión de la colonia de los insectos sociales es la comunicación, que puede llevarse a cabo a través de señales visuales, audibles, táctiles y semioquímicas, siendo la última la más efectiva, este tipo de comunicación es imprescindible para el forrajeo de la especie, comportamiento mediante el cual un animal busca alimento (Brito et al., 2015). Las horas con mayor presencia de avispas esta comprendida entre las 10:00 h a 13:00 h y en la isla Santa Cruz se ha registrado que las especies de plantas más visitada son nativas (48.2%) seguidas de plantas endémicas (28.6%)

y plantas introducidas (23.2%) (Lasso, 1997).



Figura 10. Nido de avispa *P. versicolor*.

Fuente: Dr. Oscar Mollá.

Polistes versicolor o avispa del papel es claramente identificable por su bandeado amarillo-negro en la parte abdominal, los ojos son negros y en la parte ventral oscuro verdosa, las mandíbulas son alargadas, fuertes y dentadas y se encuentran a los lados del clípeo, las alas presentan una alta especialización en su nerviación y su tamaño varia de 8 a 15 mm, las patas presentan un espolón o calcar en el ápice de la tibia anterior similar a un cuchillo (ver Anexo 3), la avispa completa mide aproximadamente mide 25 mm de largo en su etapa adulta y tiene un comportamiento agresivo con una picadura desagradable (Lasso, 1997).

Las crías en sus primeras etapas son alimentadas con néctar y miel de las flores, en la etapa adulta su dieta generalmente es a base de larvas y adultos de Lepidoptera, insectos como saltamontes y abejas mieleras. Avistada por primera vez en 1988 en la isla Floreana (Abedrabbo, 1991). Ahora se encuentra en islas habitadas y algunas islas inhabitadas y es más común en las

costas que en las partes altas de las islas (Fitter et al., 2016).

Actualmente se están realizando estudios para comprender las consecuencias que la presencia de la avispa pueda tener para los ecosistemas de las Islas Galápagos. La avispa *P. versicolor* se encuentra comúnmente en áreas habitadas por humanos (ej., sitios de visita, mercados, fincas) y su picadura puede causar reacciones alérgicas en algunas personas (Fundacion Charles Darwin, 2018). Los científicos han descubierto que la avispa se siente muy atraída por frutos fermentados y levaduras. Por esta razón, la levadura y otros atrayentes potenciales se están evaluando por investigadores para identificar un posible señuelo que sea eficaz en la atracción de avispas (Fundacion Charles Darwin, 2018).

2.10 Base Legal

Las Islas Galápagos han sido reconocidas como una prioridad de conservación por diferentes normas legales y a través de instrumentos de planificación no sólo a nivel nacional sino también a nivel mundial: La Ley Orgánica de Régimen Especial de Galápagos (LOREG), su denominación como el primer Patrimonio Natural de la Humanidad en 1978, la denominación de la Reserva Marina de Galápagos como el mayor bastión de vida silvestre en el Pacífico Oriental (PNG, 2005; SPNG –JICA, 2002) y su inclusión en la lista de Reserva de Biósfera, reconocen su singular valor natural, científico y educativo y se comprometen a conservarlo a perpetuidad para las presentes y futuras generaciones (LOREG, 1998), demostrando los esfuerzos que desde varios frentes, se concentran para lograr conservar la integridad ecológica que caracteriza a Galápagos.

El presente proyecto de investigación se fundamenta en el “Reglamento a la Ley de la Cartografía Nacional”, en mención a lo señalado en el Título I de Generalidades, Capítulo II de las Cartas y Mapas, en su artículo 18, literales d) y e), en que, consideran como plano de referencia de

las alturas, al nivel medio de los mares con su origen en La Libertad, y el uso de la Proyección Universal Transversa de Mercator (UTM) para elaborar productos cartográficos respectivamente.

2.11 Hipótesis

Es factible obtener un modelo de predicción espacial de la especie invasora *P. versicolor* de alta confiabilidad sobre la isla Santa Cruz, mediante el muestreo de presencia y ausencia de la especie, el modelamiento de variables espaciales y el uso de herramientas geoinformáticas.

2.12 Variables de la investigación

Las variables de investigación que se consideran para el desarrollo del presente proyecto de investigación se detallan en la Tabla 2.

Tabla 2
Descripción de las variables dependientes e independientes

Tipo	Nombre	Descripción	Subtipo	Unidades	Metodología de Observación	Fuente
Variable Dependiente	Presencia y ausencia de la especie	Obtenidas a través del muestreo sobre la isla Santa Cruz.	Cuantitativa continua	Ráster	Se obtuvo mediante el muestreo.	Autor
Variable Independiente	Humedad Relativa	Es la relación entre la fracción molar del vapor de agua en el aire y la fracción molar correspondiente si el aire estuviese saturado con respecto al agua (INAMHI, Anuario metereologico, 2014).	Cuantitativa Continua	Porcentaje	Estación meteorológica Bellavista isla Santa Cruz.	INAMHI
Variable Independiente	Precipitación	Cantidad de agua que cae a la superficie terrestre y proviene de la humedad atmosférica, ya sea en estado líquido o en estado sólido (INAMHI,	Cuantitativa Continua	Milímetros	Estación meteorológica Bellavista isla Santa Cruz.	INAMHI/ WORLDCLIM

CONTINÚA



		Boletín Agroclimático decadal, 2018).				
Variable Independiente	Temperatura del aire	Temperatura señalada en un termómetro expuesto al aire y protegido de la radiación solar directa, corresponde a los valores máximo y mínimo de temperatura del aire observadas durante un periodo de observación dado (INAMHI, Anuario meteorológico, 2014).	Cuantitativa Continua	Grados Celsius	Estación meteorológica Bellavista isla Santa Cruz.	INAMHI/ WORLDCLIM
Variable Independiente	Velocidad del viento	Movimiento del aire con respecto a la superficie de la tierra. La dirección se toma de donde proviene el viento (INAMHI, Boletín Agroclimático decadal, 2018).	Continua	$m\ s^{-1}$	Estación meteorológica Bellavista isla Santa Cruz.	INAMHI/ WORLDCLIM
Variable Independiente	Radiación solar	Es la energía emitida por el sol en forma de energía electromagnética que llega a la atmósfera. Se mide de forma	Continua	$kJ\ m^{-2}\ dia^{-1}$	Estación meteorológica Bellavista isla Santa Cruz.	INAMHI/ WORLDCLIM

CONTINÚA



		horizontal con la ayuda de un sensor de radiación o piranómetro, el cual se coloca en un lugar libre de sombras (Navarra, 2019).				
Variable independiente	Altitud	Superficie de referencia aproximada al geoide definida mediante el promedio del nivel de los océanos registrados por un tiempo prolongado (≥ 1 año), a través de un mareógrafo (Torge, 2001). En Ecuador es el Datum vertical de La Libertad, obtenido entre 1960 y 1961 con datos del mareógrafo de las campañas 1950 a 1959 y 1949 a 1967, con diferencia de 0.8 mm entre estos (Paredes, 1986).	Cuantitativa Continua	Metro	Fue definida mediante registro de datos de mareógrafos con el promedio del nivel del mar.	INOCAR/IGM
Variable Independiente	Uso de suelo	En Galápagos existen 560 especies de plantas nativas (180 endémicas), adicionalmente,	Cualitativa Discreta	Atributo	Base de Datos	MAE

CONTINÚA



		700 especies han sido introducidas en las islas como resultado de la colonización y las actividades antrópicas.				
Variable Independiente	Proximidad a estructuras antrópicas	Distancia existente desde los puntos de interés hasta la infraestructura existente en la isla.	Cuantitativa continua	Metro	Se obtuvo mediante la base de datos proporcionada por el IGM.	IGM
Variable Independiente	MDE	Modelo digital de elevaciones.	Cuantitativa continua	Metro	Se obtuvo de la base de datos del IGM.	IGM

CAPITULO III

METODOLOGÍA

3.0 Resumen

En este capítulo se muestra la planificación del muestreo de campo la cual consta de 3 partes: determinación de áreas con mayor probabilidad de encontrar la especie, método de recopilación de datos y muestreo por transectos. Posterior a esto el tratamiento de las variables dependientes (presencia y ausencia) y variables independientes (predictoras). También se detalla la aplicación de los modelos Maxent, MARS, BIOCLIM y Regresión Logística con el fin de encontrar las áreas con mayor probabilidad para la distribución de la especie *P. versicolor* dentro de la zona de estudio en la isla de Santa Cruz, así mismo se muestran los pasos para usar el software necesario, así como la explicación de sus insumos y sus productos más relevantes como: curva ROC, matriz de confusión, estadístico Kappa, entre otros.

3.1 Metodología de desarrollo del proyecto

La metodología del proyecto se compone de 4 fases entre las que se encuentran: selección del área de estudio, planificación del muestreo y recolección de datos, procesamiento de los datos y validación de resultados. En la Figura 11, se presenta el flujograma de la metodología que se desarrolló en el presente proyecto de investigación, para generar el modelo de distribución de la especie en la zona de estudio.

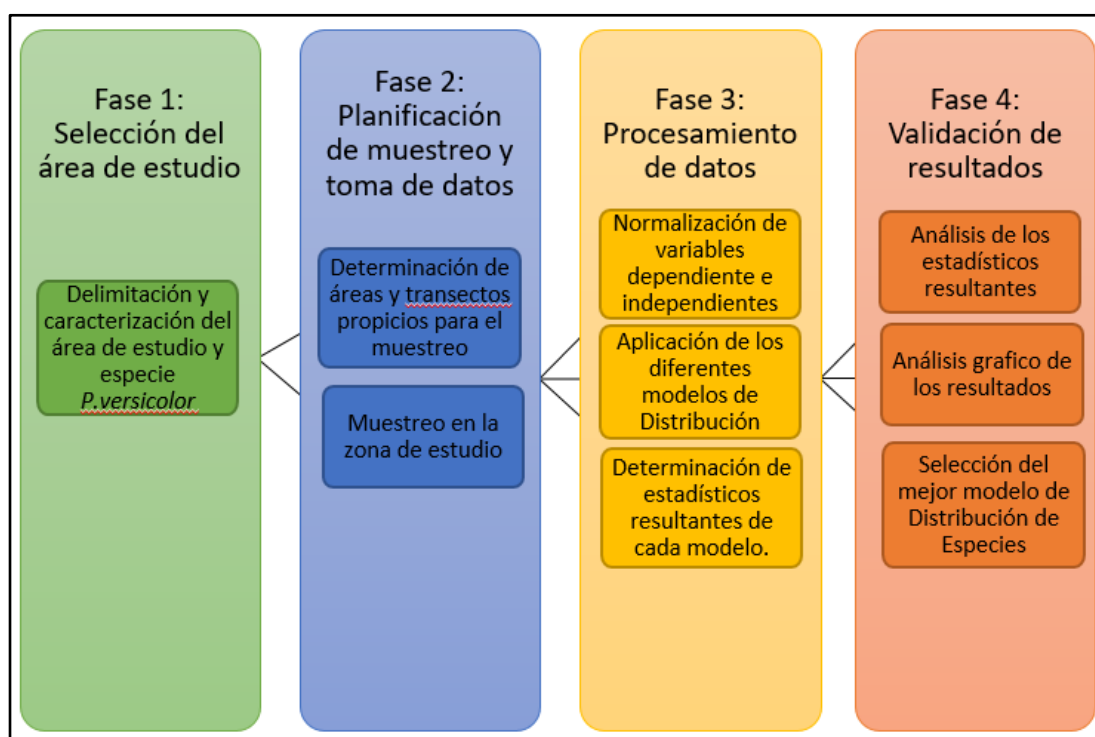


Figura 11. Flujograma de la metodología usada para el proyecto.

3.2 Planificación de toma de datos

La especie *P. versicolor* se distribuye por varios ecosistemas y no presenta un patrón de distribución específico (uniforme, aleatoria o agrupada) por lo cual, la recolección de presencias y ausencias dentro del área de estudio se dispuso en 3 etapas: Determinación de áreas con mayor probabilidad de avistamiento, método de recolección de datos y aplicación del método. Estas etapas constan de trabajo de gabinete y trabajo de campo. A continuación, se describen las etapas realizadas.

3.2.1 Determinación de áreas con mayor probabilidad de avistamiento

Como se menciona en el capítulo anterior (ver apartado 2.9) los nidos son de fácil localización y están contruidos en lugares protegidos como vegetación y construcciones hechas por el hombre o en sitios escondidos como desvanes, agujeros de árboles, cavidades en la superficie, rocas. Por

este motivo para la determinación de áreas con mayor probabilidad se establecieron las zonas vegetales que ofrecen mayor protección a los nidos y las áreas de construcciones antrópicas, de este modo se determina un área donde la *P. versicolor* tiene más posibilidades de ser avistada. Por esto, se tomó desde la zona litoral hasta la zona arbustiva incluyendo la zona de agricultura y de construcción antrópica como zonas de prioridad para poder realizar los transectos de muestreo. En el software Arcgis a partir de las curvas de nivel de la isla Santa Cruz se generó un MDT y se realizó una clasificación de la isla Santa Cruz para establecer las zonas de vegetación de prioridad y de este modo tener una guía para verificar que los puntos de muestreo abarquen las zonas de prioridad de avistamiento. Sin embargo, los puntos de ausencia también son importantes por tal motivo también se realizaron transectos de muestreo fuera de las zonas de prioridad. En las Islas Galápagos, al ser un área protegida por el Parque Nacional, no se permite el acceso a zonas que no tengan un camino o sendero previamente establecido, por este motivo no se puede realizar transectos transversales a los caminos; por lo que se usó como transectos los caminos y vías de libre acceso en los cuales se puede realizar el muestreo mediante una caminata.

3.2.2 Método de recopilación de datos

El muestreo por transectos es un procedimiento en el cual se realizan recorridos a lo largo de una o varias líneas predeterminadas y a una velocidad constante y se intenta detectar la presencia del individuo (visto u oído) (Castro, 2006). Por lo cual el avistamiento de un individuo de la especie *P. versicolor* en cualquier espacio ya sea natural o antrópico se determina como un punto de presencia. Por esto la recopilación de datos fue realizada mediante un muestreo de campo usando el método observacional y georreferenciando los puntos de muestreo (presencia o ausencia)

mediante GPS, así un punto de presencia es tomado tanto por la observación de la especie en proceso de forrajeo (ver apartado 2.9), como también por la presencia de nidos.

Los datos de las variables independientes fueron recopilados mediante el acceso a las bases de datos libre del IGM y el INAMHI y solicitud de información a los laboratorios de Geomática, Sensores Remotos de la ESPE y Estación Científica Charles Darwin, así como también las 19 variables climatológicas de Worldclim con 1 km² de resolución.

3.2.3 Aplicación del muestreo por transectos

El muestreo fue determinado de 3 maneras, en las cuales de forma visual se determinó la presencia o ausencia de nidos de la avispa o avispas en actividad de forrajeo. El punto de observación fue georreferenciado con GPS en el sistema de coordenadas UTM 15 S:

- a) Los transectos en franja fueron seleccionados en las diferentes zonas altitudinales y de vegetación representativas de la isla Santa Cruz. Estos transectos tuvieron una dimensión variable y fueron trazados a lo largo de caminos, senderos turísticos y trochas existente, las cuales se pueden recorrer con el permiso del Parque Nacional y acompañado de un guardaparque. No se ingresó a propiedades privadas, sin embargo, el recorrido fue suficiente para recoger la información necesaria.
- b) Realizando recorridos por vías entre los sitios más representativos de la isla Santa Cruz como lo son las carreteras entre centros poblados Puerto Ayora-Bellavista; Bellavista-Santa Rosa; Santa Rosa- Canal de Itabaca y Bellavista-Garrapatero.
- c) En centros poblados y zonas agrícolas registrando la presencia o ausencia de nidos de la avispa e individuos volando. En la isla Santa Cruz se lo realizó en las zonas pobladas de Puerto Ayora,

Bellavista y Santa Rosa. Las viviendas fueron seleccionadas al azar dependiendo de la disponibilidad para la observación de tejados.

De este modo con la ayuda del mapa base WorldImaginary de Arcgis y con la experiencia en las islas del Dr. Oscar Mollá, se realizó una vectorización de los transectos por donde hay accesibilidad y se pueden realizar los muestreos. Para la planificación se generaron 7 zonas las cuales fueron recorridas en 1 o 1.5 días aproximadamente cada una. Estas zonas cubren centros poblados, así como también áreas cultivadas accesibles y las áreas determinadas como áreas de mayor probabilidad. Los transectos que se encuentran fuera de estas zonas fueron recorridos de forma longitudinal ya sea a pie o bicicleta y van desde zonas bajas como Puerto Ayora o Tortuga Bay, hasta zonas más altas como los cerros “Crocker” o “el Puntudo”. A continuación, en la Figura 12 se presenta el área de estudio con las zonas determinadas con mayor probabilidad de avistamiento (litoral, transición, siempreverde, pampa y zona alta árida), 7 zonas generadas (polígonos color azul) y la totalidad de transectos y vías recorridos (líneas de color rojo y negro).

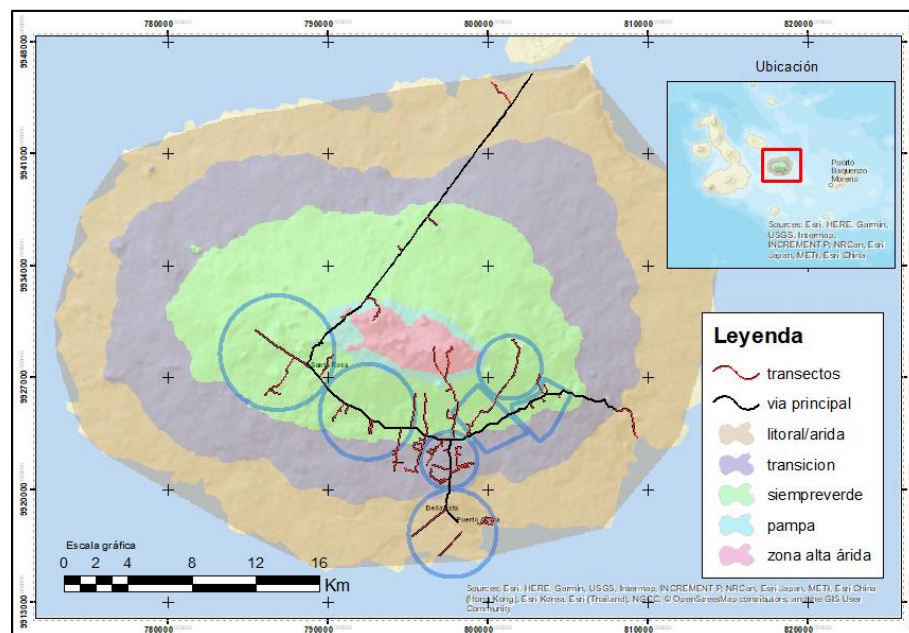


Figura 12. Área de estudio y planificación de muestreo.

3.3 Variables dependientes

3.3.1 Datos de presencia de *P. versicolor*

El trabajo fue realizado desde el 19 de junio hasta el 9 julio donde se logró recopilar los datos de presencia, con 1598 puntos registrados sobre la isla Santa Cruz entre observaciones de nidos y avispas en actividad de forrajeo. Con la ayuda del GPS Garmin modelo OREGON se registró la coordenada geográfica en sistema de coordenadas UTM 15 S (ver Anexo 4). En la Tabla 3 se muestra el cronograma de actividades durante las semanas de muestreo:

Tabla 3

Cronograma de actividades durante el muestreo

Fecha	Actividad	Resultado
19/06/2019	Llegada a las islas y recopilación bibliográfica de la avispa y área de estudio. En la biblioteca de la Estación Charles Darwin. Libros: Atlas de Galápagos y Galapagos Preserving	Descripción de comportamiento de la especie en el Manual de monitoreo y vigilancia de invertebrados introducidos.
20/06/2019	Recopilación bibliográfica de la avispa y área de estudio. En la biblioteca de la Estación Charles Darwin. Libros: Atlas de Galápagos y Galapagos Preserving	Descripción de comportamiento de la especie en el Manual de monitoreo y vigilancia de invertebrados introducidos.
21/06/2019	Muestreo en la zona central urbana de Puerto Ayora	20 puntos de presencia entre nidos y avistamientos
22/06/2019	Muestreo en el área de la estación Charles Darwin y oficinas del Parque Nacional Galápagos	211 puntos de presencia entre nidos y avistamientos

CONTINÚA



23/06/2019	Muestreo en la parroquia de bellavista	233 puntos de presencia entre nidos y avistamientos
24/06/2019	Muestreo en la parroquia de Bellavista y parte alta de Puerto Ayora	136 puntos de presencia entre nidos y avistamientos
25/06/2019	Muestreo de la vía Puerto Ayora – Bellavista y sus alrededores	35 puntos de presencia entre nidos y avistamientos
26/06/2019	Muestreo de Kroker y Puntudo	85 puntos de presencia entre nidos y avistamientos
27/06/2019	Muestreo de la Parroquia Santa Rosa	87 puntos de presencia entre nidos y avistamientos
28/06/2019	Muestreo de la vía Santa Rosa – Canal	158 puntos de presencia entre nidos y avistamientos
29/06/2019	Muestreo del camino hacia las Grietas	7 puntos de presencia entre nidos y avistamientos
30/06/2019	Descanso	
01/07/2019	Muestreo de los ranchos en las zonas agrícolas	15 puntos de presencia entre nidos y avistamientos
05/07/2019	Muestreo de la vía al Garrapatero	134 puntos de presencia entre nidos y avistamientos
06/07/2019	Muestreo de ranchos en el área entre Bellavista y Santa Rosa	50 puntos de presencia entre nidos y avistamientos
07/07/2019	Descanso	
08/07/2019	Descanso	
09/07/2019	Viaje de regreso	

En la Figura 13 se presenta los puntos de muestreo de la avispa *P. versicolor*, sobre la isla Santa Cruz.

