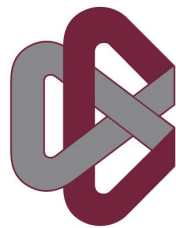

Análisis del Nicho Fundamental.



CIMAT

Centro de Investigación en Matemáticas, A.C.

TESIS DE MAESTRÍA

Desireé Fernanda Soto Galindo

Centro de Investigación en Matemáticas, A.C.

Noviembre 2016

Documento maquetado con T_EX!S v.1.0.

Este documento está preparado para ser imprimido a doble cara.

Análisis del Nicho Fundamental.

Tesis que presenta

Desireé Fernanda Soto Galindo

Para obtener el grado de

**Maestra en Ciencias con Especialidad en Probabilidad y
Estadística**

Dirigida por el Doctor

José Andrés Christen Gracia

Centro de Investigación en Matemáticas, A.C.

Noviembre 2016

Copyright © Desirée Fernanda Soto Galindo

*A mis padres, a mis hermanos
y
a mi compañero incondicional*

Agradecimientos

Querer es una gran cosa, porque la Actividad y el Trabajo son consecuencia generalmente de la Voluntad, y casi siempre el Trabajo va acompañado del Éxito. Trabajo, Voluntad y Éxito llenan la vida del hombre. La Voluntad abre las puertas del Éxito con brillantez y felicidad; el Trabajo hace pasar a través de estas puertas, y al final del viaje el Éxito corona los esfuerzos realizados.

Louis Pasteur

Gracias, de corazón, a todas las personas que, de una manera u otra, han sido claves en mi crecimiento tanto profesional como personal.

Mi más amplio agradecimiento para el Dr. Andrés Christen, por aceptarme para realizar esta tesis bajo su valiosa dirección, por su paciencia durante el desarrollo de la misma y por compartirme su conocimiento durante los cursos que pude tenerlo como profesor. Así mismo, debo agradecer a la Mtra. Laura Jiménez que siempre se mostró dispuesta a resolver o discutir cualquier duda que me surgiera en relación a este proyecto. También, gracias al Mtro. Nicolás Kuschinski por sus aportaciones con respecto a la herramienta clave en esta tesis para la interpretación de los resultados.

También quisiera hacer patente mi agradecimiento y admiración a todos los profesores que tuve en la maestría, porque gracias a ellos creció mi pasión por la estadística y la probabilidad. Gracias por su dedicación, apoyo, comprensión y por compartir sus experiencias. De igual forma, gracias a mis colegas y compañeros, aprendí de ellos y disfruté su compañía.

Desde luego, llego al final de esta etapa gracias al apoyo, amor y motivación de mis padres, hermanos y amigos. A mi padre, quien me inculcó el amor al aprendizaje y siempre me inspiró para superarme a mi misma. Agradezco, además, todo su apoyo y sus cuidados. A mi madre, por su cari-

ño, por su preocupación y por hacerme verdaderamente una mejor persona. A mis hermanos, que también influyen en la persona que soy y fueron un pilar importante para que yo pudiera llegar hasta aquí. Gracias a ellos por su alegría y su amor. Gracias también, a todos mis amigos, a los de aquí y a los de allá. Especialmente, gracias a Pedro Orozco, por sus consejos, su cercanía y por permitirme aprender de él y con él, durante estos dos años.

Para terminar, agradezco de todo corazón a Hamid, mi compañero, mi guía, por darme su apoyo incondicional, por motivarme a dar lo mejor de mí misma y porque siempre ha creído en mí. Gracias por enseñarme la importancia de la voluntad y el esfuerzo, ya que lo que no tiene mérito no se disfruta realmente.

Resumen

Si quieres conocer a una persona, no le preguntes lo que piensa sino lo que ama.

San Agustín

El *nicho ecológico* de una especie animal o vegetal se define como las combinaciones de condiciones ambientales o recursos que permiten a la especie existir. En esta tesis, el objeto de estudio se encuentra en el llamado *nicho fundamental*, el cual incluye los límites de este conjunto de combinaciones, todo lo que la especie soporta. Se presenta una propuesta basada en modelos de probabilidad para la construcción del nicho fundamental de una especie de interés, en un espacio bidimensional (dos factores ambientales). A partir de datos de presencia de la especie e información previa acerca de los rangos de tolerancia de temperatura y precipitación de la especie, se realiza un ajuste a la distribución normal sesgada bajo un enfoque bayesiano. Con esto, se obtiene una estimación puntual del nicho y una gráfica para visualizar la variabilidad. El método se implementa en el software Python. Se comentan las cuestiones computacionales. Finalmente, se realizan las interpretaciones de los nichos de 3 especies de plantas.