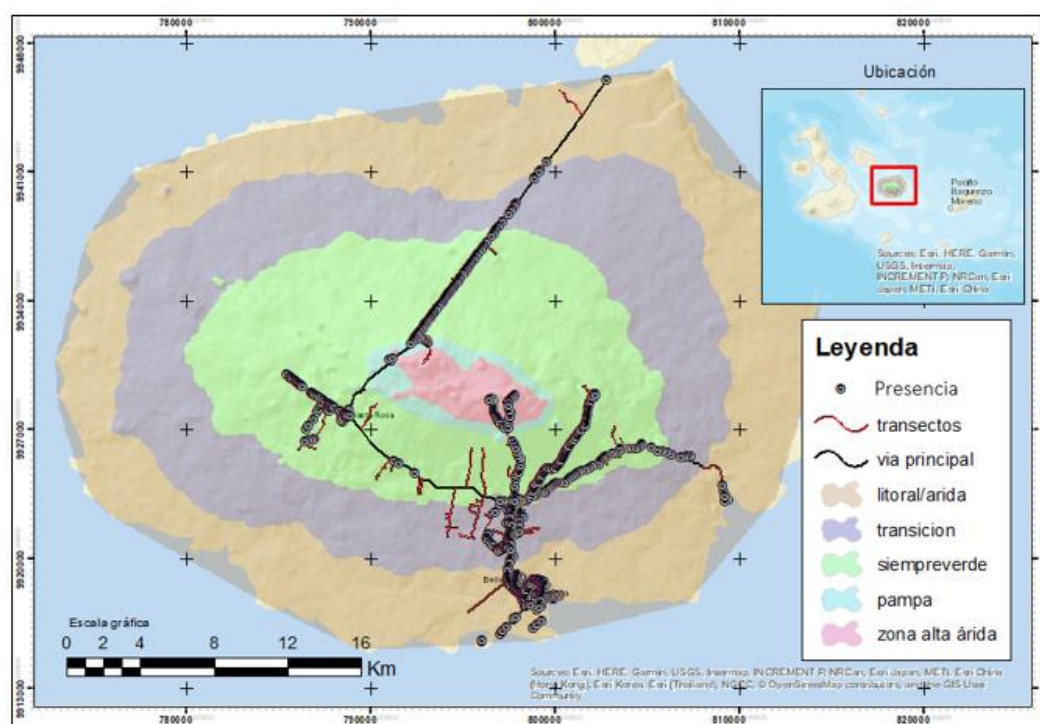


Figura 13. Puntos de muestreo sobre el área de estudio.

3.3.2 Datos de pseudo-ausencias de *P. versicolor*

Los datos de ausencia son de mucho valor al momento de realizar un modelo de distribución de especies, como se menciona en el capítulo anterior (ver apartado 2.4.2). Actualmente no existen estudios que determinen las zonas no aptas para la distribución de *P. versicolor* ya que existe relativamente muy poco conocimiento acerca de su comportamiento. Por este motivo se crearon pseudo-ausencias las cuales fueron estimadas a partir de la experiencia de expertos sobre el campo.

Se realizó una delimitación de las zonas que se consideran poco favorables para el desarrollo de *P. versicolor*, con la ayuda de la herramienta *buffer* de ArcMap se creó un área alrededor de las

presencias de nidos procurando no crear conflicto con zonas que si presentan presencia. Se escogió una distancia de 5 km ya que se considera que en esta distancia los factores ambientales pueden ser notablemente similares. En otras especies como la avispa *Vespula Germanica* Fabricius 1793 (Hymenoptera: Vespidae) se sabe que la distancia de vuelo en forrajeo puede ser de hasta 6 km según su masa corporal (Masciocchi, 2013). Con la ayuda de la herramienta *random point* se crearon puntos aleatorios fuera de estas zonas delimitadas. A continuación, en la Figura 14 se presenta el proceso usado para generar los puntos de pseudo-ausencias el cual se compone de 4 pasos: 1) Puntos de presencia (Figura 14.1). 2) Generación de buffer de 5 km (Figura 14.2). 3) Generación de puntos aleatorios fuera de los buffers de 5 km (Figura 14.3). 4) Resultado puntos de pseudo-ausencias (Figura 14.4).

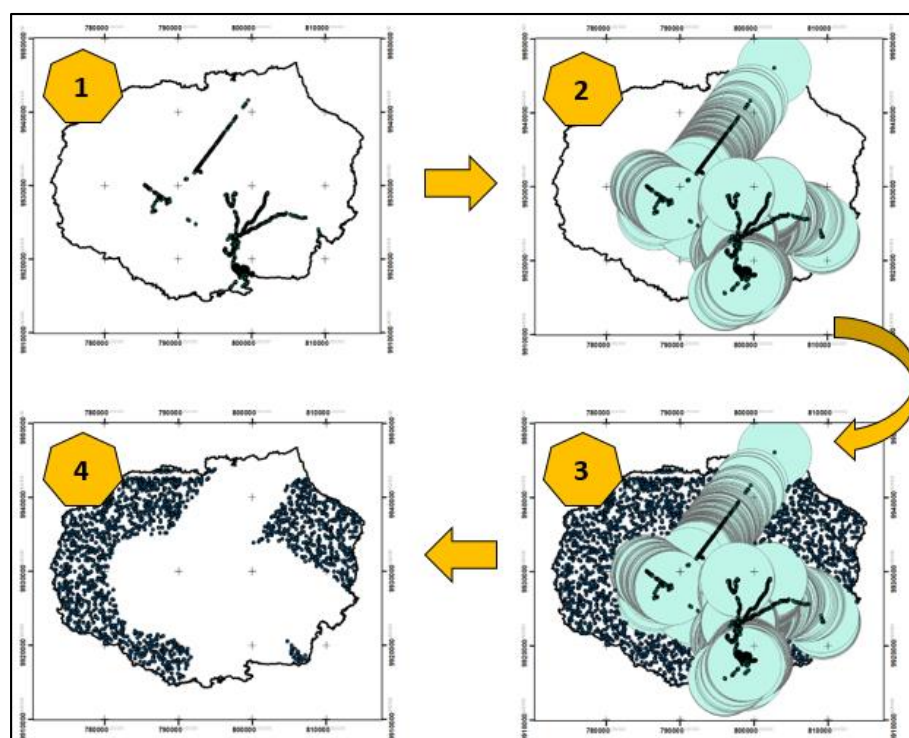


Figura 14. Proceso para generar los puntos de pseudo-ausencias.

3.4 Variables independientes

3.4.1 Variables climáticas

Las variables independientes son usadas como las variables predictoras de un modelo, es decir estas variables serán correlacionadas con presencias y ausencias y así se podrán determinar los espacios geográficos donde existe un ambiente más adecuado y potencialmente más apto para su habitabilidad (ver apartado 2.4.2). Entre estas variables se usaron 5 variables climáticas las cuales fueron tomadas del sitio Worldclim (<http://worldclim.org/>). Worldclim reúne datos de estaciones climáticas de todo el mundo y genera modelos de variables climáticas a partir de la interpolación de datos mensuales, trimestrales y anuales (Worldclim, 2019). Se encuentran a diferentes resoluciones espaciales entre las cuales está; 30 segundos que equivalen a aproximadamente a 1 km² sobre la superficie terrestre. Esta resolución fue usada para este trabajo ya que es la mejor resolución disponible en nuestra área de estudio (isla Santa Cruz) que presenta una superficie relativamente pequeña de 986 Km². Las capas que ofrece el portal se encuentran en formato TIFF por cada mes, es decir, existen 12 ráster por cada variable. Las variables que se incluyeron fueron: Temperatura media (°C), Precipitación (mm), Radicación solar (kJ m⁻² día⁻¹), Velocidad del viento (m s⁻¹) y Presión de vapor de agua (kPa). Con estas variables se realizó el recorte del área de estudio de Santa Cruz mediante la herramienta *extract by mask* usando un valor de píxel de 80 y el mismo número de filas y columnas (534x404). Es de suma importancia que estas dimensiones sean las mismas para todas las variables usadas al momento de trabajar en formato ráster.

Al tener rásters de todos los meses se realizó un promedio con la ayuda de la herramienta *raster calculator* de los meses entre junio-noviembre los cuales responden al semestre más frío del año y

entre los cuales se realizó el muestreo de campo (ver apartado 2.8). A continuación, en la Figura 15, se muestran los rásters de las variables usadas.

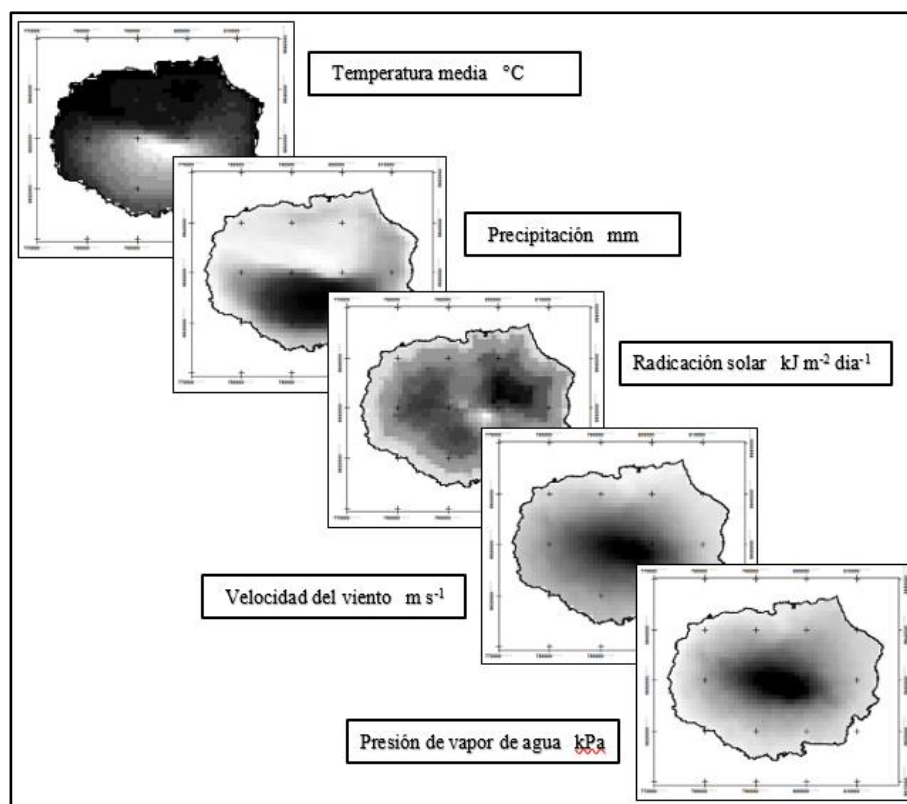


Figura 15. Rásters de las variables climáticas usadas.

Cabe mencionar que no se usó el paquete de 19 variables que ofrece el sitio web ya que estas no corresponden a una misma temporalidad por lo cual no era factible usarlas para promediar entre los meses de junio-noviembre.

Para realizar una validación de las variables climáticas se usó la base de datos de las estaciones meteorológicas de Bellavista y Puerto Ayora obtenidas de la página oficial de la Estación Charles Darwin (Foundation, 2019), con las cuales se realizó un promedio histórico de la temperatura media

(variable disponible) entre los meses de junio y noviembre y se comparó con los valores de la variable climatológica de Worldclim.

Como se puede ver en la Tabla 4 la desviación estándar es de 0.09°C por lo cual la precisión de la variable temperatura de Worldclim en las estaciones climatológicas de la isla Santa Cruz es muy aceptable.

Tabla 4

Comparación de la variable temperatura entre Worldclim y estaciones de la isla Santa Cruz

	E. Bellavista	E. Puerto Ayora
Estación	20.99	22.35
Worldclim	20.91	22.25
Diferencia	0.08	0.1
Desviación estándar		0.09

3.4.2 Altitud

La información correspondiente a la altitud fue recopilada del sitio web del Instituto Geográfico Militar (IGM), el cual proporciona las curvas de nivel de las Islas Galápagos con un desnivel de 20 m. Mediante el software ArcMap se creó el modelo digital de elevaciones en formato ráster y sobre este se realizó el recorte del área de estudio es decir la isla Santa Cruz, como se puede observar en la Figura 16.

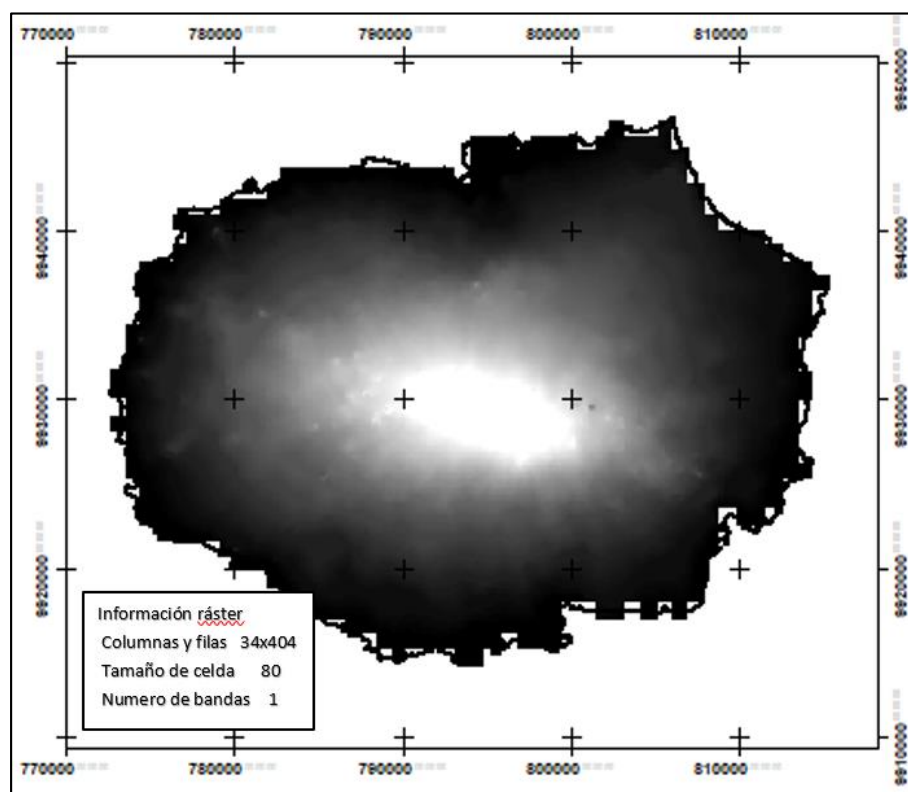


Figura 16. Variable altitud generada a partir del modelo digital de elevaciones. Mayor altitud (blanco), menor altitud (negro).

3.4.3 Infraestructura

La información fue recogida del geportal del Instituto Geográfico Militar (IGM) y se encuentra en formato shape file, por lo cual se utilizó la herramienta *Euclidian Distance* de ArcMap para crear una variable que represente la distancia más cercana a una infraestructura antrópica. Esta variable servirá para determinar cuán importante es la cercanía de la infraestructura antrópica para la presencia de la *P. versicolor*. Esta herramienta crea un ráster con las distancias en metros, desde cualquier parte del ráster al punto de infraestructura más cercano (Figura 17).

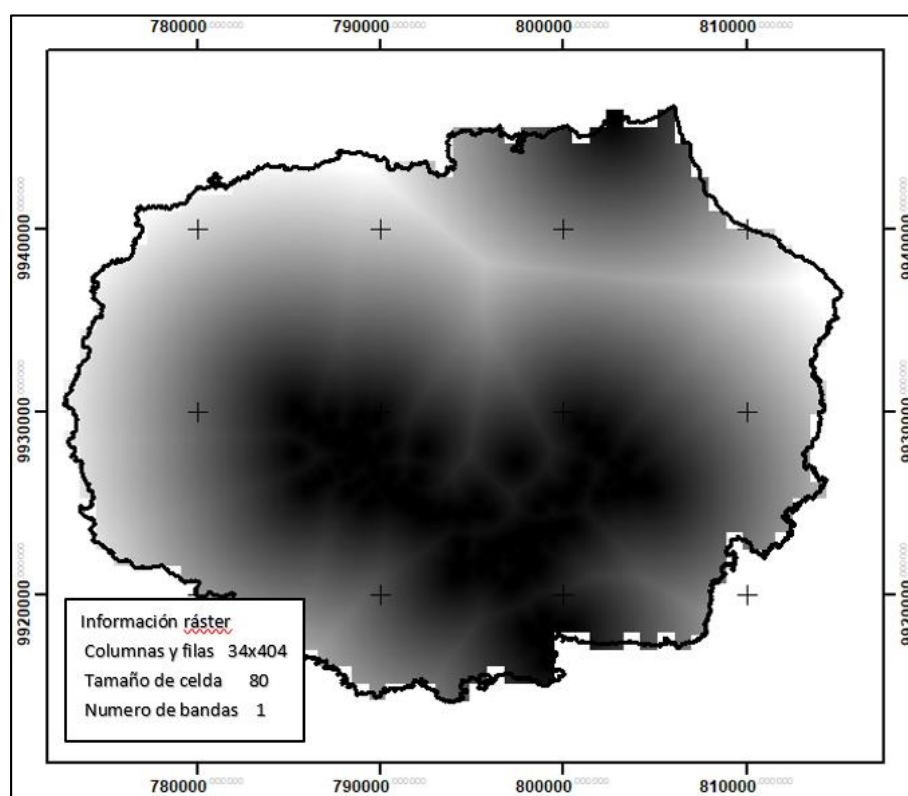


Figura 17. Variable distancias a infraestructura antrópica. Menor distancia (negro), mayor distancia (blanco).

3.4.4 Cobertura vegetal y uso de suelo

Para la variable uso de suelo se realizó el método analítico Thomas L. Saaty, el cual se basa en ponderar numéricamente cada componente o variable en función de la importancia que esta representa frente a las demás. En este caso la presencia de la avispa sobre el terreno según el muestreo realizado, para obtener un modelo jerárquico que permite determinar qué tipo de suelo tiene más importancia para la presencia de la especie. En la Tabla 5 se muestra la ponderación que se realizó para determinar el peso de cada componente de uso de suelo sobre la isla Santa Cruz y en la Tabla 6 los parámetros ya normalizados de la variable donde a mayor porcentaje indica más probabilidad de encontrar la especie.

Tabla 5*Matriz Thomas L. Saaty para la variable uso de suelo*

VARIABLES	V	V1	V2	V3	V4	V5
Agropecuario	V1	1.00	7.00	2.00	1.00	5.00
Eriales	V2	0.14	1.00	0.14	0.14	0.20
Rasgos culturales	V3	0.50	7.00	1.00	1.00	5.00
Vegetación invasora	V4	1.00	7.00	1.00	1.00	5.00
Vegetación natural	V5	0.20	5.00	0.20	0.20	1.00

Tabla 6*Parámetros normalizados de la variable uso de suelo*

VARIABLES	V	V1	V2	V3	V4	V5	Peso	%
Agropecuario	V1	0.35	0.26	0.46	0.30	0.31	0.34	33.59
Eriales	V2	0.05	0.04	0.03	0.04	0.01	0.04	3.51
Rasgos culturales	V3	0.18	0.26	0.23	0.30	0.31	0.25	25.46
Vegetación invasora	V4	0.35	0.26	0.23	0.30	0.31	0.29	28.98
Vegetación natural	V5	0.07	0.19	0.05	0.06	0.06	0.08	8.46

A partir de estos resultados se transformó el shape de uso de suelo a un ráster en función de los pesos, el resultado es un ráster con valores de celda entre 0 y 1 (Figura 18).

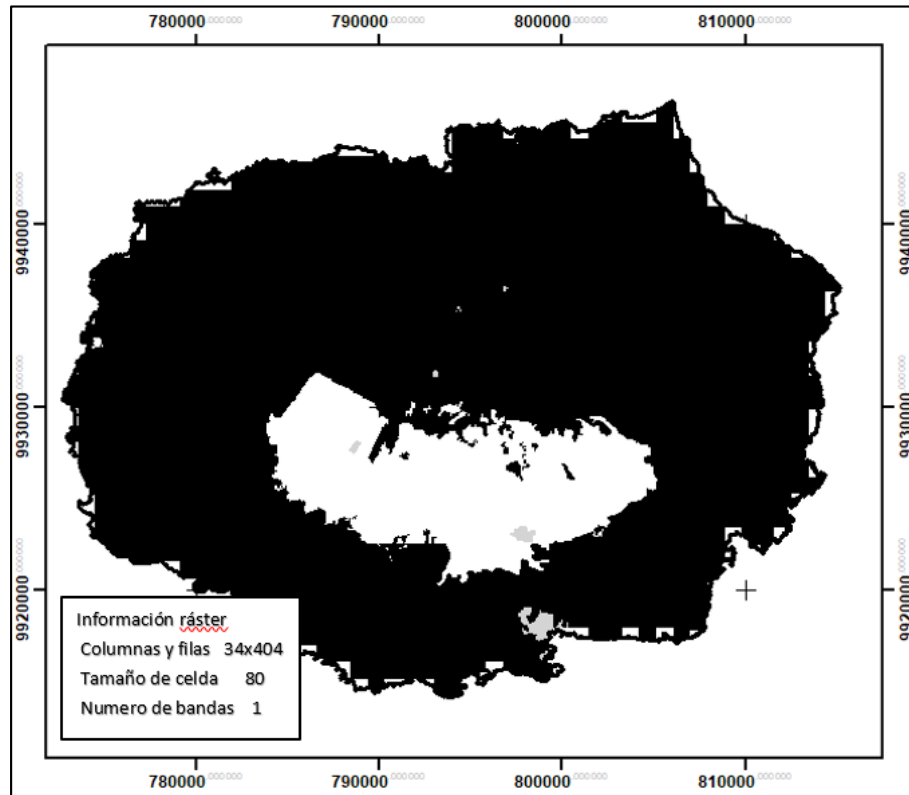


Figura 18. Ráster de la variable uso de suelo normalizada.

Las variables utilizadas para el modelamiento de *P. versicolor* fueron una combinación entre variables bioclimáticas, variables topográficas y variables culturales como se muestran en la Tabla 7.

Tabla 7
Variables independientes o predictoras usadas

Variable	Descripción	Tipo
Temperatura (°C)	Temperatura media entre los meses de junio-noviembre	Bioclimática
Precipitación (mm)	Precipitación media entre los meses de junio-noviembre	Bioclimática
Radicación solar (kJ m⁻² día⁻¹)	Radiación media entre los meses de junio-noviembre	Bioclimática
Velocidad del viento (m s⁻¹)	Velocidad media entre los meses de junio-noviembre	Bioclimática
Presión de vapor de agua (kPa)	Presión media entre los meses de junio-noviembre	Bioclimática
Altitud (m)	Altura referida al nivel medio del mar	Topográfica
Infraestructura	Distancia a infraestructura antrópica	Cultural
Uso de suelo	Estado de la distribución de uso de suelo sobre la Isla Santa Cruz	Cultural

Estas variables fueron seleccionadas por su disponibilidad procurando cubrir los aspectos que conforman el medio ambiente de las Islas Galápagos, es decir, aspecto bioclimático, topográfico y cultural. La importancia de cada una de estas variables podrá variar según los estadísticos que use cada modelamiento.

3.5 Normalización de las variables

Como se había mencionado anteriormente (ver apartado 2.6), la normalización de las variables es necesario en ciertos modelos de distribución que requieren que las variables tanto dependientes como independientes tengan una misma escala, por este motivo se normalizaron las variables en una escala de 0 a 1 en función de la probabilidad de encontrar a la especie dentro de este rango, es decir, existe mayor probabilidad de presencia mientras más cercano este el valor a 1.

Con la ayuda de los puntos de muestreo y todas las variables en formato ráster se encontró los máximos y mínimos de cada variable en los puntos de muestreo, esto gracias a la ayuda de la herramienta *Extract Multi Value Points* de ArcMap. De esta manera se pudo considerar los valores extremos de supervivencia de la especie dentro de la muestra, donde se considera que la media de este rango será el lugar óptimo para el desarrollo de la especie, por lo cual la normalización se la realiza hacia la media, se asignan valores cercanos a 1 mientras más cercanos estén los valores a la media del rango, teniendo la media un valor de 1 es decir 100% de probabilidad de encontrar a la especie. En la Tabla 8 se muestran los valores máximos y mínimos tanto del muestreo como los valores de la variable en general.

Tabla 8

Valores máximos y mínimos de cada variable

Variable	General		Muestras	
	Min	Max	Min	Max
Altitud	10.00	840.00	10.00	840.00
Viento	4.13	8.52	4.22	4.52
Vapor	1.86	2.29	1.86	2.27
Temperatura	18.45	23.35	18.45	23.35
Solar	12884.00	16617.00	12929.00	16614.00
Precipitación	2.00	17.00	2.00	17.00
Edificios	0.00	15242.90	0.00	9661.46

Ya que se obtuvo los valores máximos y mínimo de la muestra sobre cada variable se utiliza la siguiente fórmula para poder ajustar la probabilidad hacia la media del rango:

$$x_n = 1 - \frac{|max+min-2*xi|}{max-min} \quad (11)$$

Donde x_n es el valor normalizado, max es el máximo del rango, min es el valor mínimo del rango y xi es la variable por normalizar. Cabe mencionar que cualquier valor que este fuera del rango de max y min tendrá un valor negativo por lo cual con la ayuda de *ráster calculator* para la normalización de los ráster y usando el condicional “Con” de la herramienta, se logra que cualquier valor negativo resultante de la formula sea automáticamente devuelto como 0, es decir, sin probabilidades de encontrar a la especie (Figura 19).

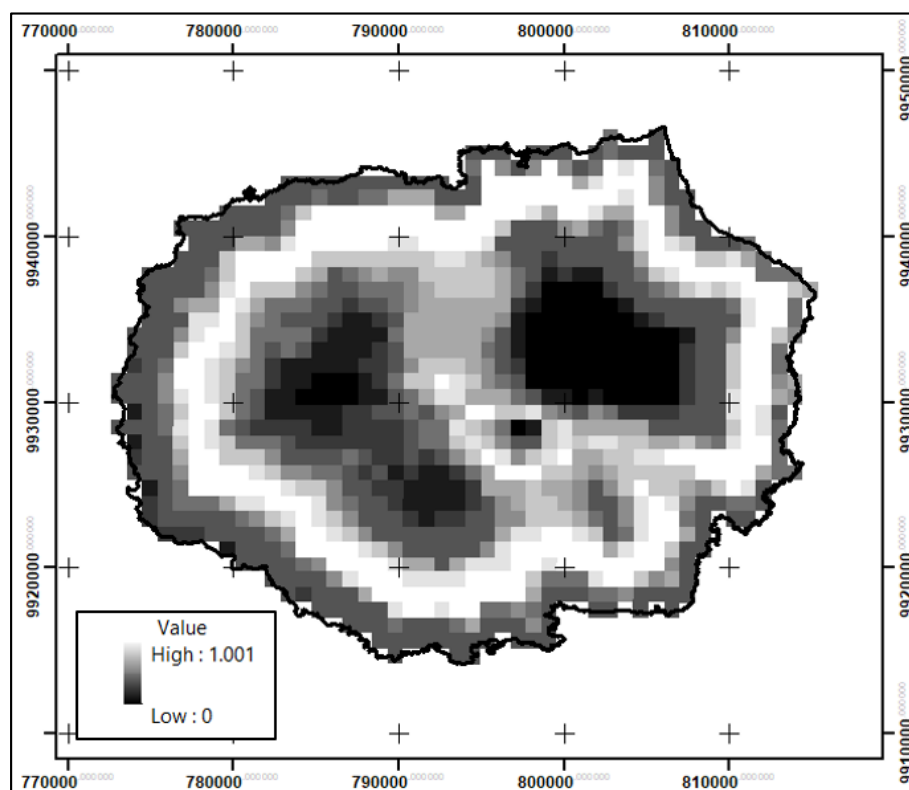


Figura 19. Ejemplo de variable normalizada entre los valores de 0 a 1. Mientras mayor sea el valor de pixel, mayor será la probabilidad de encontrar a la especie.

3.6 Modelo de Distribución de Especies

Se usaron 4 modelos (Maxent, MARS, BIOCLIM y Regresión Logística) para determinar la distribución espacial de la especie *P. versicolor*, con el fin de probar cuál de estos modelos es el más adecuado. La aplicación de los modelos empleados se detalla a continuación.

3.6.1 Maxent (Máxima Entropía)

La aplicación del modelamiento de Máxima Entropía se lo realizó en el software Maxent versión 3.4.1 desarrollado por Biodiversity Informatics. Ésta es una metodología de modelamiento estadístico discriminante que busca la menor pérdida de información posible, usando presencias y ausencias como insumo de entrada. Sin embargo, dichas ausencias las genera automáticamente el

software es su *background*. Adicionalmente no existe la necesidad de normalizar las variables ya que este software realiza la normalización de las variables en su proceso.

3.6.1.1 Aplicación del modelo

Para que los datos de entrada al programa puedan ser reconocidos, se usó la base de datos “specie.csv” (ver Anexo 5.1), donde las presencias deben estar en un formato CSV (delimitado por comas) de Excel con tres columnas, las cuales contengan la especie y coordenadas de estas (especie, X, Y). Es muy importante que tanto las variables dependientes como independientes se encuentren en el mismo sistema de referencia sin importar cuál sea este, en este caso se encuentran en coordenadas UTM 15 S, como se muestra en la Tabla 9.

Tabla 9

Formato de los datos de presencia para Maxent

species	longitude	latitude
p_versicolor	799277	9918113
p_versicolor	799235	9918241
p_versicolor	799223	9918263
p_versicolor	799220	9918281
p_versicolor	799220	9918282
p_versicolor	799247	9918378
p_versicolor	799248	9918378
p_versicolor	799182	9918366
p_versicolor	799172	9918413
p_versicolor	799282	9918465
p_versicolor	799302	9918476

A continuación, se ingresan las variables predictoras las cuales deben estar en formato ASCII (ver Anexo 5.1). Estas fueron obtenidas mediante la ayuda de la herramienta *Ráster to ASCII* de ArcMap. La plataforma de Maxent es muy sencilla y amigable con el usuario donde: en la sección “samples” se ingresa la variable objetivo, es decir, la variable dependiente y en la sección

“Environmental Layer” se ingresa las variables predictoras en formato ASCII (Phillips, 2019). En la Figura 20 se muestra la interfaz del programa y las variables ingresadas. Como se puede ver todas las variables predictoras son variables continuas a excepción de la variable uso de suelo, la cual es elegida como variable categórica, esto debido a que esta variable representa una característica cualitativa del terreno. Entre las opciones disponibles en la sección inferior derecha se activa las tres opciones disponibles las cuales corresponden a: crear una curva de respuesta, crear un gráfico de la predicción y realizar el análisis Jackknife, el cual analiza la importancia de cada variable en la predicción y será descrito más adelante. El formato de salida acumulativo puede ser interpretado como una predicción de las condiciones adecuadas para la especie, en este caso de 0 a 100. En el tipo de formato de salida es ASCII ya que es un formato amigable para la lectura en ArcMap, por último, se elige la carpeta donde los resultados serán guardados.

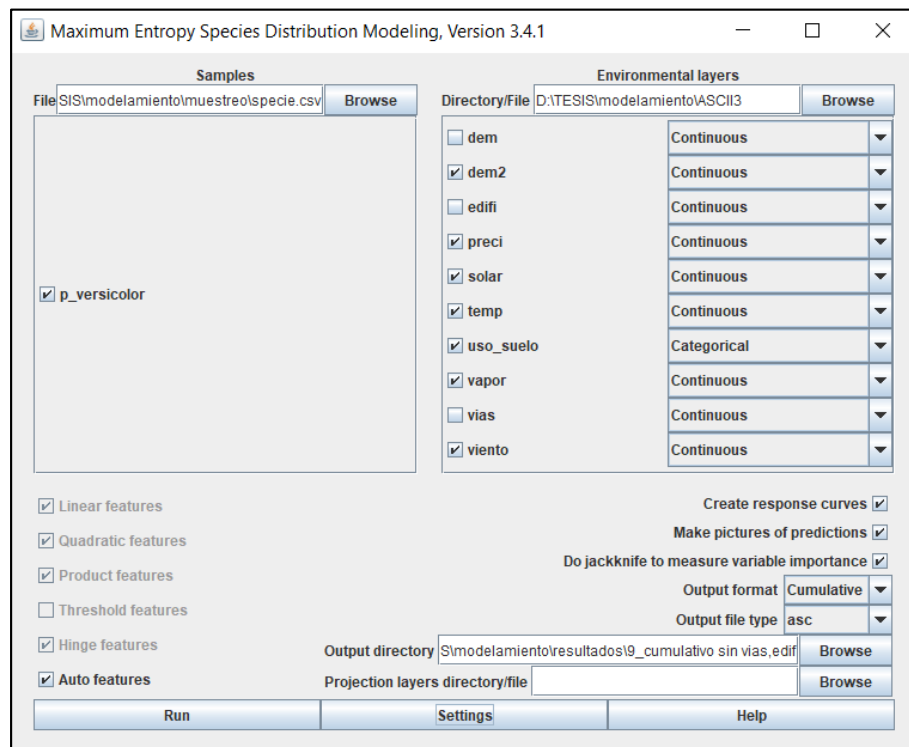


Figura 20. Interfaz principal Maxent.

En la pestaña “settings” el porcentaje de prueba al azar escogido es de 30 esto quiere decir que se usarán 30% de los datos de manera aleatoria para comprobar la validez del modelamiento y generar una curva ROC. Las demás opciones se usan por defecto ya que se usan para estudios ecológicos más específicos y en este caso no son necesarios.

Finalmente, el modelamiento de la distribución comienza con la pestaña RUN, uno de los errores más comunes es que puntos de la variable dependiente no se encuentran dentro del área especificada por las variables independientes, por lo cual el programa dará la opción de omitir estos puntos, caso contrario se deberá corregir este problema en ArcMap procurando que todos los puntos de muestreo queden dentro del área de las variables predictoras. Al terminar el modelamiento entre los archivos resultantes se encuentra un HTML el cual es un resumen de los estadísticos resultantes, así como la imagen de la predicción resultante en formato ASCII (ver Anexo 6).

3.6.1.2 Contribución de las variables usadas.

Entre los siguientes estadísticos devueltos por el software se encuentra la importancia de las variables en el proceso de entrenamiento, eso se determina en función de la contribución que tuvo cada variable en el modelo. Cada paso del algoritmo de Maxent incrementa la ganancia del modelo mediante la modificación del coeficiente para una única característica; el programa asigna el incremento en la ganancia a las variables ambientales de las cuales depende dicha característica y al final se realiza una conversión a porcentajes (Tabla 10).

Tabla 10*Contribución de cada variable al modelo Maxent*

Variable	Percent contribution	Permutation importance
edifi	68.3	72.5
uso_suelo	16.7	0
viento	6.4	4.7
preci	2.9	5.5
temp	2.4	0.6
solar	1.6	3.9
dem2	1.4	12.5
vapor	0.2	0.2

Con el fin de estimar cuales son las variables más importantes en el modelo también se hace uso de otro análisis en este caso el resultado de la prueba Jackknife, con esta prueba se crea un conjunto de modelos, en cada oportunidad se excluye una variable y se crea un modelo con las demás, por tanto, también se crea un modelo con cada variable aislada (Phillips, 2019).

En la figura 21, en barras de color celeste se observa la ganancia que obtienen el modelo sin la variable, en color azul la ganancia que obtiene el modelo solo con esa variable y las barras rojas representan la ganancia dada por todas las variables (Phillips, 2019).

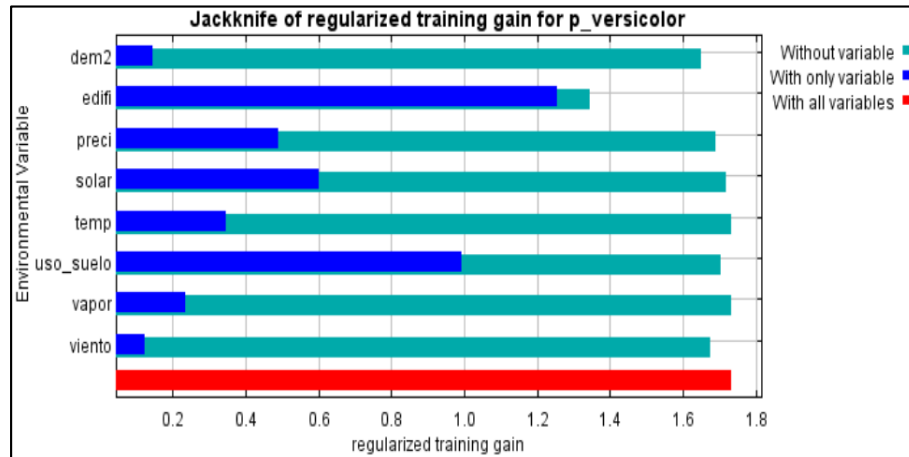


Figura 21. Diagrama del análisis Jackknife para las variables del entrenamiento. Entrenamiento sin la variable (celeste), entrenamiento solo con la variable (azul).

Otra forma de analizar la correlación de las variables son las curvas de respuestas (Figura 22), las cuales representan la influencia de cada variable sobre el modelo en conjunto con la correlación que existe entre ellas.

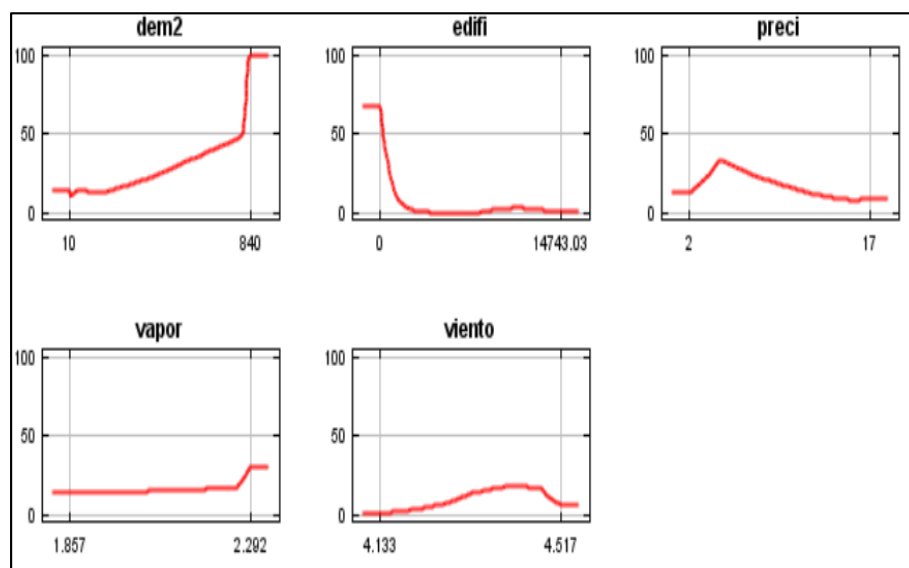


Figura 22. Curvas de respuesta del modelamiento (Eje y: ganancia de la variable en el algoritmo, eje x: dimensión de la variable).

El la Figura 23 representa la curva de respuesta de la variable distancia a infraestructura antrópica, el eje “y” muestra la ganancia acumulada producto del modelamiento, mientras que el eje “x” muestra la dimensión de la variable, en este caso “metros”.