Índice

Agradecimientos	VII
Resumen	IX
1. Introducción	1
2. Antecedentes generales del problema	5
2.1. Nichos ecológicos	5
2.2. Datos e información a priori	6
2.3. Inferencia bayesiana	9
2.3.1. Métodos Markov Chain Monte Carlo	9
2.4. Distribución normal–sesgada	11
2.4.1. \mathcal{SN} Univariada	11
2.4.2. \mathcal{SN} Bivariado	12
3. Solución propuesta	15
3.1. Supuestos	16
3.2. Nichos no elipsoides	17
3.3. Inferencia	18
3.3.1. Verosimilitud	18
3.3.2. Distribución a priori	21
3.3.3. Función Energía	27
3.4. Gráfica de incertidumbre	29
3.5. Programación	31
3.5.1. T–walk en Python	32
3.5.2. Comparación de velocidad: Python vs. R	33
3.5.3. Problemas computacionales	34
4. Ejemplos	37
4.1. Género <i>Arachis</i>	37
4.2. Género <i>Chenopodium</i>	38
4.3. Género <i>Panicum</i>	40

4.4. Otros experimentos	40
4.4.1. Ponderaciones más contrastantes	40
4.4.2. Sin ponderaciones	41
4.4.3. A priori con mayor incertidumbre	41
5. Discusión y conclusiones	49
Bibliografía	51
A. Densidad \mathcal{SN} bivariada	53
B. Jacobiano de la transformación para el parámetro de sesgo.	55
Índice alfabético	57

Índice de figuras

2.1. Ejemplo datos	7
2.2. Biomas propuestos por Whittaker (1970)	8
3.1. Estructura de un nicho como respuesta a las condiciones X y Y	17
3.2. Efecto en las ponderaciones para distintos valores de a	21
3.3. Valor esperado a priori de (μ, A, δ)	22
3.4. Distribución a priori para el valor esperado de $X \in \mathbb{R}^2$	23
3.5. Sesgo para distintos valores de β	25
3.6. Variabilidad a priori de la matriz de precisión A	26
3.7. Histograma de eigenvalores para las elipses (score 0.15) simuladas	27
3.8. Intensidad del color según el número de simulaciones.	30
3.9. Ejemplo gráfica de incertidumbre	31
3.10. Comparación de resultados: t-walk vs. RWMH	33
3.11. Comparación de velocidad con R y Python	34
4.1. Datos género <i>Arachis</i>	38
4.2. Resultado género <i>Arachis</i>	39
4.3. Datos género <i>Chenopodium</i>	40
4.4. Resultado género <i>Chenopodium</i>	42
4.5. Modas de la posterior para la especie <i>Chenopodium</i>	43
4.6. Datos género <i>Panicum</i>	43
4.7. Resultado género <i>Panicum</i>	44
4.8. Estimaciones para $a = 1$	45
4.9. Estimaciones para $a = 0$	46
4.10. Estimaciones con mayor varianza a priori para μ y A	47

Capítulo 1

Introducción

Mira profundamente en la naturaleza y entonces, comprenderás todo mejor.

Albert Einstein

Una población de una especie, animal o vegetal, reacciona a distintos tipos de factores ambientales que afectan su crecimiento, reproducción o supervivencia. El término “nicho” surge en este contexto, haciendo referencia al lugar que ocupa la especie en un espacio p -dimensional, donde cada dimensión es un factor ambiental o recurso. Conocer la estructura del nicho permite a los ecólogos entender los requerimientos de la especie, su interacción con el ambiente, cómo se relaciona con otros organismos y cómo se adapta a distintos ecosistemas.

Ciertamente, los factores ambientales que afectan a una especie son variados. Éstos comprenden todas las características biológicas y físico-químicas del lugar dónde la población crece, vive e interactúa. Con frecuencia, algunos autores han identificado a las variables temperatura y cantidad de lluvia como los dos factores que afectan fuertemente en el tipo de animales o plantas que crecen en una región. Whittaker (1970), por ejemplo, propone una clasificación de la flora y la fauna en base a estas dos variables. Esta clasificación constituye el concepto de Biomas.

Existen varios conceptos que se relacionan con el de nicho, lo que ha causado confusión y malinterpretaciones. Así mismo, refiriéndose al nicho de una especie se puede hablar de varios tipos. En este trabajo, el objeto de estudio consiste en el llamado *nicho fundamental* que, según la definición de Hutchinson (1957), delimita la tolerancia de la especie a las variables ambientales consideradas. Es decir, es el conjunto p -dimensional que contiene todas las combinaciones con las que una especie puede sobrevivir. Reducciones de este nicho se identifican al tomar en cuenta otros factores. Generalmente,

para un tiempo y región fijos, la especie ocupa un nicho más angosto.

En estudios de ecología, se ha abordado el problema de modelar estadísticamente la estructura de diferentes tipos de nichos. Para ello, son necesarios datos de ocurrencia de la especie y los registros de las condiciones ambientales del lugar donde se observó. Estas ocurrencias están restringidas a las condiciones ambientales existentes para la región de observación, provocando que no necesariamente se observe más a la especie en las combinaciones ambientales de mayor preferencia para ésta. En otras palabras, la accesibilidad que tiene la especie a las distintas combinaciones ambientales cambia según la región y el tiempo de observación, lo que altera la probabilidad de observar un espécimen en las distintas combinaciones. Conociendo y entendiendo la manera en que se altera esta probabilidad, es posible contrarrestar el efecto en la modelación. En este sentido, es una ventaja que actualmente se tenga acceso a la información geográfica y climática de amplias regiones geográficas vía satélite. Esto permite conocer el espacio de “fondo” para la región de observación, es decir, las condiciones ambientales que realmente estuvieron disponibles para la especie.

Además de la accesibilidad a las condiciones ambientales, existen otros factores que también influyen en la probabilidad de observar un espécimen. Un ejemplo son las relaciones de competencia con otras especies. Otro ejemplo es el hecho de que algunas zonas de difícil acceso (montañas, zonas pantanosas, etc.) no son visitadas por humanos. Como resultado, habrá conjuntos de combinaciones ambientales para las cuales no se registre la presencia de la especie aunque pertenezcan a su nicho fundamental. Esta falta de información se puede recompensar incluyendo conocimiento previo acerca de la tolerancia de la especie a los factores ambientales considerados, que se consigue por medio de un “experto” que ha estudiado a la especie de interés.

Basada en las ideas de Christen y Soberón (2016), esta tesis tiene como objetivo medir la preferencia de una especie a cualquier conjunto (en \mathbb{R}^2) de condiciones ambientales usando modelos de probabilidad. Se propone delimitar el nicho fundamental de la especie con el contorno (de cierto “nivel de preferencia”) de una densidad estimada. Más aún, se propone una representación gráfica de la variabilidad de dicha estimación. Lo anterior, a partir de datos de presencia de una especie de interés e información previa acerca de la preferencia de la especie a las condiciones ambientales. Para incluir ambos tipos de información se adapta un enfoque bayesiano.

En contraste con lo descrito en Christen y Soberón (2016), se busca que las estructuras estimadas para los nichos no sean necesariamente elípticas. Para lograr la flexibilidad deseada, la distribución normal sesgada bivariada

resulta útil. Es preciso comprender sus propiedades básicas para una correcta asignación de distribución a priori a sus parámetros.

El nicho fundamental de una especie es un conjunto que no cambia con el tiempo y es independiente de la región geográfica. Conocerlo es de especial relevancia para entender la respuesta de la especie a distintos ambientes. En base al nicho fundamental de las especies se puede predecir, por ejemplo, el cambio de flora o fauna de una región como respuesta al cambio climático.

?