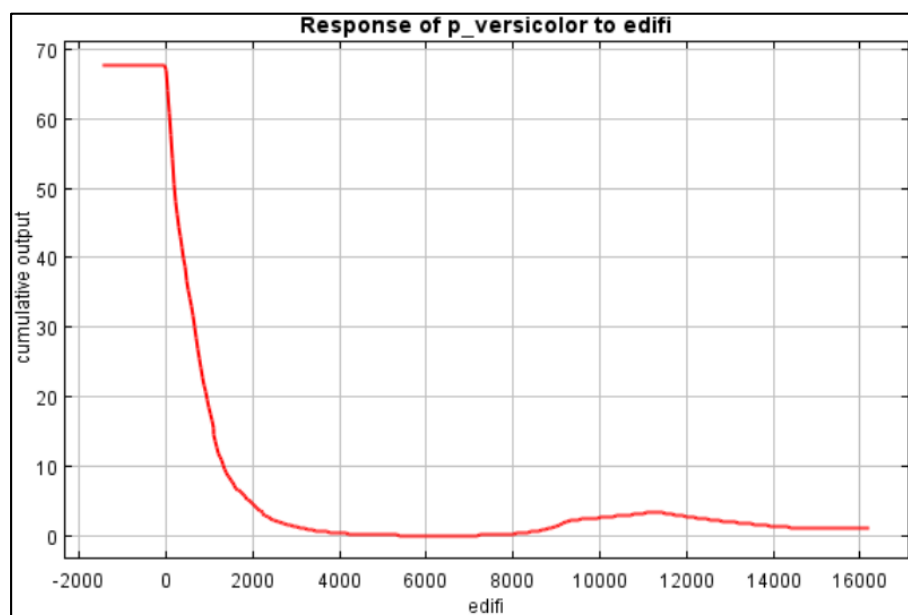


Figura 23. Curva de respuesta de la variable distancias a infraestructura antrópica.

3.6.2 Splines de Regresión Multivariada Adaptativa (MARS)

Creado por la Universidad de Salford Inglaterra en la plataforma Salford Predictive Modeler (SPM), MARS es un módulo enfocado en la elaboración de modelos predictivos y descriptivos que ayudan en el análisis de bases de datos de cualquier tamaño y con diferente objetivo, entre otros programas creados con el mismo fin están herramientas estadísticas como CART, TreeNet y Random Forest.

Como se menciona en el capítulo anterior (ver capítulo 2.4.5), MARS usa un método no paramétrico, que no hace ninguna suposición sobre la relación fundamental existente entre las variables respuesta e independiente, así como otros supuestos de la regresión lineal múltiple. De este modo MARS construye relaciones basándose solamente en una serie de coeficientes asociados

a cada función base que son encontradas a partir de las regresiones de los datos de entrada. A este modo de trabajar también se lo puede entender como un algoritmo que trabaja con múltiples pedazos de regresiones lineales.

3.6.2.1 Aplicación del modelo

En el módulo MARS de Salford Predictive Modeler 8, se usó como insumo de entrada al archivo “total.csv” (ver Anexo 5.2), que es un formato delimitado por comas (.CSV) con los datos de cada punto de presencia y pseudo-ausencia, junto a esta una matriz que contenga los valores de las variables independientes, estos datos se obtuvieron gracias a la herramienta *Extract Multi Values to Point* de ArcMap. En la figura 24 se muestra la base de datos de entrada para MARS.

presencia	temp	vapor	viento	dem2	solar_1	precip_1	dist_edif2	uso_suelo2
0	22.3167	2.2816701	4.46667	21.4244995	16300		5	7354.77979
0	21.7833004	2.2183299	4.3666701	145.585999	15508		6	7120
0	22.0499992	2.1866701	4.1833301	136.067001	15463		7	5552.37012
0	22.5	2.22333	4.2666702	84.4133987	15224		6	6676.04981
0	22.0667	2.2583301	4.46667	37.6493988	15207		8	7265.91992
0	22.0832996	2.26	4.46667	43.9012985	16072		6	7463.62012
0	22.1000004	2.2149999	4.3499999	139.382996	16439		3	7474.75977
0	22.4167004	2.2883301	4.4833298	10.0956001	16494		4	10608.7002
0	22.3999996	2.2783301	4.5	19.4631004	16496		4	10451.9004
0	22.0667	2.2583301	4.46667	39.8160019	15538		7	7581.02979
0	22.8500004	2.2516699	4.3666701	59.9878998	15845		4	11622.2998
0	22.4333	2.25	4.46667	22.6623001	15760		7	6322.02002
0	23.1833	2.27	4.4499998	10	16291		2	8809.08984
0	21.1499996	2.175	4.28333	200	14723		9	4252.06006
0	23.1833	2.27	4.4499998	10	16291		2	8033.93018
0	22.7166996	2.25	4.3666701	19.1163998	16060		5	8033.93018
0	21.7667007	2.2283299	4.3499999	72.1680984	13898		10	3617.72998
0	22.0499992	2.2550001	4.46667	40	15435		8	6547.29981
0	22.0832996	2.2483301	4.46667	80	16113		5	9200.34961
0	22.7667007	2.25333	4.3666701	69.0569	16185		3	10898.7998
0	21.9666996	2.24667	4.4499998	73.5870972	15859		6	7549.72998
0	22.7332993	2.27	4.46667	35.5581017	16352		3	12603.7002
0	22.7833004	2.2316699	4.2666702	61.7132988	15484		6	9794.37012
0	22.1833	2.1800001	4.2666702	163.707993	16093		5	8222.88965
0	22.0667	2.2583301	4.4166698	44.7549019	15033		8	6432.91992
0	22.0667	2.19333	4.333302	160	16194		4	8714.49023
0	22.2166996	2.27	4.5	10	16174		5	8272.55957

Figura 24. Base de datos de entrada para MARS (columna de puntos de presencia junto a la matriz de valores independientes en cada punto).

Para la definición de variables la interfaz de MARS permite escoger cual será la variable dependiente y variables independientes, en este caso la variable presencia es la variable objetivo y debe ser elegida de tipo logístico binario por ser presencias y ausencias, las demás variables se les

asigna como predictores, procurando escoger a uso de suelo como una variable categórica (Figura 25).

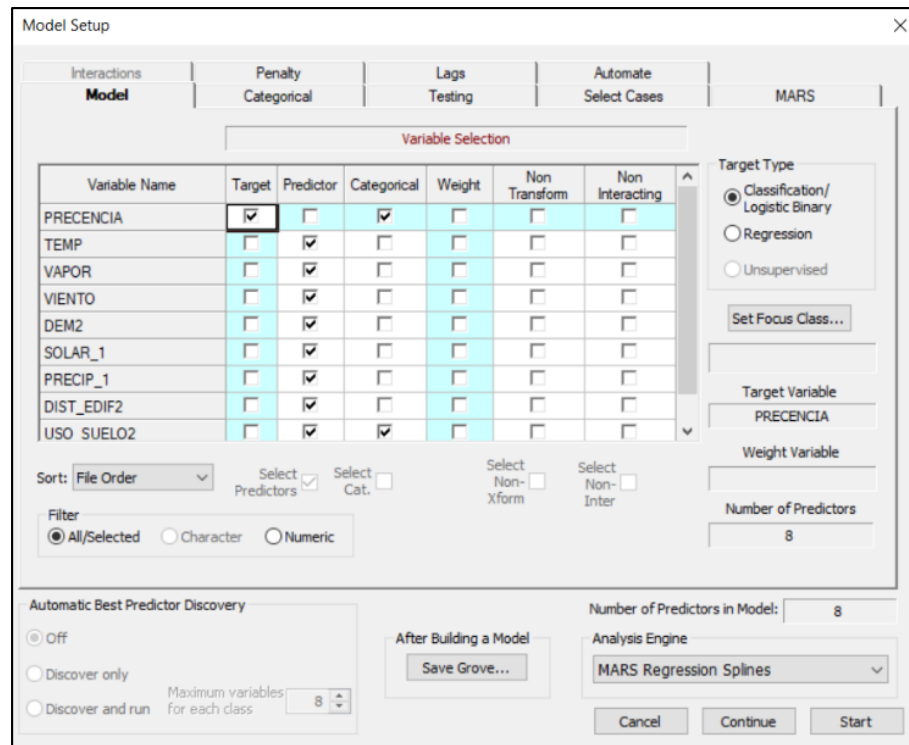


Figura 25. Interfaz principal del software MARS. Selección de variables dependientes (Target) e independientes (Predictor).

Una vez que se ingresó y designó las variables necesarias, se define el porcentaje de prueba que se usará para realizar el test, en este caso se aplicó el 30% (Figura 26).

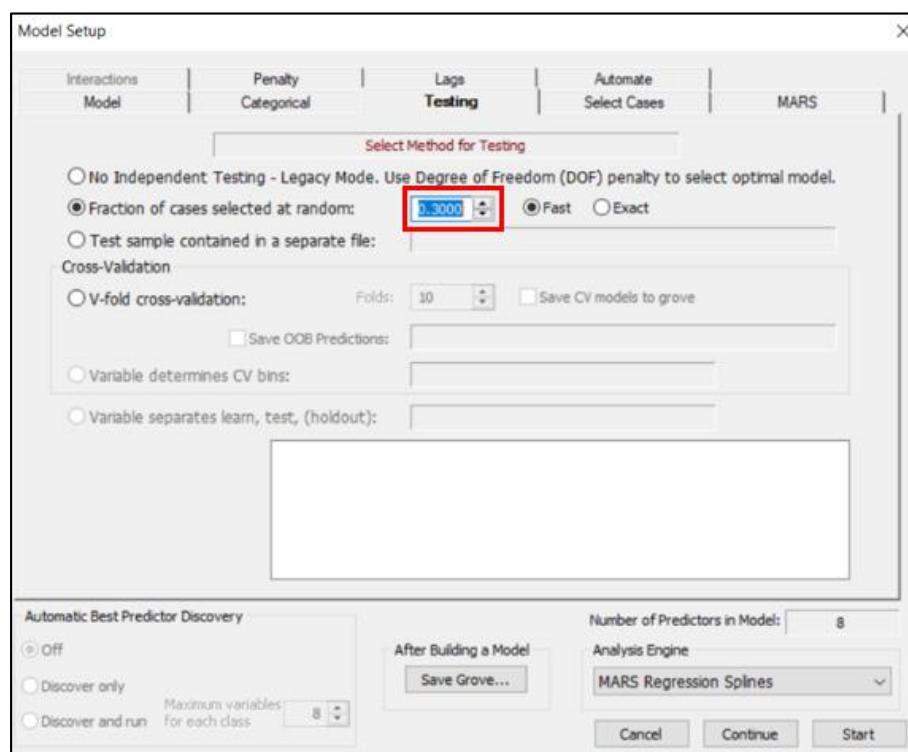


Figura 26. Configuración del test de MARS.

Para la definición de la cantidad de funciones base se hizo referencia al manual de MARS (Minitab, 2018), donde se recomienda que el número de funciones bases sea 3 o 4 veces el número de predictores, por tal motivo se usó un máximo de 32 funciones base, así también recomienda usar 10 iteraciones en caso de usar el modelo con fines de minería, al no ser este el caso, se usa el valor por defecto 1 (Figura 27).

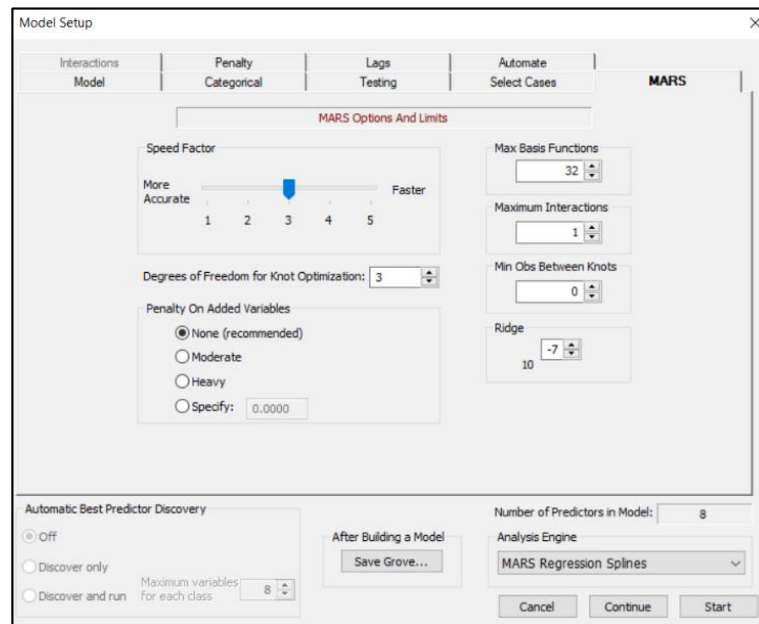


Figura 27. Configuración del modelamiento MARS.

El modelo MARS usa las llamadas funciones base que junto con parámetros del modelo (estimados a partir de mínimos cuadrados), crean predicciones de la información de entrada. Una vez procesado el modelo se observa una ventana donde se muestra estadísticos importantes como R^2 y curvas de aprendizaje (Figura 28). En la pestaña “Summary” se encuentra la contribución de variables, matriz de confusión y funciones básicas resultantes.

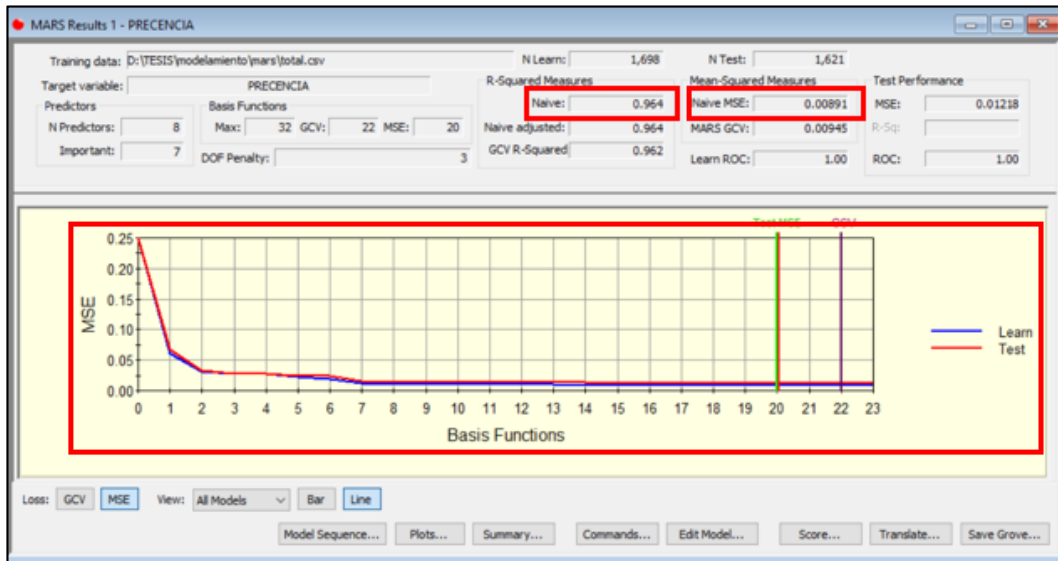


Figura 28. Ventana resultante de modelamiento con las curvas y estadísticos de aprendizaje.

3.6.2.2 Funciones básicas

Como se menciona en el capítulo anterior (ver apartado 2.4.5) las funciones básicas son el principal producto de este modelador, estas funciones son usadas en una ecuación predictora que junto a coeficientes determinaran el valor de cada píxel sobre el mapa final. En la Tabla 11 se muestran las funciones básicas resultantes y a continuación la ecuación predictora.

Tabla 11*Funciones básicas resultantes del modelo MARS*

BF3	max(0, SOLAR_1 - 16458);
BF4	max(0, 16458 - SOLAR_1);
BF5	max(0, VAPOR - 2.09333);
BF6	max(0, 2.09333 - VAPOR);
BF7	max(0, TEMP - 20.6167);
BF8	max(0, 20.6167 - TEMP);
BF9	max(0, DEM2 - 204.743);
BF10	max(0, 204.743 - DEM2);
BF11	max(0, PRECIP_1 - 13);
BF112	max(0, 13 - PRECIP_1);
BF15	max(0, DIST_EDIF2 - 1712.07);
BF17	max(0, VAPOR - 2.125);
BF19	max(0, VAPOR - 2.065);
BF21	max(0, DIST_EDIF2 - 4722.71);
BF23	max(0, SOLAR_1 - 16253);
BF25	max(0, TEMP - 20.3167);
BF27	max(0, VIENTO - 4.26667);
BF28	max(0, 4.26667 - VIENTO);
BF29	max(0, DEM2 - 320);
BF31	max(0, SOLAR_1 - 13777);

Ecuación predictora:

$$\begin{aligned}
 Y = & 0.723257 + 0.00199082 * BF3 + 0.000142066 * BF4 - 31.3491 * BF5 \\
 & + 14.2421 * BF6 - 0.990317 * BF7 - 1.34252 * BF8 \\
 & + 0.0011504 * BF9 + 0.00103011 * BF10 - 0.0409515 * BF11 \\
 & + 0.0449618 * BF12 - 0.000336015 * BF15 + 22.309 * BF17 \\
 & + 8.38919 * BF19 + 0.000321367 * BF21 - 0.000489379 * BF23 \\
 & + 0.915586 * BF25 - 0.357162 * BF27 - 1.16999 * BF28 \\
 & - 0.00196543 * BF29 + 0.000122577 * BF31;
 \end{aligned}
 \tag{12}$$

Donde Y es la probabilidad del modelo.

Hay 20 funciones básicas, las cuales serán transformadas en ráster y posteriormente operadas en la ecuación predictora, esto usando la herramienta *Raster Calculator*. En la Figura 29 se muestra la herramienta *Raster Calculator* para la transformación de las funciones básicas a ráster con la ayuda del condicional “Con” que reemplaza a la función “Max”. El condicional “Con” funciona como un condicional SI estructurado de la siguiente manera: Con (condicional, si cumple condición, caso contrario). Para finalizar se aplica la ecuación final con los rústers creados (ver Anexo 7).

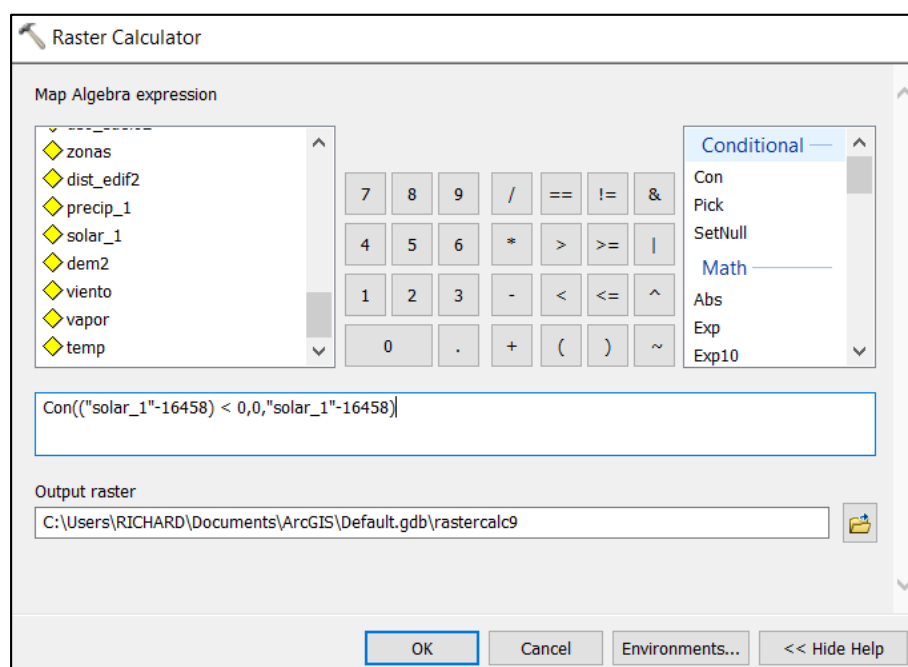


Figura 29. Herramienta *Raster Calculator*: transformación de las funciones básicas a ráster.

3.6.3 BIOCLIM

La aplicación del modelo BIOCLIM se lo realiza en el software DIVA-GIS versión 7.5, el cual es un programa informático gratuito para el mapeo y el análisis de datos espaciales lo cual lo convierte es un sistema de información geográfica (SIG). DIVA-GIS es un programa con un alto soporte online, ya que en su página web oficial se encuentran datos espaciales locales, regionales y mundiales gratuitos de todo nivel como: límites administrativos, densidad poblacional, alturas, cobertura de suelo, entre otros. Así mismo los datos climáticos provenientes de Worldclim adaptados al formato necesario para DIVA-GIS el formato GRID (.grd), también hace una recopilación de datos de ocurrencia de especies, datos de cultivos e imágenes satelitales de alta resolución provenientes de Landsat.