Capítulo 2

Antecedentes generales del problema

2.1. Nichos ecológicos

En ecología, se habla del *nicho ecológico* de una especie (animal o vegetal) para referirse al conjunto de condiciones del ambiente y/o de los recursos en el que la especie sobrevive. Este término se presenta con frecuencia acompañado de otros conceptos. Uno muy relacionado es el de *distribución geográfica*. La diferencia radica en que la distribución de una especie se ubica en una región geográfica (latitud y longitud), mientras que el nicho se encuentra en el llamado *espacio ambiental* (\mathcal{A}). Este espacio es un subconjunto de \mathbb{R}^d , donde cada dimensión corresponde a una variable ambiental. Para un tiempo t fijo, a cada región geográfica le corresponde un conjunto de variables ambientales. El inverso no necesariamente se cumple ya que no todas las condiciones ambientales necesariamente existen en un tiempo dado. El subconjunto en el espacio ambiental inducido por una región geográfica G al tiempo t se conoce como *espacio existente* o *espacio disponible* ($\mathcal{E}(t; G)$). (Peterson, Soberón, Pearson, Anderson, Martínez-Meyer, Nakamura y Araújo, 2011)

Tomando en cuenta la última definición se pueden identificar 2 nichos distintos:

N_F. Nicho Fundamental: Conjunto de combinaciones de condiciones ambientales y/o recursos ($\mathbf{N}_F \subset \mathcal{A}$) que permitirían existir a una especie indefinidamente. (Hutchinson, 1957)

N_P(t**; **G**)**. Nicho Existente, Disponible o Potencial: Puntos del nicho fundamental

que existen al tiempo t y la región G .

Además, hay un tercer nicho contenido en los 2 anteriores:

$\mathbf{N}_R(\mathbf{t}; \mathbf{G})$. Nicho Realizado: Reducción del Nicho Existente debido a otros factores e interacciones de competencia.

El nicho fundamental se refiere a los límites en el espacio ambiental que soportaría la especie, sin que importe aclarar un tiempo fijo o un lugar en específico, y suponiendo la ausencia de otras especies, de tal forma que puede hacer uso completo de las condiciones ambientales y recursos. El nicho disponible esta conformado por las condiciones ($\in \mathcal{N}_F$) que están accesibles para determinado tiempo y región geográfica, y el realizado considera la interferencia de otras especies y otros factores.

La siguiente expresión resume la jerarquía de los 3 nichos:

$$\mathbf{N}_R(\mathbf{t}; \mathbf{G}) \subseteq \mathbf{N}_P(\mathbf{t}; \mathbf{G}) = \mathbf{N}_F \cap \mathbf{E}(\mathbf{t}; \mathbf{G}) \subseteq \mathbf{N}_F.$$

(Christen y Soberón, 2016)

Teóricamente, la estructura de los tres tipos de nichos son muy distintas. El nicho existente y el realizado pueden ser conjuntos complejos, no conexos o no convexos debido a los huecos que se forman por la inexistencia de ciertas combinaciones ambientales, sesgos en el muestreo, etc. Por otro lado, es razonable suponer que en el nicho fundamental estos huecos desaparecen, resultado una forma más sencilla o uniforme en algún sentido.

2.2. Datos e información a priori

Para aplicar la metodología propuesta a alguna especie de interés, son necesarios datos de presencia de la especie en un espacio ambiental de dos dimensiones, información del espacio disponible y conocimiento previo acerca del área que abarca el nicho fundamental de la especie en el espacio ambiental.

Para los ejemplos del capítulo 4 se utilizan los valores observados de temperatura media anual ($^{\circ}\text{C} \times 10$) y precipitación total anual (mm) de 3 especies vegetales, y los niveles máximos y mínimos de tolerancia que tiene la especie a cada variable. El espacio disponible consiste en las condiciones ambientales existentes en el continente de América (G).

Toda la información de una especie se resume como en la figura 2.1. A cada coordenada geográfica (longitud y latitud) le corresponde un punto en

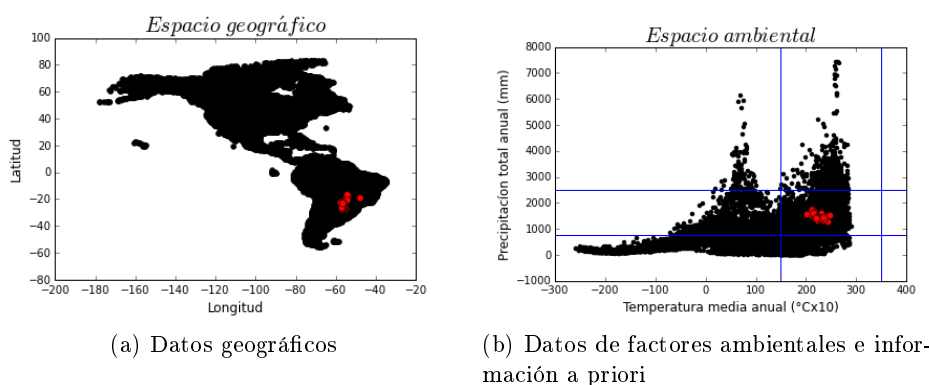


Figura 2.1: Ejemplo datos
Especie *Arachis*.

el espacio ambiental (temperatura y precipitación). Los puntos negros representan el espacio disponible, los rojos las ocurrencias de la especie y las líneas azules los rangos de tolerancia. Para conocer el espacio disponible se toma la información ambiental de una rejilla suficientemente fina que cubra América.

La base de datos para el espacio disponible se encuentra en <http://www.worldclim.org/current> (Hijmans et al., 2005). Consiste en una rejilla del clima global con resolución espacial de un kilómetro cuadrado. Para generar este tipo de datos se realiza un muestreo sobre algún área geográfica, dividiendo el área en cuadrantes de tamaños aproximadamente iguales y registrando algunas características del cuadrante, como por ejemplo, precipitación anual total, altitud, etc.

Las ocurrencias de la especie corresponden a colecciones de datos históricos. Al observar a la especie, se registran sus coordenadas geográficas para luego identificar los valores correspondientes de las variables climáticas en el espacio disponible. Ciertamente, estas observaciones pertenecen al nicho realizado. Es por esto que para obtener conclusiones acerca del nicho fundamental, es necesario incluir conocimiento previo de éste, de forma que compense la falta de información de la parte empírica.

Observando la densidad de puntos que representan el espacio disponible, se percibe que en la región de observación G hay más abundancia de ciertas condiciones ambientales que de otras. Más aún, se observan “huecos de información”, regiones ambientales amplias que no estuvieron accesibles para la especie pero que posiblemente pertenezcan al nicho fundamental de ésta. Visto de otra forma, las observaciones de la especie pertenecen al nicho realizado y no precisamente al nicho fundamental. Lo anterior puede ser tomado

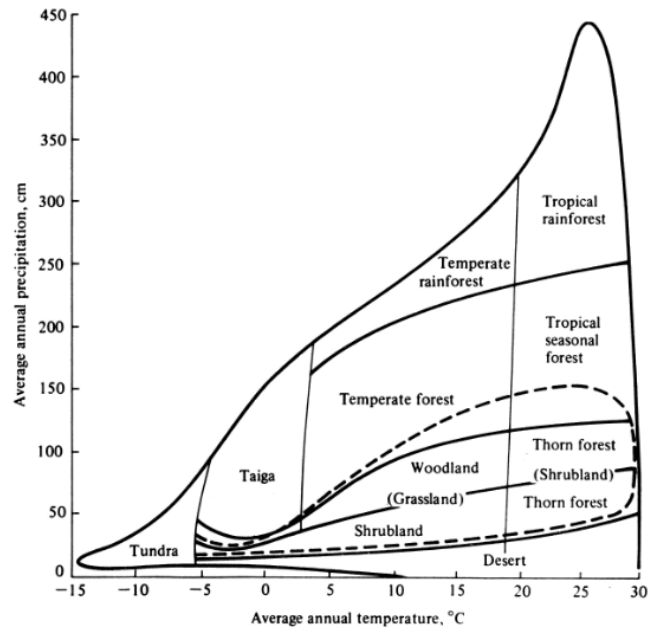


Figura 2.2: Biomas propuestos por Whittaker (1970)
 Imagen extraída de <http://www.zo.utexas.edu/courses/bio301/chapters/Chapter4/Chapter4.html>.

en cuenta en la modelación, como se revisa en el siguiente capítulo. Sería conveniente, en cambio, conocer la presencia de la especie en todo un conjunto ambiental “grande”, pero acotado, en donde sea razonable suponer que esta contenido el nicho de la especie. Comparando la forma de los datos con el conjunto que contiene a los Biomas de Whittaker (figuras 2.1b y 2.2) es posible hacerse una idea del tipo de espacio disponible deseable e identificar las zonas en donde la ausencia de datos es especialmente perjudicial.