DIVA-GIS es particularmente usado para mapear y analizar datos de biodiversidad, así como la distribución de especies y otros puntos de distribución. Es compatible con varios formatos de ESRI ArcMap, así que se puede trabajar con datos tratados en otros programas SIG. Para realizar mapas de distribución de diversidad biológica y modelamiento de nicho ecológico se puede usar los algoritmos de BIOCLIM y DOMAIN que ya vienen integrados en el software, así como modelos externos los cuales deberán ser programados previamente en R.

3.6.3.1 Aplicación del modelo

Es muy importante establecer que DIVA-GIS lee las variables independientes en formato GRID (.grd), por lo cual la base de datos de 8 variables establecidas anteriormente en formato ASCII de ESRI (ver Anexo 5.3), debe ser importado a formato GRID, se usa la herramienta *import to Gridfile* de DIVA-GIS, como se muestra en la Figura 30 usando como entrada la base de datos en formato ASCII.

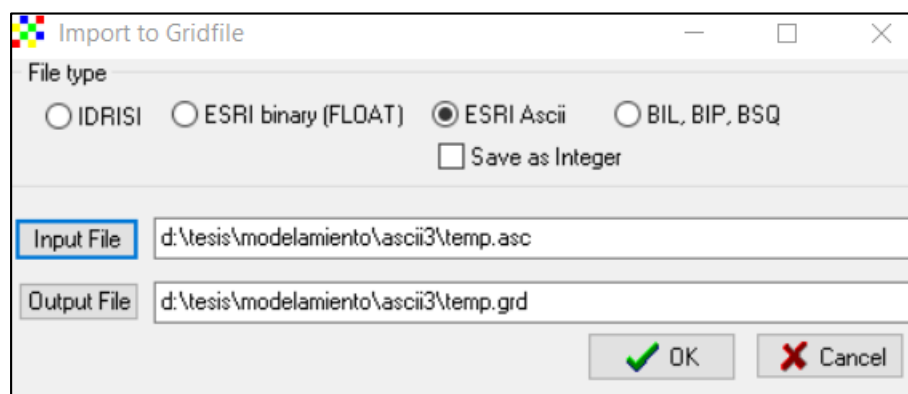


Figura 30. Importación de variables formato ASCII a formato GRID.

Una vez creadas todas las variables en formato GRID se debe realizar un Stack de estas variables, lo cual es apilar todas las variables predictoras en un solo archivo de formato Stack (.gri), esto se realiza mediante la herramienta de *Make Stack*, como se muestra a continuación en la Figura 31. Con esto se habrá creado las entradas climatológicas o variables predictoras necesarias para la interfaz del modelador.

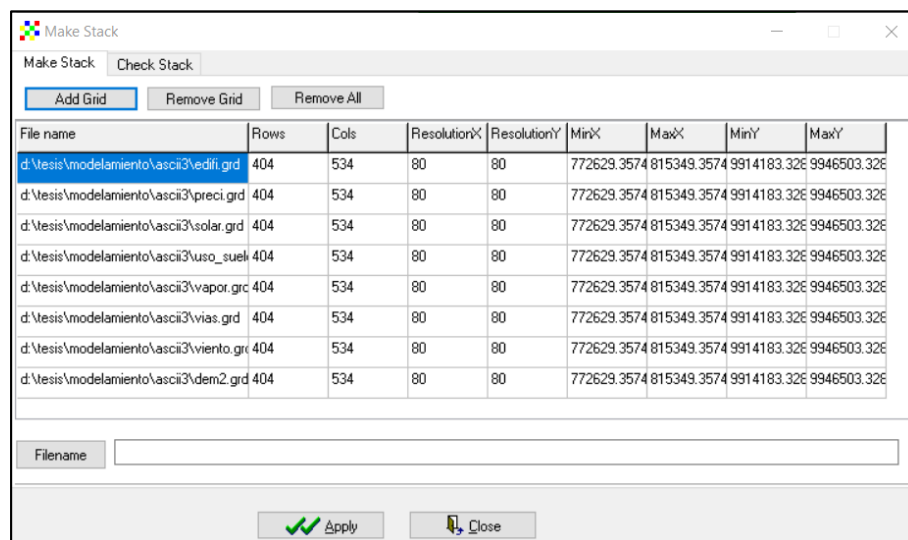


Figura 31. Creación de Stack de variables predictoras.

Para ingresar los puntos de muestreo a DIVA-GIS se usa el text file “presencia.txt” donde se incluya el nombre de la especie, sus coordenadas, y su presencia (ver Anexo 5.3). Esto se lo puede realizar simplemente con la exportación de la tabla de datos del shape de puntos en ArcMap. Ahora se usa la herramienta *Create Shapefile from Text File* donde se determina el tipo de separación y las columnas correspondientes a ESTE y NORTE (Figura 32).

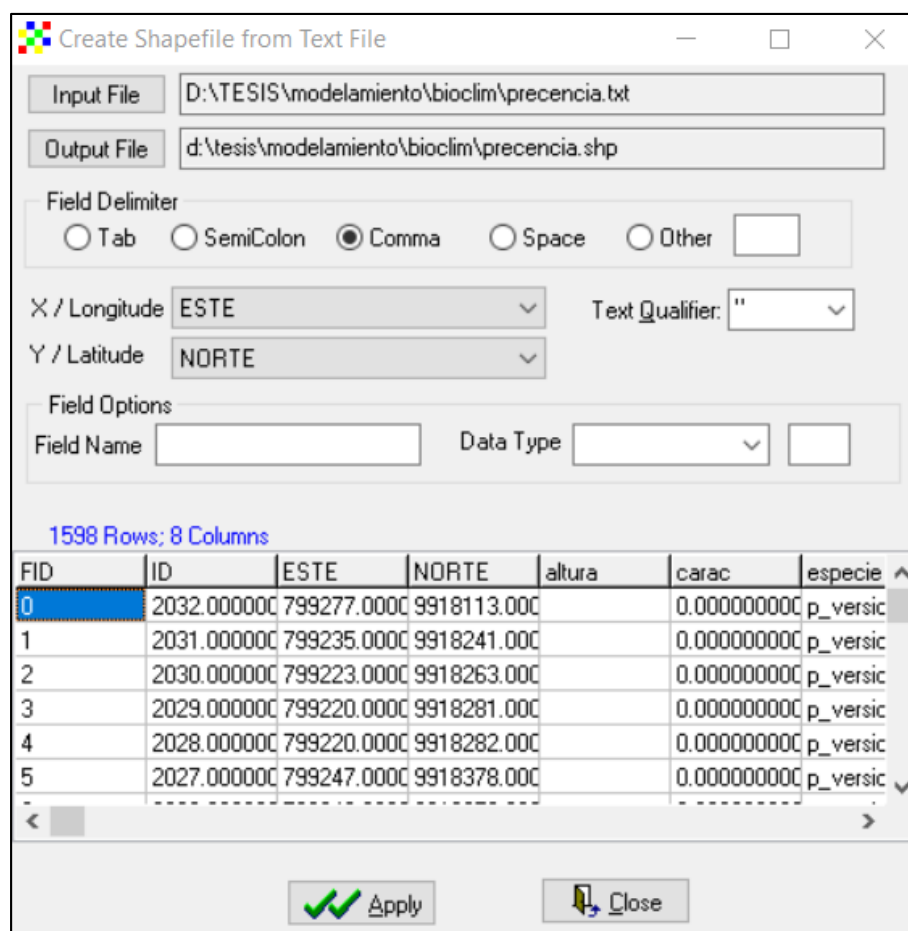


Figura 32. Ingreso de puntos de muestreo a DIVA-GIS.

En la preparación de puntos de muestreo para el entrenamiento y la prueba, se ingresa el shapefile de puntos de presencia con la herramienta *add layer*. Y luego en la herramienta *prepare*

points se usa el 70% de los puntos de entrada para el entrenamiento y se activa datos de ausencia aleatoria (Figura 33).

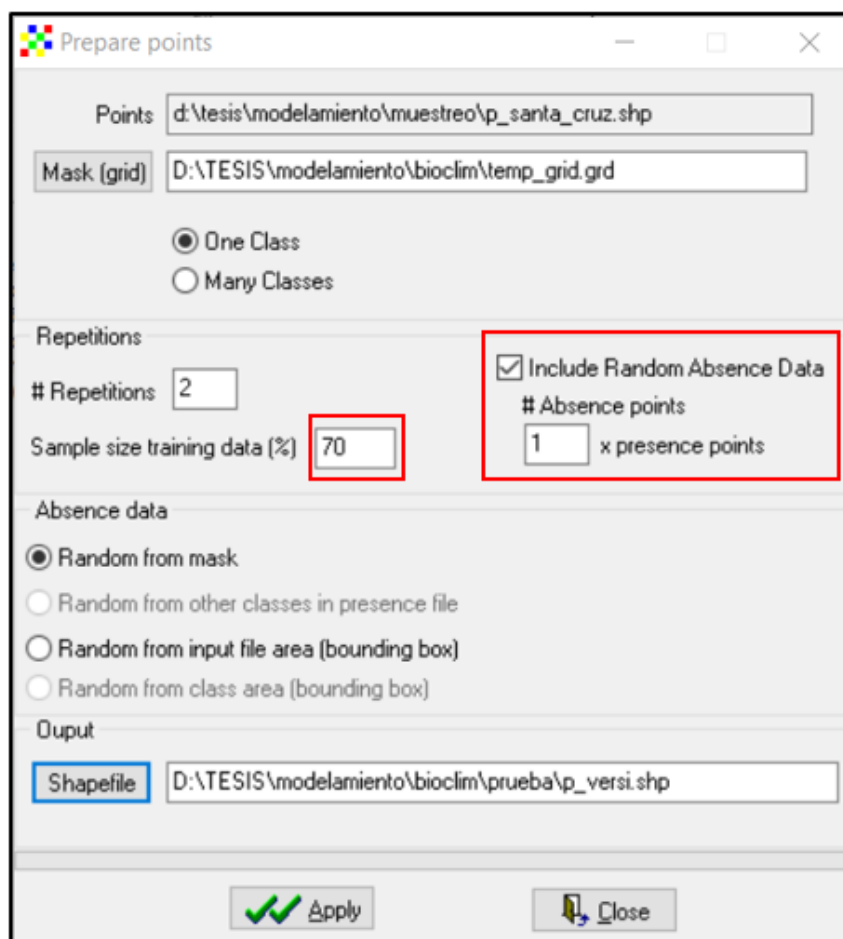


Figura 33. Preparación de puntos para el entrenamiento y la prueba.

Seguidos estos pasos el resultado serán dos archivos shapefile, uno destinado al entrenamiento del modelador y otro para su posterior prueba. Para la modelación se recurre a la herramienta *Distribution Modeling*, donde por insumos de entrada se ingresa el Stack de las variables predictoras realizado anteriormente y el shapefile de puntos de entrenamiento.

En la pestaña “predict”, se ingresa la extensión sobre la cual se desea realizar la predicción, así como las variables que influye dentro de la misma, en el tipo de salida se puede elegir entre los algoritmos de BIOCLIM o DOMAIN. Las demás pestañas sirven para realizar un análisis previo de los puntos de presencia sobre las variables predictoras (Figura 34).

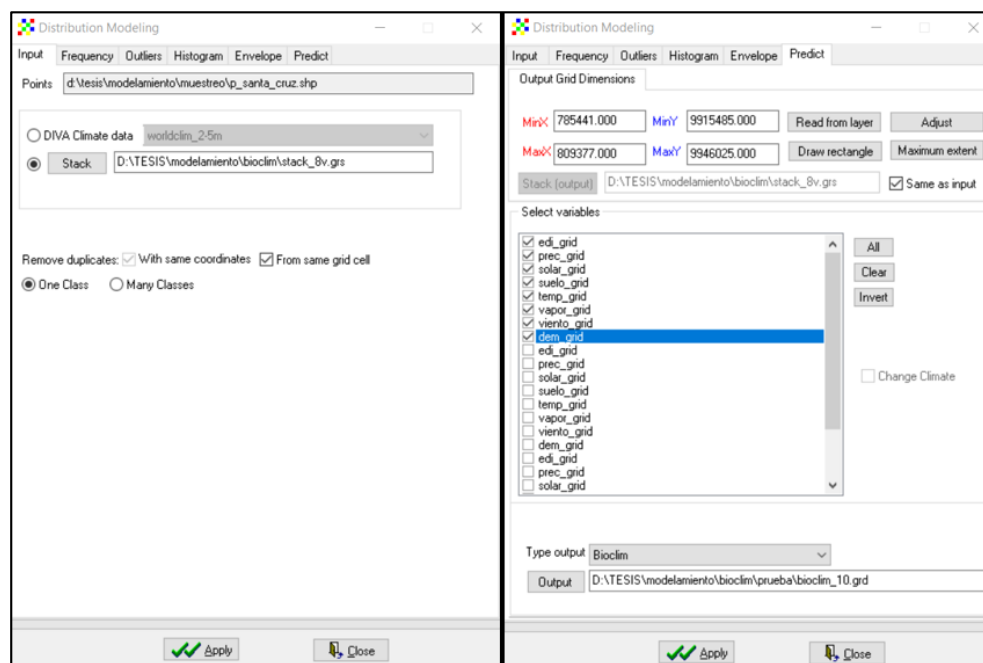


Figura 34. Configuración de la interfaz de BIOCLIM.

Una vez que todos los parámetros estén correctos se inicia la modelación y como resultado se obtiene un GRID con la modelación de nuestra especie (ver Anexo 8).

3.6.3.2 Curva ROC

Como se menciona anteriormente el archivo de prueba resultante de la preparación de puntos de presencia, sirve como insumo para generar un archivo de evaluación mediante la herramienta *extract values points from grid*. En esta herramienta los insumos de entrada son los puntos de prueba, la máscara la cual será el archivo GRID resultante del modelamiento y la dirección para el archivo de salida (Figura 35).

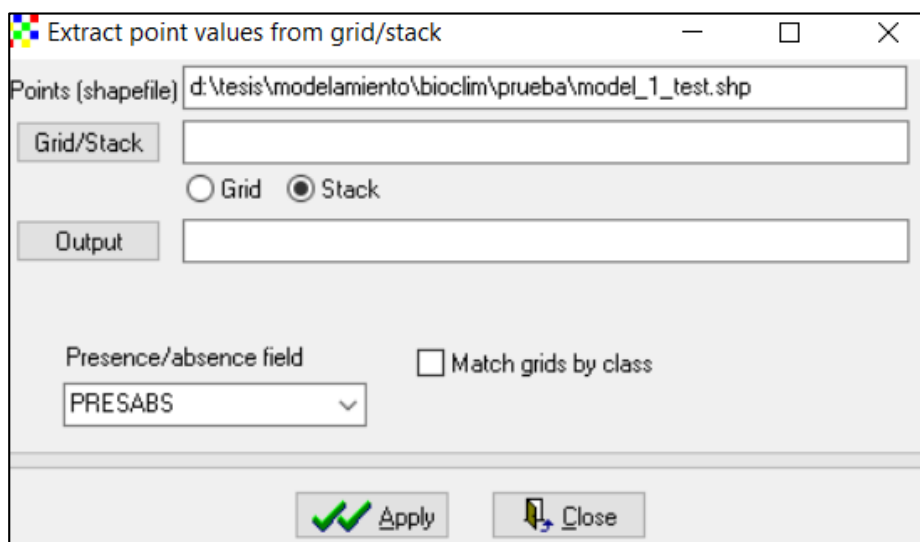


Figura 35. Extracción de archivo de evaluación de los puntos de prueba.

El archivo resultante en formato ROC (.roc) será el insumo para la herramienta *Show ROC/Kappa*, que muestra la curva ROC y estadístico Kappa (Figura 36).

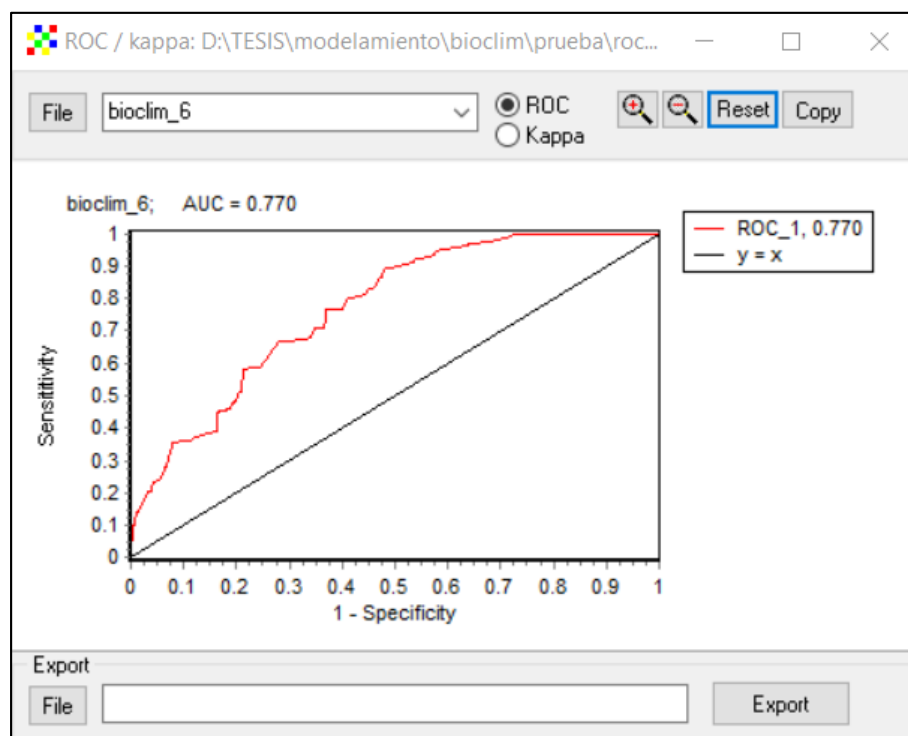


Figura 36. Curva ROC resultante de modelamiento.

3.6.4 Regresión logística

La aplicación de la Regresión Logística para el modelo de la distribución de *P. versicolor* se lo realizó en SPSS 25 por sus siglas en inglés “Statistical Package for the Social Sciences”, el cual es un formato creado por IBM para el análisis estadístico de datos. Este software incluye estadística descriptiva como la tabulación y frecuencias de cruce, estadística de dos variables, así como pruebas T-STUDENT, ANOVA y de correlación. En este caso es muy útil ya que permite crear la correlación existente entre la variable objetivo (presencia) con las variables predictoras.

3.6.4.1 Aplicación del modelo

El uso de la herramienta regresión logística binaria disponible en SPSS usa como datos de entrada el archivo “total.csv” que contiene los datos de presencia y ausencia junto a los valores para cada variable en una matriz (ver Anexo 5.4). En la Figura 37 se muestra el ingreso de la base de datos.

	presencia	temp	vapor	viento	dem2	solar_1	precip_1	dist_edif2	uso
1714	0	23.0667000	2.2449999	4.4166698	40.0000000	16094	4	11161.4003900	
1715	0	22.7667007	2.2466700	4.3666701	52.6254997	15970	4	12080.2998000	
1716	0	22.4167004	2.2883301	4.4833298	10.0000000	16450	4	10893.2002000	
1717	0	22.8332996	2.2700000	4.4666700	41.0105019	16344	2	11409.5000000	
1718	0	23.0167007	2.2550001	4.4166698	13.2797003	16348	2	14330.9003900	
1719	0	22.1499996	2.2033300	4.2666702	100.8020020	15156	7	6191.0898440	
1720	0	22.0667000	2.2583301	4.4666700	40.0000000	15538	7	7491.8598630	
1721	0	22.8500004	2.2516699	4.3666701	41.7326012	15845	4	12118.0996100	
1722	1	22.2500000	2.2616701	4.4666700	10.0000000	14840	8	80.0000000	
1723	1	22.2500000	2.2616701	4.4666700	13.5359001	14840	8	160.0000000	
1724	1	22.2500000	2.2616701	4.4666700	13.5359001	14840	8	160.0000000	
1725	1	22.2500000	2.2616701	4.4666700	13.5359001	14840	8	178.8849945	
1726	1	22.2500000	2.2616701	4.4666700	13.5359001	14840	8	178.8849945	
1727	1	22.2500000	2.2616701	4.4666700	18.0114002	14840	8	226.2740021	
1728	1	22.2500000	2.2616701	4.4666700	18.0114002	14840	8	226.2740021	

Figura 37. Ingreso de base de datos a SPSS.

En la Figura 38 se muestra la configuración, donde se identifica la variable dependiente y las variables independientes.