Por otro lado, para ejemplificar la metodología se aprovecha información proveniente de invernaderos que consiste en los rangos (máximos y mínimos) de temperatura y precipitación que soporta la especie. Estos valores reflejan la respuesta de la especie a cada variable bajo el supuesto de que no fue afectada por ningún otro factor ambiental o de competencia y que tuvo acceso a, prácticamente, todo el espacio ambiental. Por esta razón, estos rangos se piensan como información acerca del área que abarca el nicho fundamental en el espacio ambiental. Es de esperarse, entonces, que el cuadrado a priori para algunas especies (azul en la figura 2.1), se perciba muy grande en comparación con los datos.

2.3. Inferencia bayesiana

Supóngase que se tiene una muestra aleatoria X_1, \dots, X_n para la cual se formula un modelo $\mathcal{P}(X|\theta)$, donde $\theta \in \mathbb{R}^p$ es el parámetro. Al hacerlo, se está suponiendo que el fenómeno puede ser descrito satisfactoriamente mediante dicho modelo paramétrico.

Bajo el enfoque bayesiano (el trabajo de Erdely y Gutiérrez consiste en un breve resumen del paradigma), la inferencia de un modelo paramétrico se basa en la función de probabilidad de θ condicionada a que se observó una muestra $D = (x_1, \dots, x_n)$. Así, es posible expresar en términos de probabilidades la incertidumbre que se tiene acerca del parámetro. Esta función recibe el nombre de distribución *posterior* (a los datos) de θ . Primero, se formula la distribución *a priori* de θ , procurando que refleje la información “previa” (sin datos) que se tiene acerca del parámetro. Para ello, se debe interpretar y traducir el conocimiento que un “experto”, del fenómeno aleatorio de interés, tiene acerca de θ . La distribución posterior toma en cuenta esta información y los datos (D). Esto se logra aplicando el Teorema de Bayes.

Se tiene que,

$$\mathcal{P}(\theta|D) = \frac{\mathcal{P}(D, \theta)}{\mathcal{P}(D)} = \frac{\mathcal{P}(D|\theta)\mathcal{P}(\theta)}{\mathcal{P}(D)} = \frac{\mathcal{P}(D|\theta)\mathcal{P}(\theta)}{\int_{\mathbb{R}^p} \mathcal{P}(D|\theta)\mathcal{P}(\theta)d\theta}. \quad (2.1)$$

En la expresión (2.1), $\mathcal{P}(\theta|D)$ es la función de densidad de probabilidad de θ dado que se observó D , $\mathcal{P}(\theta)$ la densidad de θ y $\mathcal{P}(D|\theta)$ la probabilidad de observar D dado θ . Estas funciones corresponden a las llamadas distribución posterior, distribución a priori y función de verosimilitud, de θ ; respectivamente. Las dos últimas se mezclan resultando en una función $\mathcal{P}(\cdot|D)$, $\Theta \subset \mathbb{R}^p$, que cumple ser una densidad para θ . En algunos casos, es conveniente trabajar con la función $U(\theta) = -\log(\mathcal{P}(\theta|D))$, llamada *función energía*.

Sin embargo, no siempre es posible obtener la expresión analítica de la distribución posterior, razón por la que se recurren a métodos numéricos para simular de ésta. Dadas simulaciones $\theta_1, \dots, \theta_M$ se pueden aproximar estimadores puntuales del parámetro. Más aún, al explotar dichas simulaciones, se consigue sintetizar información acerca de la variabilidad. Por tanto, se logra una comprensión del fenómeno más precisa.

2.3.1. Métodos Markov Chain Monte Carlo

Las dificultades computacionales que surgen en el enfoque bayesiano, generalmente se pueden resolver usando métodos de Monte Carlo basados en Cadenas de Markov (Gamerman y Lopes, 2006) – en inglés, Markov Chain

Monte Carlo (MCMC). Estos métodos sirven para simular de una distribución objetivo $f(x), x \in \mathcal{X}$.

La idea básica consiste en diseñar una cadena de Markov $(X_n)_{n \geq 1}$ que tenga como distribución límite $f(\cdot)$. Es decir, dadas las simulaciones x_1, \dots, x_T , se cumple que $x_i \sim f(\cdot)$, con $i \geq l$ para algún $l \in \{1, 2, \dots, T\}$ y T suficientemente grande. La cadena se inicializa en un punto x_0 , preferentemente simulándose de una distribución *inicial*, $\pi(\cdot) : \mathcal{X} \rightarrow (0, 1)$. Este punto inicial l influye en la rapidez de convergencia de la distribución de la cadena. Suponiendo que la cadena converge a la distribución objetivo en un tiempo $l \in \{1, 2, \dots, T\}$, los valores x_1, \dots, x_l no deben ser tomados como simulaciones de la variable. Además, al ser x_l, \dots, x_T una cadena de Markov, x_l y x_{l+1} están correlacionadas. Para formar un subconjunto pseudo independiente, se incluyen las simulaciones con un rezago τ adecuado: $x_{l+i\tau}$, con $i \in \{0, \dots, T\}$. De esta forma, con T iteraciones se obtiene una *muestra efectiva* de tamaño $N = (T - l)/\tau$. Se debe escoger τ tal que x_j y $x_{j+\tau}$ estén lo menor correlacionadas posible, para toda $j \in \{1, \dots, N - \tau\}$. En el contexto de inferencia bayesiana, $\mathcal{X} = \Theta$ y $f(\cdot)$ es la posterior de θ . Es decir, $f(\theta) = \mathcal{P}(\theta|D)$. En este caso, es recomendable tomar la distribución a priori de θ como distribución inicial.

Con el objetivo de formar cadenas de Markov con distribución límite $f(\cdot)$ eficientes, existen varios métodos numéricos. Dada una distribución objetivo, existen infinidad de cadenas de Markov que convergen a ésta. La dificultad radica en diseñar una cadena para la cual l y τ sean pequeños. Entre menor sean estos valores, se necesitarán menos iteraciones T para un mismo tamaño de muestra efectiva N .

Uno de los métodos más usados es algoritmo Metrópolis–Hastings (M–H). Para implementarlo, se propone una *distribución instrumental* $q(\cdot|x)$, de la cual se pueda simular para todo $x \in \mathcal{X}$. La eficiencia del algoritmo depende de la similitud que tenga esta función con la objetivo. Según la forma de la distribución objetivo, algunas propuestas suelen favorecer significativamente más a la rapidez del método que otras.

Una de las ventajas del algoritmo M–H es que, para simular de una distribución posterior $\mathcal{P}(\theta|D)$, no es necesaria la expresión analítica completa de esta función. En específico, no es necesario calcular la integral en (2.1).

2.4. Distribución normal–sesgada

El siguiente, es el lema clave para el desarrollo de una variedad de distribuciones con parámetro de sesgo. Para un estudio más profundo del tema consultar los trabajos de Azzalini (2005) y Azzalini y Valle (1996).

Lema 2.1. *Si f_0 es una función de densidad d -dimensional tal que $f_0(x) = f_0(-x)$ para $x \in \mathbb{R}^d$, F es una función de distribución 1-dimensional tal que F' es una densidad simétrica con respecto al 0, y w es una función real tal que $w(-x) = -w(x)$ para toda $x \in \mathbb{R}^d$, entonces*

$$f(z) = 2f_0(z)F(w(z)), \quad z \in \mathbb{R}^d$$

es una función de densidad en \mathbb{R}^d .

Usando la densidad Normal d -variada, la función de distribución normal univariada y $w(x) = \beta^T x$, con $\beta \in \mathbb{R}^d$, se obtiene la llamada distribución normal–sesgada.

2.4.1. \mathcal{SN} Univariada

Una variable aleatoria $Z \in \mathbb{R}$ se dice ser normal–sesgada con parámetro $\lambda \in \mathbb{R}$, $Z \sim \mathcal{SN}(\lambda)$, si su función de densidad es

$$f_Z(z; \lambda) := 2\phi(z)\Phi(\lambda z), \quad z \in \mathbb{R},$$

donde $\phi(x) = \frac{\exp(-x^2/2)}{\sqrt{2\pi}}$ y $\Phi(\lambda z) = \int_{-\infty}^{\lambda z} \phi(t)dt$.

El parámetro λ es de forma ya que regula el sesgo de la distribución. Cuando $\lambda = 0$, desaparece el sesgo y se obtiene la distribución normal estándar, $\mathcal{N}(0, 1)$. Conforme λ crece en valor absoluto, aumenta el sesgo de la distribución. Si $\lambda > 0$ el sesgo es positivo, y si $\lambda < 0$ es negativo. De esta forma, cuando $\lambda \rightarrow +\infty$, la distribución converge a la normal “doblada”, folded–normal en inglés. Además se cumple que si $Z \sim \mathcal{SN}(\lambda)$, entonces $-Z \sim \mathcal{SN}(-\lambda)$.