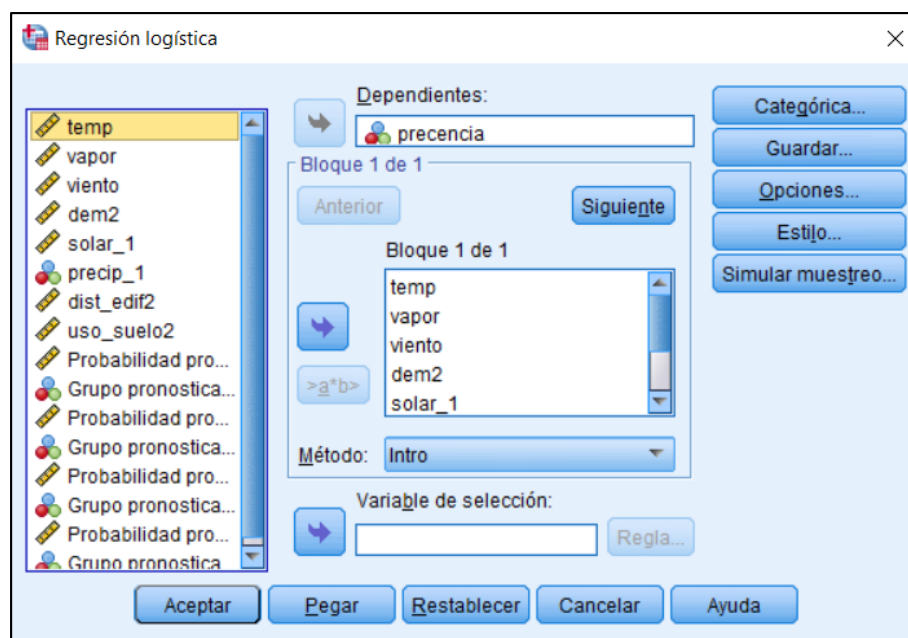


Figura 38. Configuración de la Regresión Logística. Las variables independientes se seleccionan en la sección de “bloque 1 de 1”.

En el resultado de la regresión se obtienen varios estadísticos entre los cuales se muestra R^2 y tabla de confusión. El principal producto resultante de la Regresión Logística para este estudio es la tabla de coeficientes, los cuales como se mencionan en el anterior capítulo ayudan a formular la ecuación predictora y se podrá generar un mapa de predicción, con la ayuda de la herramienta *Raster Calculator* de ArcMap. A continuación, en la Tabla 12 se muestran los coeficientes, estadísticos resultantes y la ecuación que dará paso al mapa de predicción. La columna denotada por B muestra los coeficientes asignados para cada variable junto con su respectivo error estándar, en la columna Wald se puede ver la influencia que ha tenido cada variable en el proceso de la regresión.

Tabla 12

Coefficientes y estadísticos resultantes de las variables en el modelo Regresión Logística

	B	Error estándar	Wald	gl	Sig.	Exp(B)
temp	2.461	.850	8.388	1	.004	11.721
vapor	-111.454	17.356	41.236	1	.000	.000
viento	19.791	2.484	63.501	1	.000	393722325.770
dem2	-.012	.011	1.192	1	.275	.988
solar_1	.001	.001	2.359	1	.125	1.001
precip_1	.621	.269	5.341	1	.021	1.860
dist_edif2	-.001	.000	167.083	1	.000	.999
uso_suelo2	38.645	17.714	4.759	1	.029	60726327696556312.000
Constante	91.585	50.101	3.342	1	.068	5.952E+39

$$y = \frac{1}{1 + e^{-(\beta_0 + \beta_1 * temp + \beta_2 * vapor + \dots + \beta_n * uso_{suelo})}} \quad (13)$$

Donde y es la probabilidad de encontrar la especie y β_n es el coeficiente calculado.

Se aplica la ecuación junto con los coeficientes de la Tabla 12 y con la ayuda de *Raster Calculator* de ArcMap se crea el ráster resultante (ver Anexo 9).

3.6.4.2 Curva ROC

Para la generación de la curva ROC se usó la herramienta *Curva COR*, donde se ingresó la variable de estado (presencias) y la variable de prueba es decir el resultado de la regresión logística binaria realizada anteriormente (Figura 39).

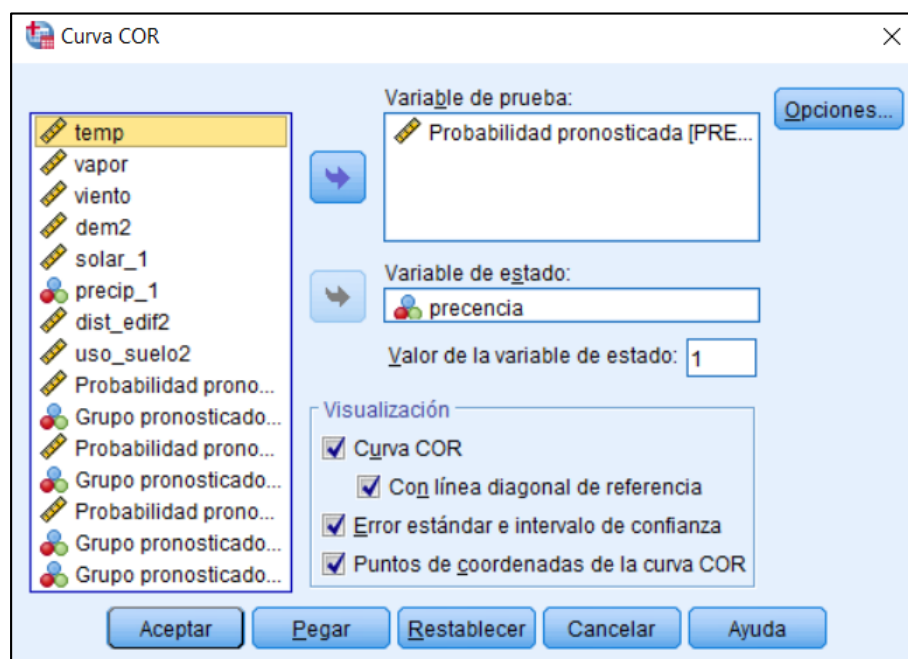


Figura 39. Generación de la curva ROC.

En la Figura 40 se muestra la curva ROC (azul) y la curva de lo esperado por el azar (rojo).

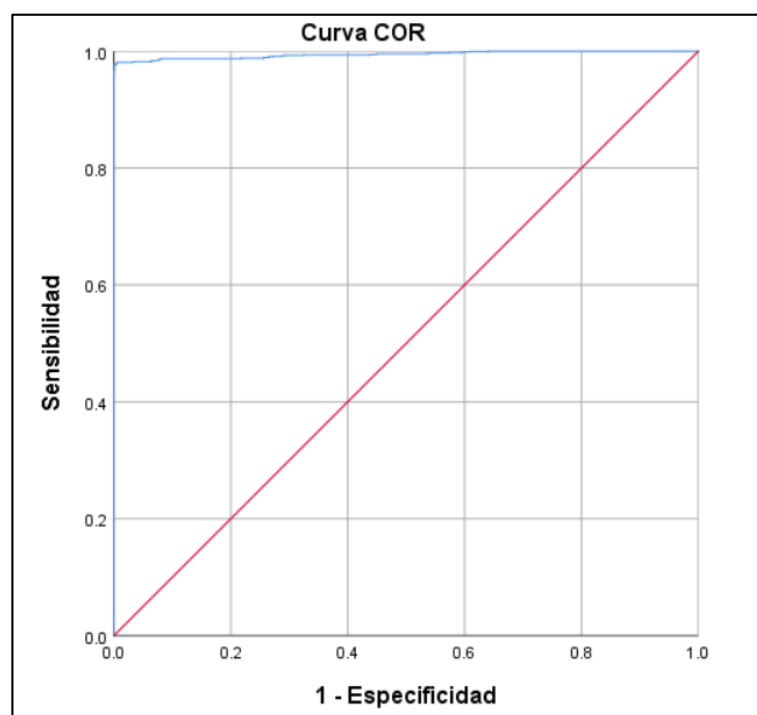


Figura 40. Curva ROC resultante de la Regresión Logística.

CAPITULO IV

RESULTADOS Y DISCUSIONES

4.0 Resumen

En este capítulo se presenta los resultados y el análisis de los estadísticos de los modelos realizados: Maxent, MARS, BIOCLIM y Regresión Logística. Los estadísticos muestran que los modelos con mejores ajustes son MARS y Regresión Logística ya que ambos presentan un AUC de 0.99. Se realiza el análisis de la curva ROC con estadísticos previos como: matriz de confusión, estadístico Kappa y análisis de omisión y comisión. Para los modelos MARS y Regresión Logística el resultado de la matriz de confusión es de 98.91% y 98.8% respectivamente lo cual valida el modelamiento y presenta un ajuste muy satisfactorio.

El modelo Maxent presenta un AUC de 0.94 y un análisis de omisión y comisión muy satisfactorio. Por último, está el modelo BIOCLIM el cual presenta un AUC de 0.77 y un estadístico Kappa de 0.41, demostrando que el modelo es poco satisfactorio. Adicional a esto, se realiza un análisis gráfico donde de manera visual se determina la satisfacción de cada modelo con probabilidades alta, media y baja frente a puntos de prueba entre presencia y pseudo-ausencias elegidos al azar. Este análisis contrasta con los estadísticos ya que los modelos visualmente más satisfactorios son MARS y Regresión Logística seguido del modelo Maxent y por último con poca aceptación BIOCLIM.

4.1 Análisis estadísticos

Ya que la curva ROC es el estadístico más comúnmente usado para evaluar un modelo de distribución de especies (ver apartado 2.5.2), es de suma importancia asegurarse que la curva ROC

resultante de cada modelo es de alta confiabilidad y sea la mejor posible. El análisis de resultados previos a la curva ROC como: análisis de omisión y comisión, matriz de confusión y estadístico Kappa, permiten validar la curva ROC resultante.

4.1.1 Resultados del modelo Maxent

- Omisión y comisión

La figura 41 muestra la tasa de omisión y el área pronosticada en función del umbral cumulativo (umbral usado por default). Como se puede observar, las curvas de tasa de omisión de entrenamiento y de prueba se acercan satisfactoriamente a la curva de omisión prevista por lo cual el análisis de omisión y comisión es satisfactorio.

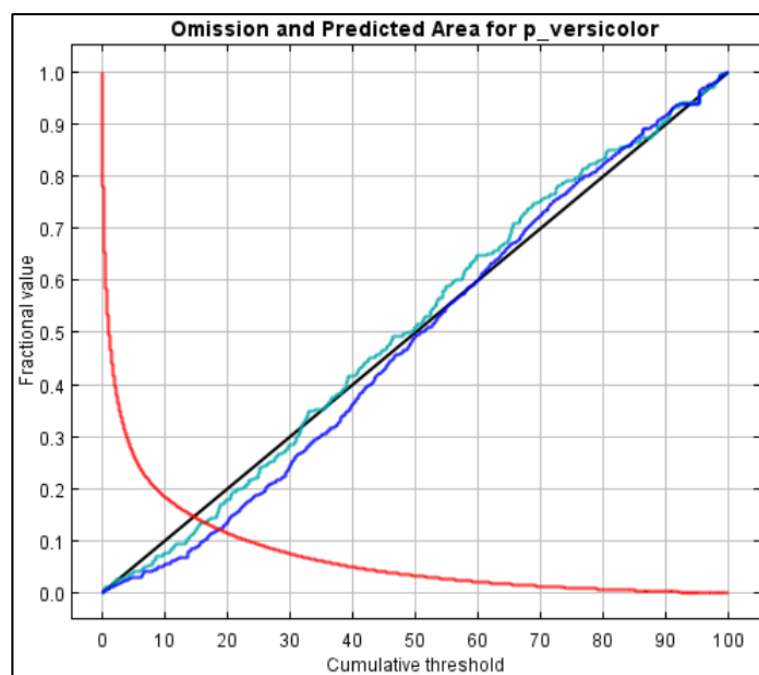


Figura 41. Diagrama de omisión: entrenamiento (azul); prueba (celeste); prevista (negra); fracción de pseudo-ausencias pronosticadas (rojo).

4.1.2 Resultados del modelo MARS

- Coeficiente de determinación (R^2)

En la Figura 42 se muestran los estadísticos resultantes tanto del modelamiento como del test que se usó en el software MARS. Como se puede ver la medida del R^2 es de 0.964, siendo muy satisfactoria y en el diagrama de líneas las curvas de aprendizaje (azul) y la curva de prueba (rojo), ambas se encuentran muy cercanas mostrando la disminución del error cuadrático medio (MSE) representado en el eje de la “y”, a través del entrenamiento de las funciones básicas este llega a 0.0089. Los estadísticos muestran que el ajuste del entrenamiento MARS es muy satisfactorio.

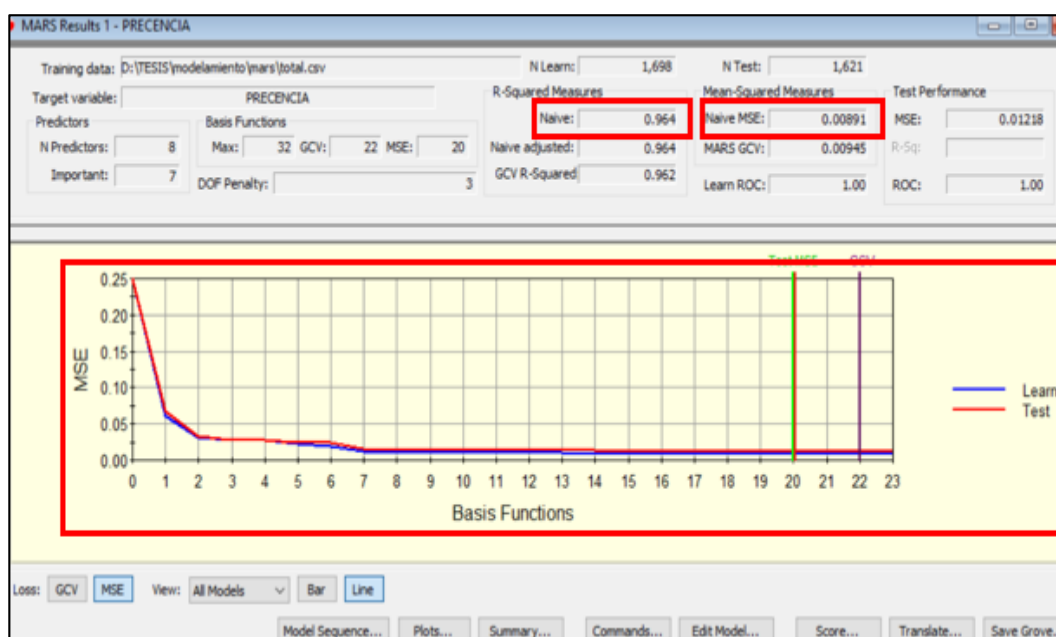


Figura 42. Estadísticos resultantes y curvas de aprendizaje en el modelo MARS.

- Matriz de confusión

En la Figura 43 se presenta la relación entre presencia y ausencias con los errores de omisión y comisión, como se observa existe error por omisión (más crítico) en 17 casos de 762 y el error

por comisión es de 0 dando un porcentaje de aciertos de 98.95% lo cual verifica que el resultado del modelamiento es satisfactorio.

Actual Class	Total Class	Percent Correct	Predicted Classes	
			0 N = 859	1 N = 762
0	842	100.00%	842	0
1	779	97.82%	17	762
Total:	1,621			
Average:		98.91%		
Overall % Correct:		98.95%		
Specificity		100.00%		
Sensitivity/Recall		97.82%		
Precision		100.00%		
F1 statistic		98.90%		

Figura 43. Matriz de confusión resultante del modelo MARS.

4.1.3 Resultados del modelo BIOCLIM

- Estadístico Kappa

En la Figura 44 se muestra el resultado del estadístico Kappa del modelamiento BIOCLIM, donde el resultado muestra un Kappa de 0.41 lo cual es poco satisfactorio, el entrenamiento está más aproximado al azar por lo cual el modelamiento es poco exacto, esto se contrasta con la curva ROC analizada más adelante.

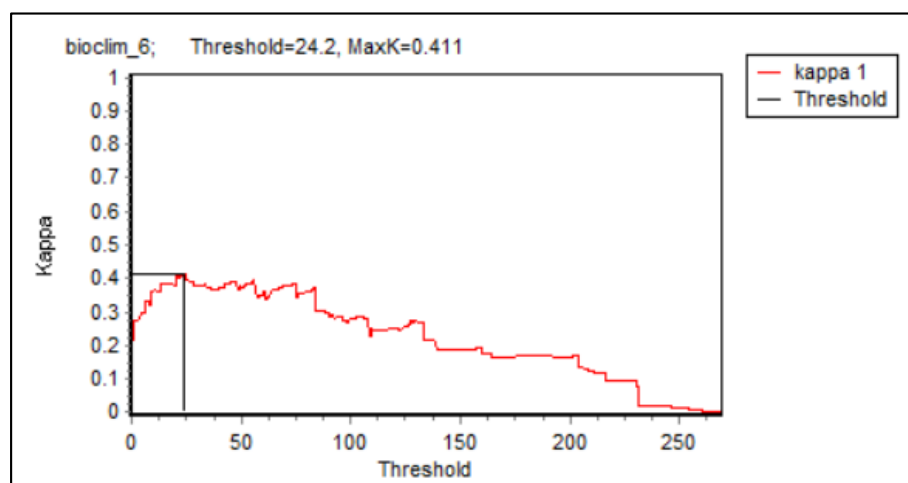


Figura 44. Estadístico Kappa del modelamiento BIOCLIM.

4.1.4 Resultados del modelo Regresión Logística

- Matriz de confusión

En la Figura 45 se muestra los resultados, como se puede observar en la matriz el error por comisión es de 9 casos mientras que el error de omisión es de 30 casos, teniendo un porcentaje global de 98.8%. Aunque el resultado de Regresión Logística está por debajo del resultado de MARS ya que existe mayor error de omisión (más crítico), este sigue siendo muy satisfactorio.

		Pronosticado		
		presencia		Porcentaje correcto
Observado	0	1		
Paso 1	presencia 0	1712	9	99.5
	1	30	1521	98.1
Porcentaje global				98.8

Figura 45. Matriz de confusión resultante de la Regresión Logística.

4.2 Análisis de la curva ROC

Las gráficas de las curvas ROC muestran que los modelos más satisfactorios son; el modelo MARS y el modelo Regresión Logística, ambos con valor de AUC de 0.99, seguidos está el modelo Maxent con 0.94 y al final el modelo BIOCLIM con un 0.77 (Figura 46, 47, 48, 49). Esto se contrasta con los estadísticos analizados anteriormente.

- Modelo Maxent

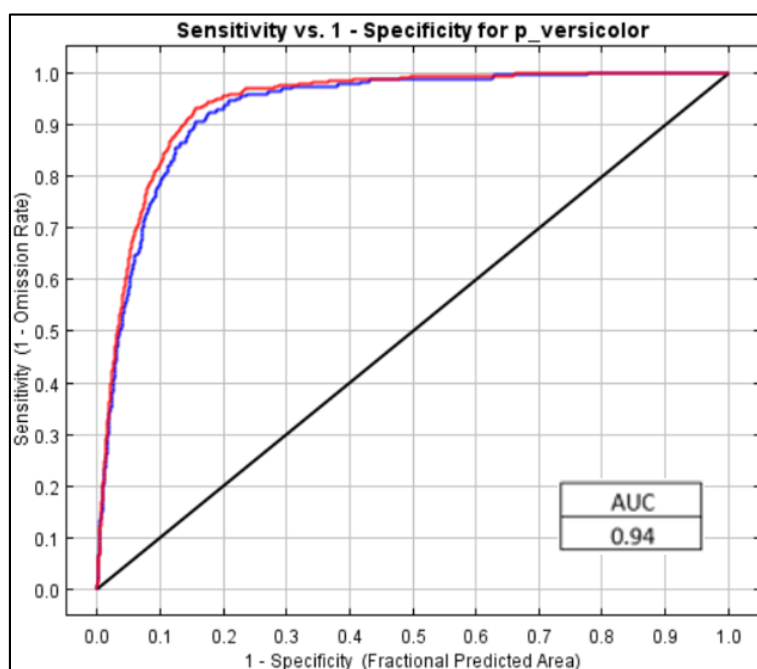


Figura 46. Curva ROC del modelo Maxent. curva entrenamiento (rojo), curva de entrenamiento (azul).

- Modelo MARS

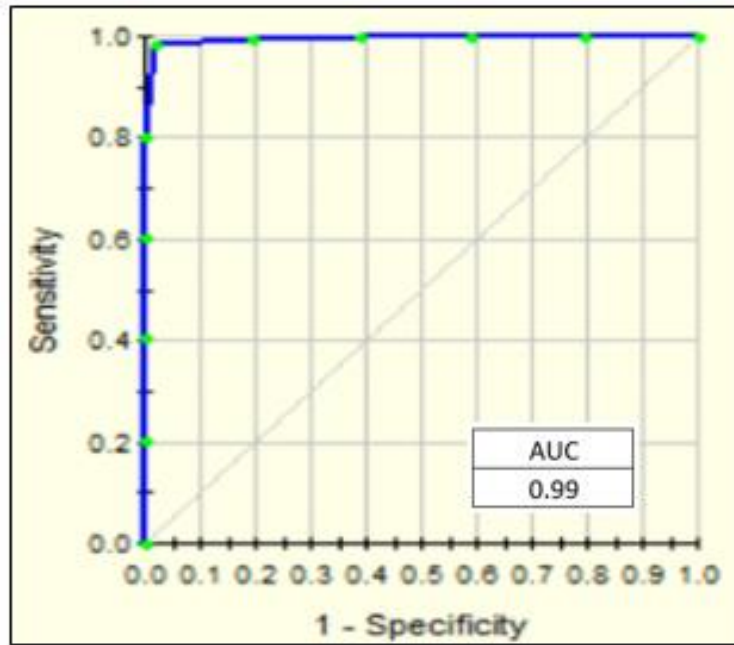


Figura 47. Curva ROC del modelo MARS.

- Modelo BIOCLIM

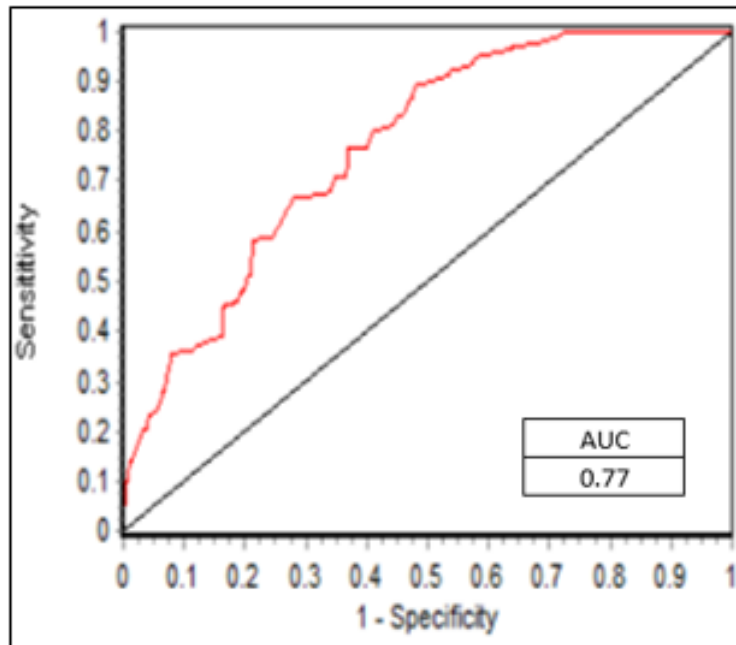


Figura 48. Curva ROC del modelo BIOCLIM.

- Modelo Regresión Logística

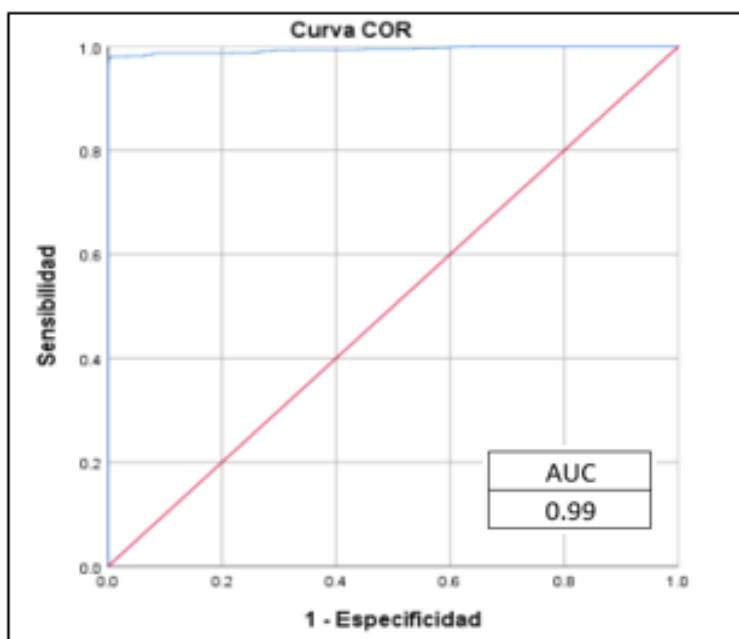


Figura 49. Curva ROC del modelo Regresión Logística.

4.3 Análisis contribución de variables independientes

La importancia de cada variable independiente o predictora al algoritmo influye en la cantidad de información que esta entrega al proceso y la cantidad de ganancia que el algoritmo final obtiene de las variables en el entrenamiento, el programa asigna el incremento en la ganancia a las variables predictoras de las cuales depende dicha característica, los modelos Maxent y MARS expresan la contribución en porcentaje y para la regresión lineal se transformó el estadístico Wald a porcentaje. Como se puede ver en la Tabla 13 la variable que más contribuyó en todos los modelos es la distancia a infraestructura antrópica, seguida a esta variable están las variables: uso de suelo en Maxent, temperatura en MARS y velocidad del viento en Regresión Logística.

Tabla 13*Contribución de las variables independientes*

Variable	Maxent	MARS	Regresión Logística
Distancia a infraestructura antrópica	68.3	100	56.85
Uso de suelo	16.7	0	1.61
Velocidad del viento	6.4	7.54	21.6
Precipitación	2.9	10.12	1.81
Temperatura	2.4	28.28	2.85
Radiación solar	1.6	11.45	0.8
Altitud	1.4	9.4	0.4
Presión de vapor de agua	0.2	22.42	14.03

En la Tabla 12 los coeficientes dados por el modelo Regresión Logística muestran que el coeficiente B es negativo para distancia a infraestructura antrópica, por lo cual, a menor distancia de la infraestructura antrópica mayor la probabilidad para encontrar la especie, y como el coeficiente para velocidad del viento es positivo se puede decir que, a mayor velocidad del viento mayor será la probabilidad de encontrar la especie. En los 3 modelos la variable con mayor influencia en la distribución de la especie es de origen antrópico.

No se presenta análisis de contribución de variables independientes del modelo BIOCLIM ya que el software DIVA-GIS no genera un producto donde se presente la ganancia que cada variable predictora tuvo en el modelamiento.

4.4 Análisis gráfico de los modelos de distribución de especies

En el análisis gráfico de los modelos se analiza las áreas donde existe alta, media y baja probabilidad de presencia de la especie, esto con la ayuda de puntos de presencia (rojo) y pseudo-ausencias (azul) elegidos al azar (Figura 50, 51, 52, 53). En la comparación gráfica de los 4 modelos se observa que, los modelos indicados por los estadísticos como más satisfactorios MARS y Regresión Logística, son los modelos que gráficamente se ajustan más a los puntos de prueba (Figura 51 y 53), pues visualmente estos modelos son los que más contienen puntos de presencia dentro de las áreas de alta y media probabilidad de presencia. En la Figura 53 se puede ver que el modelo Regresión Logística generaliza mucho más las áreas de alta probabilidad en comparación con el modelo MARS, en el cual las áreas con alta probabilidad son visualmente menores. Seguido a estos dos modelos, se encuentra el modelo Maxent, en el cual el ajuste de los puntos de prueba es muy similar a los dos modelos anteriores, con excepción del sector noroeste donde se puede apreciar que existen puntos de ausencia en sectores donde la probabilidad es media (Figura 50), esto puede deberse a que Maxent no necesita el ingreso de puntos de ausencia por lo cual el software genera de forma automática sus propias pseudo-ausencias (ver apartado 3.6.1), los cuales no toman las consideraciones mencionadas acerca de los buffer (ver sección 3.3.2). Por último, está el modelo BIOCLIM en el cual de manera evidente se puede ver que los puntos de prueba no se ajustan a las áreas de alta probabilidad, siendo este modelo poco satisfactorio (Figura 52).

- Modelo Maxent

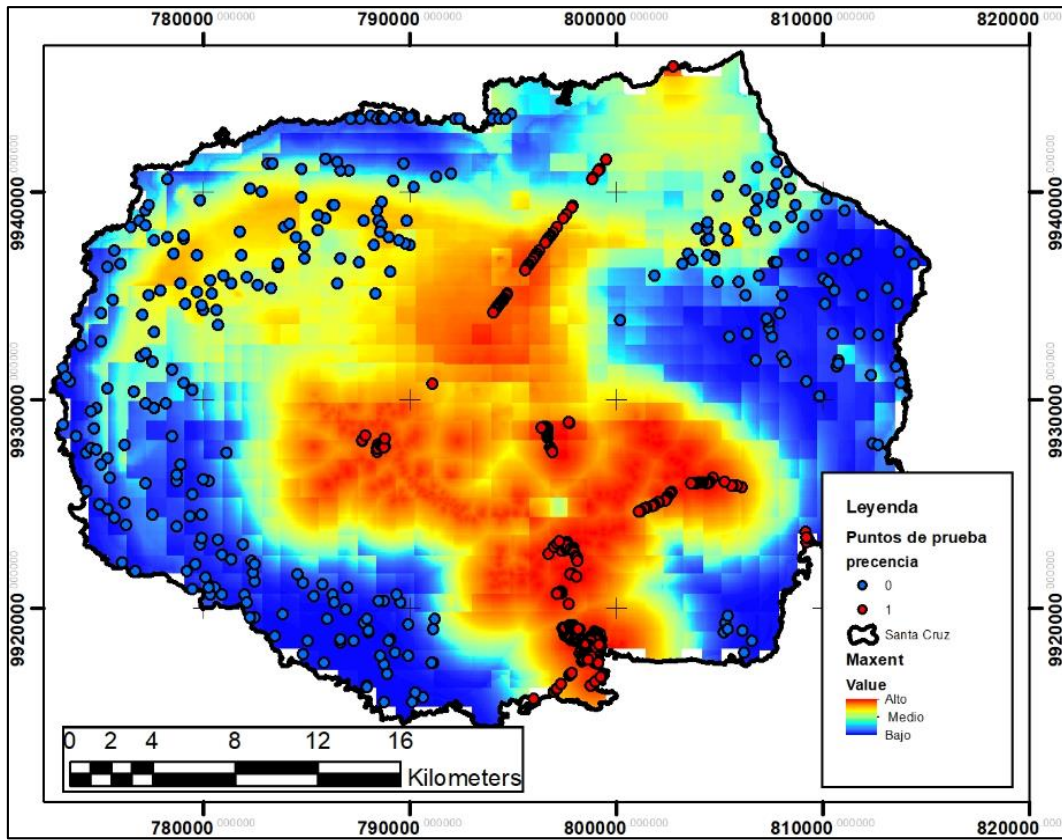


Figura 50. Modelo Maxent.

- Modelo MARS

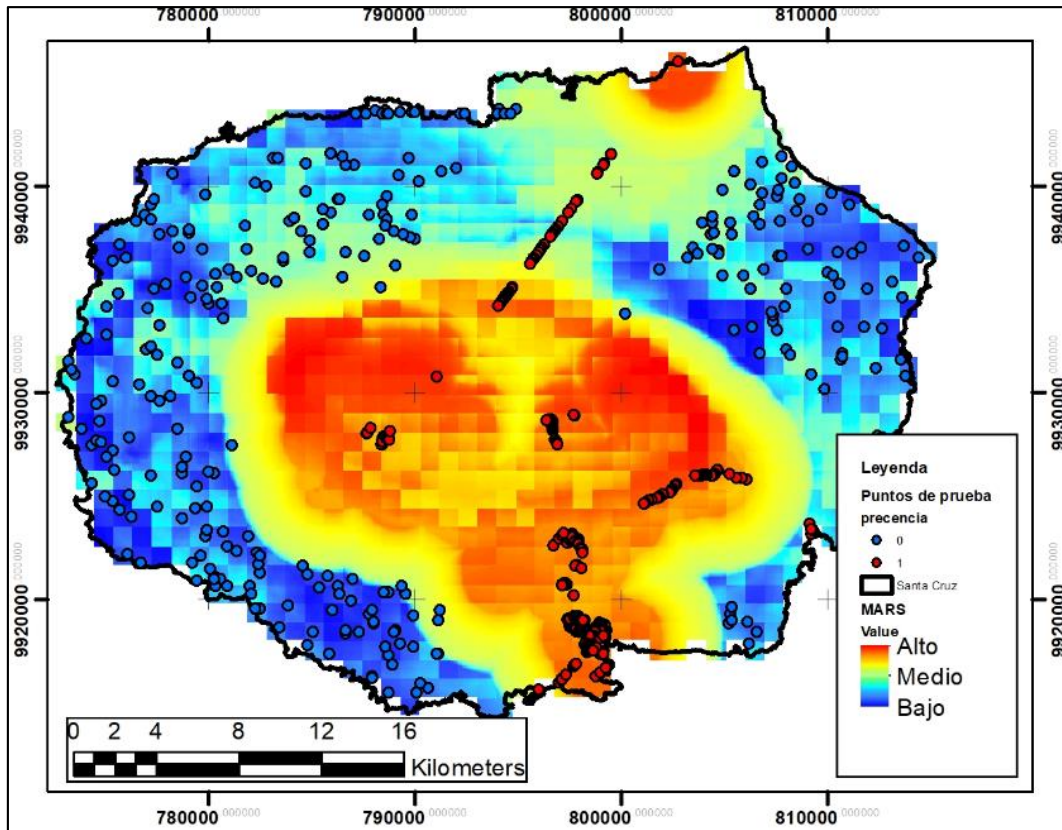


Figura 51. Modelo MARS.

- Modelo BIOCLIM

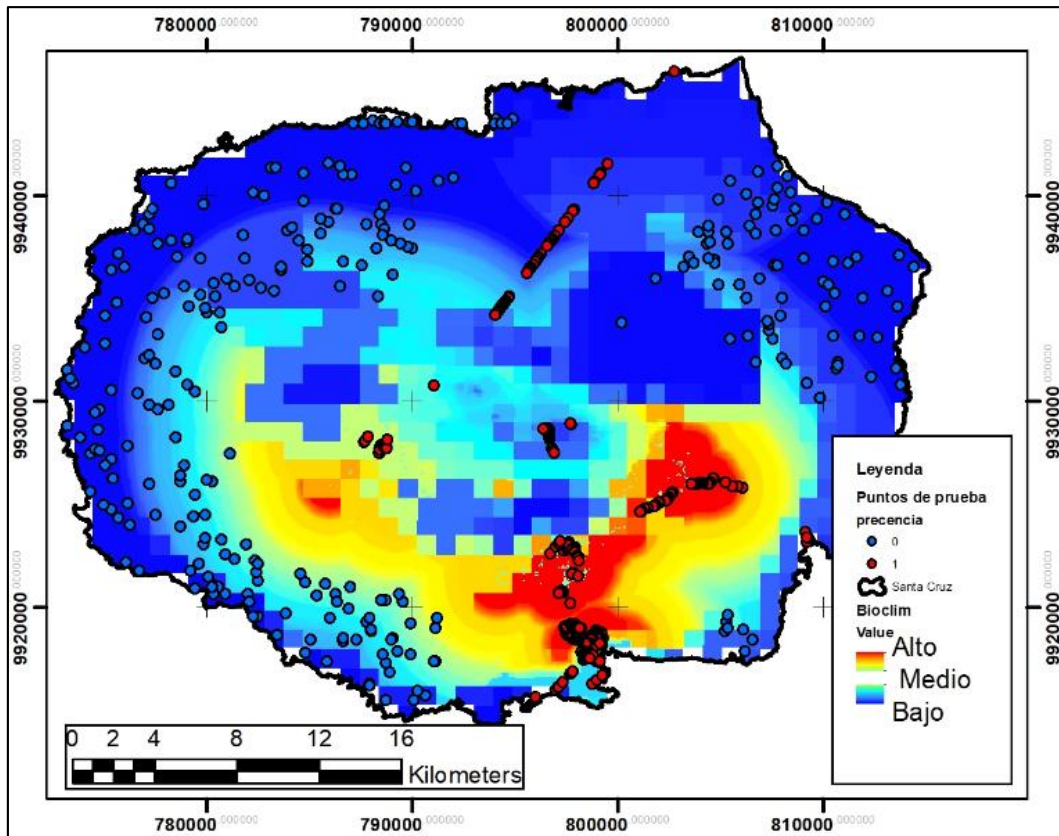


Figura 52. Modelo BIOCLIM.

- Modelo Regresión Logística

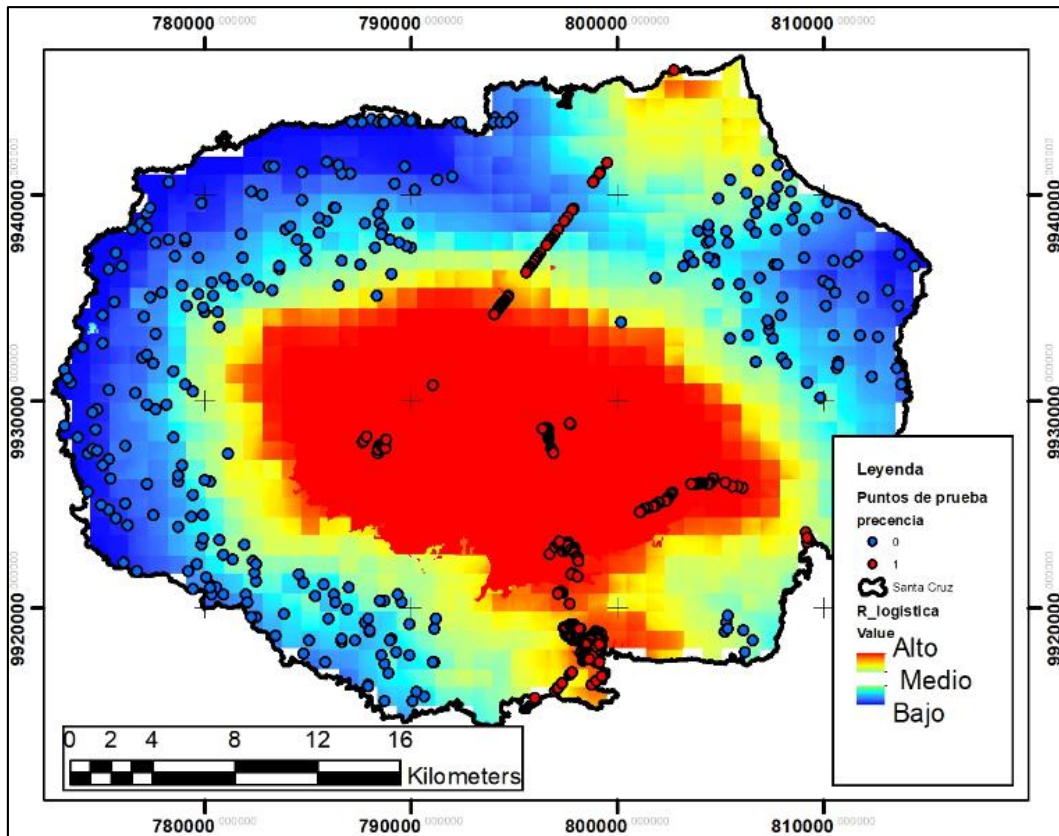


Figura 53. Modelo Regresión Logística.

También en el modelo Maxent se puede apreciar de forma más clara como la variable distancias a infraestructura antrópica influye sobre el resultado final. Como se puede apreciar en la Figura 54a, la probabilidad alta (rojo) parte de los puntos infraestructura antrópica hasta alejarse a la probabilidad baja (azul), esto se asemeja mucho a la variable distancias a infraestructura antrópica (Figura 54b), lo cual concuerda con los resultados estadísticos, donde la variable distancia a infraestructura antrópica es la variable con mayor contribución al entrenamiento y el análisis Jackknife muestra que esta variable por si sola aporta la mayor ganancia, muy por encima de las

demás (Figura 21). Por este motivo se puede deducir que en el modelo Maxent, la ganancia o contribución de la variable distancia a infraestructura antrópica es demasiado alta.

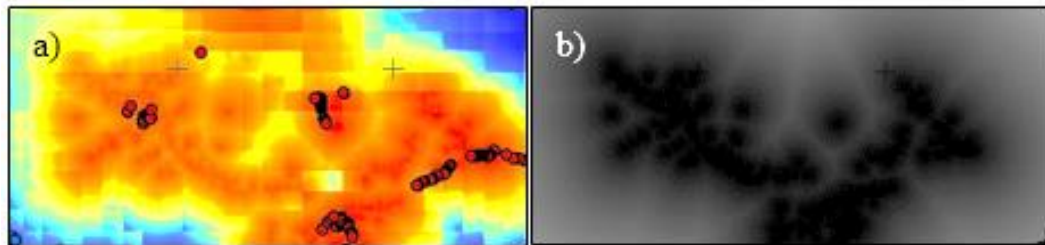


Figura 54. a) Sección del ráster del modelo Maxent. b) Sección del ráster de variable distancias a infraestructura urbana.

CAPITULO V

CONCLUSIONES Y RECOMENDACIONES

5.0 Resumen

Se aplicó 4 modelos de distribución de especie (MARS, Maxent, BIOCLIM y Regresión Logística), entre los cuales MARS fue elegido por su alto rendimiento al evaluar sus estadísticos y realizar el análisis gráfico. Además, se analizó las variables que más influyen en el comportamiento de la especie *P. versicolor* y el área donde se distribuye, concluyendo que la presencia de las variables antrópicas presenta las mejores condiciones para la presencia de la especie.

5.1 CONCLUSIONES

- MARS fue el modelo con mejor rendimiento predictivo tras tener estadísticos satisfactorios (ver apartado 4.1.2 y 4.2) y en el análisis gráfico este modelo sintetiza mejor las zonas de alta probabilidad de encontrar la especie (Figura 51), permitiendo diferenciar áreas dentro de estas zonas de alta probabilidad, que se asemejan a los puntos de presencia recogidos en campo (Figura 13). En la práctica estas áreas pueden representar una mejor sectorización de la especie, optimizando así esfuerzos y recursos.
- El método de Regresión Logística determina predicciones bastante aceptables con un rendimiento muy cercano al modelamiento MARS (ver apartado 4.1.4), pero en el análisis gráfico del modelo Regresión logística generaliza demasiado las zonas de alta probabilidad (Figura 53), lo cual se puede traducir en una mala interpretación de las áreas donde existe mayor probabilidad de encontrar a la especie (Figura 53).

- Maxent presenta estadísticos satisfactorios (ver apartado 4.1.1), sin embargo; en el análisis gráfico se puede ver que la variable distancia a infraestructura antrópica influye demasiado en la clasificación de áreas con mayor probabilidad, a tal punto que la mayor probabilidad de encontrar a la especie parte desde los puntos de infraestructura (Figura 54). Esto puede deberse a que Maxent usa sus propios puntos de ausencias (background), generados de forma aleatoria sin tomar en cuenta el buffer para generación de pseudo-ausencias que se usa en los demás modeladores.
- BIOCLIM presenta valores estadísticos poco satisfactorios (ver apartado 4.1.3), por lo cual el rendimiento de predicción para la especie *P. versicolor* es malo, esto se contrasta con el análisis gráfico donde existen muchos puntos de presencia que no se encuentran dentro de las áreas de alta y media probabilidad de encontrar a la especie (Figura 52).
- Las zonas con mayor probabilidad de presencia de la especie invasora *P. versicolor* cubre de manera general un área de 285 Km², que se extienden principalmente en las áreas pobladas de Puerto Ayora, Bellavista y Santa Rosa. Estas áreas cubren todas las zonas vegetales y ecológicas de Santa Cruz, dando a entender que en la isla no existen barreras geográficas para su distribución, sin embargo, los resultados muestran que mientras más árida y deshabitada sea la zona, menor será la probabilidad de encontrar presencia de la especie (ver Anexo 7).
- Se establece a las áreas urbanas como sitios favorables para la presencia de la especie *P. versicolor* en la isla Santa Cruz, ya que la variable con mayor contribución en los modelos de mejor rendimiento (MARS y Regresión Logística) fue la variable distancia a infraestructura antrópica, la cual presentó mayor porcentaje de contribución y tienen relaciones proporcionales a la probabilidad de encontrar a la especie.

- Las áreas de vegetación invasora y de uso agropecuario también son favorables para la presencia de la avispa *P. versicolor* ya que la variable uso de suelo, aunque contribuía poco porcentaje al entrenamiento el coeficiente era positivo en el modelo Regresión Logística.
- Al no existir barreras geográficas, reportes de extinción o control eficaz y ningún reporte de enemigos naturales que restrinjan la distribución de la especie *P. versicolor*, se puede determinar que en la isla Santa Cruz un modelo de nicho ecológico (MNE) puede ser tomado como un modelo de distribución de especies (MDE) ya que sus resultados de predicción son las mismas.

5.2 RECOMENDACIONES

- Para mejorar la base de datos se puede contar con registros georreferenciados de presencia obtenidos en investigaciones anteriores y registros históricos, así como una actualización de las variables independientes para de esta forma obtener una predicción más precisa de las zonas con mayor probabilidad de presencia de la especie.
- Ya que este estudio se centró en modelar el comportamiento de la especie en el semestre más frío del año (junio - noviembre) es pertinente realizar un análisis de la distribución de la especie con las 19 variables que proporciona Worldclim (las cuales no responden al semestre más frío) y realizar una comparación.
- Ya que esta es una especie que se adapta fácilmente al entorno de las Islas Galápagos es necesario incluir en líneas futuras de investigación, estudios entomológicos los cuales aporten a las variables predictoras y de esta manera mejorar la predicción de áreas con mayor probabilidad de presencia.
- Es pertinente realizar un modelamiento de la distribución de la especie bajo un escenario de cambio climático con la ayuda de las variables ambientales de escenarios futuros proporcionadas por Worldclim de este modo se podrá entender como la especie se esparcirá espacialmente en el futuro.
- Para el control y erradicación de la especie *P. versicolor* se recomienda focalizar los recursos y estrategias en las zonas pobladas y de agricultura ya que estas áreas son el principal foco de donde parte la mayor probabilidad de encontrar presencias.

REFERENCIAS BIBLIOGRÁFICAS

- Abedrabbo, S. (1991). *Nueva avispa introducida en las Islas. Carta informativa 31:4*. Santa Cruz, Galápagos. Ecuador.
- Albelo, R., & Causton, C. (1999). *El Nino and introduced insects in the Galapagos Islands: different dispersal strategies, similar effects*. Noticias de Galápagos.
- Baert, L. (1995). *Notas sobre el estado de los artrópodos terrestres en galapagos*. Obtenido de <https://core.ac.uk/download/pdf/11021808.pdf>
- Benavides, A. (2019). *Curvas ROC y sus aplicaciones*. Sevilla: Universidad de Sevilla.
- Berlanga, V., & Vilá, R. (2014). *Como obtener un modelo de regresión logística binaria con SPSS*. Revista d'Innovació i Recerca en Educació.
- Bioestadística. (2019). *Introduccion a la regresion logistica*. Obtenido de <http://halweb.uc3m.es/esp/Personal/personas/amalonso/esp/bstat-tema9.pdf>
- Bonilla, G. (2006). *Estado de conocimiento de especies invasoras propuesta de lineamiento para el control de los impactos*. Instituto de investigacion de recursos biológicos Alexander Von Humboldt.
- Booth, T. (2018). *Porque entender las continuas contribuciones de Bioclim a los modelos de distribucion de especies es importante*. Wiley online library.
- Brito, S., Montagana, T., & Maia, F. S. (2015). *Cuticular signature in the development of Polistes versicolor*. Obtenido de <http://www.funpecrp.com.br/gmr/year2015/vol14-4/pdf/gmr6676.pdf>
- Busby, J. (1991). *BIOCLIM – a bioclimate analysis and prediction system*. Plant Protect.
- Castro, F. (2006). *Técnicas para el inventario y muestreo de anfibios: Una Compilacion*. Colombia: Universidad del Valle.
- Conabio. (1 de mayo de 2019). *Validacion del modelo*. Obtenido de <http://nicho.conabio.gob.mx/home/proposito-y-guia-del-usuario/validacion-del-modelo>
- Consejo de Gobierno del Régimen Especial de Galápagos. (2016). *Plan de Desarrollo Sustentable y Ordenamiento Territorial del Régimen Especial de Galápagos*. Puerto Baquerizo Moreno, Galápagos, Ecuador.
- Cortez, E. (2010). *Métodos estadísticos de evaluación de la concordancia y la reproducibilidad de pruebas diagnósticas*. Medellín: Revista colombiana de educación médica.

- Cruz, G. s. (10 de Febrero de 2019). *Turismo gad Santa Cruz*. Obtenido de <http://turismo.gadsantacruz.gob.ec/index.php>
- Cubero, C. (2019). *Sistemas inteligentes de gestión*. Granada: Universidad de granada.
- Darwin, C. (1980). *Lórigine des especes*. París: La Découverte.
- Della Pietra & Lafferty, J. (1997). *Inducting features of random fields*. IEEE Tran.Pattern Anual.Mach.Intell.
- Desender, Baert, & Maelfait. (1989). *Resultados de la expedicion Belga 1986: Aranae, and provisional cheklist of the spiders of the Galápagos Archipelago*. Boletín van het Koninklijk Belgisch Instutuut voor Natuuwetenschappen Etmologie.
- Dirección del Parque Nacional Galápagos. (2017). *Informe Gestión 2017*. Santa Cruz, Galápagos. Obtenido de http://www.galapagos.gob.ec/wp-content/uploads/downloads/2018/04/2017_informe_rendicion_de_cuentas_dpng.pdf
- Dubik, M. &. (2004). *Performance guarantees for regularized maximum entropy density estimation*. New York: Proceedings of the 17th Annual Conference on Computational learnings theory, ACM Press.
- Elith, J. (2006). *Novel methods improve prediction of species' distributions from occurrence data*.
- Escuela Superior Politécnica de Chimborazo. (2019). *Científicos de la epoch registran una nueva familia de insectos en las islas galápagos "el hallazgo de sclerogibbidae"*. Riobamba, Ecuador: Eepoch.
- Estación Científica Charles Darwin. (2006). *Especies invasoras de las Galápagos*.
- Ezcurra, & Ruggiero. (2003). *Regiones y transiciones biogeograficas: Comlementaridad de los analisis en biogeografica historica y ecologica.Una perspectiva latinoamericana de la biogeografía*. UNAM, MEXico, D.F.: Las Prensas de Ciencias.
- Fabián, B., & Lincango, P. (2006). *Manual de monitoreo y vigilancia de invertebrados terrestres introducidos*. Santa Cruz, Islas Galápagos: Departamento de invertebrados terrestres Fundacion Charles Darwin.
- FAO. (2000). *El AHP y su aplicacion para determinar los usos de las tierras*. Santiago de Chile.
- Felicísimo, A., & Muñoz, J. (2011). *Modelos de distribución de especies: Una revisión sintética*. Revista chilena de historia natural.
- Fernández, S. d. (2011). *Regresion logistica*. Madrid: Universidad autonoma de Madrid.

- Fitter, J., Fitter, D., & Hosking, D. (2016). *Wild life of the Galápagos*. Princeton. Foundation, C. D. (1 de mayo de 2019). *Datos meteorologicos*. Obtenido de <https://www.darwinfoundation.org/es/datazone/clima>
- Foundation, C. D. (20 de junio de 2019). *Fundación charles darwin* . Obtenido de <https://www.darwinfoundation.org/en/about>
- Franklin. (1995). *Predictive vegetation mapping: Geographic modelling of biospatial patterns in relation to enviromental gradients*.
- Friedman, H. (1991). *Multivariate adaptative regression splines*. The annals of statistics.
- Fundacion Charles Darwin. (2018). *InformeAnual_2018*. Obtenido de <https://www.darwinfoundation.org/es/publicaciones/informe-anual>
- Galápagos, C. d. (2016). *Plan de Desarrollo Sustentable y Ordenamiento Territorial del Régimen Especial de Galápagos*. Santa Cruz.
- Galapagos, N. (1 de mayo de 2019). *Clima de Galápagos*. Obtenido de <https://naturegalapagos.com/es/clima-de-galapagos/>
- Gidahatari. (2017). *Modelamiento de distribucion de especies de flora y fauna*. Obtenido de <http://gidahatari.com/ih-es/modelamiento-de-distribucion-de-especies-de-flora-y-fauna-con-maxent-codigo-libre>
- Graham, Ferrier, Huettman, Mortis, & Peterson. (2004). *New development in museum-based informatics and applications in biodiversity analysis*. Trends in Ecology and Evolution.
- Grenier, C. (2007). *Conservación contra natura. Las islas Galápagos*. Edotorial Abya Yala.
- Guisan, A., & Thuiller, W. (2005). *Predicting species distribution*. En *offering more than simple habitat models* (págs. 993-1009).
- Guzmán, R. (2016). *La riqueza biológica de los insectos: análisis de su importancia multidimensional*. Obtenido de http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0065-17372016000300370
- Heraty, J., & Stewart, B. (1998). *Introduced insect fauna of an oceanic archipelago: The Galápagos Islands, Ecuador*. American Entomologist.
- Hutchinson, M. (1995). *Interpolating mean rainfall using thin plate smoothing splines*.
- Hurchinson, G. (1978). *An introduction to population ecology*. New Haven: Yale University Press.

- Hurtado, T. (2019). *Proceso de analisis jerárquico (AHP)*. San Marcos: Universidad nacional mayor de San Marcos.
- Hutchinson, G. (1957). "Concluding remarks" *Cold spring harbor Symposia on Quantitative Biology*.
- Hutchison, T. X. (2000). *Anuclim version 6.1 User guide*. the australian national university.
- Illoldi, P., & Tanla, E. (2008). *De los modelos de nicho ecológico a las áreas de distribución geográfica*. Obtenido de https://www.researchgate.net/profile/Escalante_Tania/publication/286336295_De_los_modelos_de_nicho_ecologico_a_las_areas_de_distribucion_geografica/links/568a9b4008aebccc4e1a0c38.pdf
- INAMHI. (2014). *Anuario metereologico*. Quito: INAMHI.
- INAMHI. (2018). *Boletin Agroclimatico decadal*. Quito: INAMHI.
- Jackson, M. H. (1993). *Galapagos a nature history*. Alberta, Canada : University of Calgary Press.
- Jaynes. (1957). *Information theory and statistical mechanics*.
- Academy, K. (5 de septiembre de 2019). *Khan Academy*. Obtenido de <https://es.khanacademy.org/science/biology/ecology/population-ecology/a/population-size-density-and-dispersal>
- Lara, E., & Simeón, C. (2006). *Los sistemas de informacion geográfica*. Geoenseñanza.
- Lasso, T. (1997). *Ecología e impacto de la avispa introducida (Polistes versicolor, Vespidae-Hymenoptera) en las islas Floreana y Santa Cruz*. Quito, Ecuador: Universidad Central del Ecuador.
- Llorente, J. (2001). *Introducción a la biogeografía en latinoamerica: teorías, conceptos, métodos y aplicaciones*. (J. Llorente, & J. Morrone, Edits.) México, D.F., México: Facultad de ciencias UNAM.
- López, F. (2015). *Control estricto de matrices de confusión por medio de distribuciones multinomiales*. España: Geofocus.
- López, M. (2007). *Descripción y caracterización de nichos ecológicos: una visión mas cuantitativa del espacio ambiental*. Guanajuato: Centro de investigacion en matemáticas.
- Maciel, C., & Manríquez, N. (2015). *El área de distribución de las especies: revisión del concepto*. Pachuca-Tulancingo: Universidad Autónoma del Estado de Hidalgo.

- Martillo, C., Pazmiño, A., Chunga, K., & Resl, R. (2011). *Descripción morfológica de la plataforma continental del Ecuador y análisis comparativo con la plataforma jurídica mediante la aplicación de sistemas de información geográfica*. Acta oceanográfica del pacífico .
- Masciocchi, M. (2013). *Dinámica de la dispersión y actividad de la avispa exótica *Vespula germanica* F. (Hymenoptera: Vespidae) en Argentina*. Obtenido de <http://rdi.uncoma.edu.ar/bitstream/handle/123456789/181/Masciocchi-TEISIS%20DOCTORAL.pdf?sequence=1&isAllowed=y>
- McGlynn. (1999). *The worldwide transfer of ants: geographical distribution and ecological invasions*. Journal of Biogeography.
- Meffe, C. &. (1997). *Principles of conservation*. Sunderland, Massachusetts.: Sinauer Associate.
- Milborrow, S. (2008). *Plot of hinge functions to illustrate Friedman's Multivariate Adaptive Regression Splines (MARS)*.
- Ministerio. (14 de noviembre de 2014). *Estrategia de gestión, control y posible erradicación del avispón asiático o avispa negra (*vespa velutina ssp nigrithorax*) en españa* . Obtenido de https://www.miteco.gob.es/gl/biodiversidad/publicaciones/estrategia_vespavelutina_tcm37-69976.pdf
- Ministerio del Ambiente. (1 de enero de 2016). *Proyecto de control y erradicación de especies invasoras prioritarias para reducción de la vulnerabilidad de especies endémicas de las islas galapagos*. Obtenido de http://www.galapagos.gob.ec/wp-content/uploads/downloads/2016/08/Proyecto_control_y_errad.pdf
- Minitab. (2018). *Introducing Mars*. Obtenido de http://media.salford-systems.com/pdf/spm8/IntroMARS_v_8_2.pdf
- Nakamura, M., & Soberon, J. (2007). *Some thoughts on the notions of areas of distribution and environmental niches as related to availed data*.
- Navarra, M. (3 de Junio de 2019). *Radiación solar*. Obtenido de http://meteo.navarra.es/definiciones/radiacion_solar.cfm
- Nix, H. (1986). *A biogeographic analysis of Australian elapid snakes*. In *Atlas of Elapid Snakes of Australia: Australian Flora and Fauna Series* . Canberra.
- Olaya, V. (2014). *Sistemas de Información Geográfica*. Obtenido de https://www.icog.es/TyT/files/Libro_SIG.pdf
- Paredes, N. (1986). *Determinación del datum vertical en La Libertad, Ecuador*. Guayaquil, Ecuador: INOCAR.

- Parent, C. (2000). *Life-cycle and ecological impact of Polistes versicolor (Olivier) (Hymenoptera: Vespidae), an introduced predatory wasp on the Galápagos Islands, Ecuador*. Ottawa: Carleton University.
- Parque Nacional Galápagos. (26 de octubre de 2019). *Especies invasoras*. Obtenido de <http://www.galapagos.gob.ec/especies-invasoras/>
- Peck, S. (1996). *Diversity and distribution of the orthopteroid insects of the Galápagos Islands*.
- Peck, S. (2006). *The beetles of the Galapagos Islands, Ecuador: Evolution, ecology, and diversity (Insecta: Coleoptera)*. Ontario, Canada: NRC-CNRC.
- Phillips, S. (2007). *Modeling of species distribution with Maxent: next extension and a comprehensive evaluation*. Princeton.
- Phillips, S. (2019). *Guía didáctica Maxent*. Biodiversity informatics.
- Pleguezuelos. (2002). *Las especies introducidas de anfibios y reptiles*. Madrid: Dirección General de la Conservación de.
- Rubén, M., & Felicísimo, Á. (2011). *Modelos de distribución de especies: Una revisión sintética*. Chile: Revista Chilena de Historia Natural.
- Salamanca, U. d. (mayo de 2019). *Servicio transfronterizo de información geográfica*. Obtenido de <http://www.stig.usal.es/quehacemos.php>
- Sánchez, J. (2016). *Análisis de calidad cartográfica mediante el estudio de la Matriz de Confusión*. Medellín. Colombia: Pensamiento matemático.
- Sánchez, V. (2001). *Modelado de la distribución de especies y la conservación de la diversidad biológica*. Mexico, D.F.: UNAM.
- Schapire, S. P. (2005). *Maximum entropy modeling of species geographic distributions*. Princeton: Computer Science Department, Princeton University. Obtenido de <https://www.cs.princeton.edu/~schapire/papers/ecolmod.pdf>
- Shannon, C. (1948). *A mathematical theory of communication*. Bell Syst. Tech.
- Star, J., & Estes, J. (1990). *Geographic Information Systems: An Introduction*. Prentice-Hall.
- Swets. (1988). *Measuring the accuracy of diagnostic systems*. Science.
- Tomlin. (1990). *Geographic information systems and cartographic modelling*. Prentice Hall.
- Torge, W. (2001). *Geodesy* (3era. ed.). Berlín - New York: W. de Gruyter.

Vanegas, J. (2017). *Multivariate Adaptive Regression Splines(MARS), una alternativa para el análisis de series de tiempo*. Gaceta Sanitaria.
doi:<https://dx.doi.org/10.1016/j.gaceta.2016.10.003>

Worldclim. (2 de agosto de 2019). *Worldclim métodos*. Obtenido de
<https://www.worldclim.org/methods1>

Zimmermann, & Guisan. (2000). *Predictive habitat distribution models in ecology*. *Ecological Modelling*.

ANEXOS