En la práctica, es necesario añadir parámetros de localización ($\mu \in \mathbb{R}$) y escala ($\sigma \in \mathbb{R}^+$). Esto se logra por medio de una transformación lineal: $X = \mu + \sigma Z$, y se escribe $X \sim \mathcal{SN}(\mu, \sigma^2, \lambda)$. Con eso, el caso $\lambda = 0$, corresponde a la distribución $\mathcal{N}(\mu, \sigma^2)$.

La esperanza y varianza de esta variable quedan dadas por las siguientes expresiones:

$$\mathbb{E}(X) = \mu + \sigma\sqrt{\frac{2}{\pi}}\delta, \quad (Var)(X) = \sigma^2\left(1 - \frac{2}{\pi}\delta^2\right), \quad (2.2)$$

donde la reparametrización está dada por las relaciones:

$$\delta = \frac{\lambda}{(1 + \lambda^2)^{1/2}} \in (-1, 1) \quad \text{y} \quad \lambda = \frac{\delta}{(1 - \delta^2)^{1/2}}.$$

Esta familia de distribuciones incluye a la normal como un caso especial, aunque en general es un conjunto de distribuciones con densidades sesgadas. Es frecuentemente usada para ajustar datos observados cuya distribución empírica tiene una forma similar a una Normal pero con cierta asimetría.

2.4.2. \mathcal{SN} Bivariado

La extensión al caso d -dimensional se construye de tal forma que las marginales resultan ser \mathcal{SN} univariadas. Fijando $d = 2$, se dice que una v.a. $Z = (Z_1, Z_2)$ se distribuye Normal-Sesgada, con parámetros $\omega \in (-1, 1)$ y $\beta \in \mathbb{R}^2$, si su densidad es

$$f_Z(z; \omega, \beta) := 2\phi_2(z; \omega, \beta)\Phi(\beta^T z), \quad z \in \mathbb{R}^2,$$

donde $\phi_2(z) = \frac{1}{2\pi|\Omega|^{1/2}} \exp(-\frac{1}{2}z^T\Omega^{-1}z)$, $\Omega = \begin{pmatrix} 1 & \omega \\ \omega & 1 \end{pmatrix}$

y $\Phi(\beta^T z) = \int_{-\infty}^{\beta^T z} \phi(t)dt$.

Para el caso bivariado, hay dos parámetros que regulan el sesgo, β_1 y β_2 . De manera análoga al caso univariado, $\beta = (0, 0)$ corresponde a la densidad normal bivariada, de marginales estándares y con correlación ω . Así mismo, al incrementar las componentes de β , el valor esperado de la distribución se aleja de la moda.

Los parámetros de localización ($\mu \in \mathbb{R}^2$) y escala ($\sigma_1, \sigma_2 \in \mathbb{R}^+$) se aumentan mediante las transformaciones: $X_i = \sigma_i Z_i + \mu_i$, $i = 1, 2$.

Es fácil ver (apéndice A) que la densidad de X es

$$f_X(x; \mu, A, \beta) = 2\psi(x; \mu, A)\Phi\left(\beta_1 \frac{x_1 - \mu_1}{\sigma_1} + \beta_2 \frac{x_2 - \mu_2}{\sigma_2}\right), \quad z \in \mathbb{R}^2, \quad (2.3)$$

donde $\psi(x; \mu, A)$ es la densidad normal bivariada con media $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ y matriz de precisión

$$A = \begin{pmatrix} \sigma_1^2 & \omega\sigma_1\sigma_2 \\ \omega\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2(1-\omega^2)} & -\frac{\omega}{\sigma_1\sigma_2(1-\omega^2)} \\ -\frac{\omega}{\sigma_1\sigma_2(1-\omega^2)} & \frac{1}{\sigma_2^2(1-\omega^2)} \end{pmatrix}.$$

Por otro lado, en ocasiones conviene regular el sesgo con otro parámetro en vez de β . La relación del parámetro en cuestión, $\delta \in (-1, 1)^2$, con β está dada por las siguientes expresiones:

$$\beta_1 = \frac{\delta_1 - \delta_2\omega}{[(1 - \omega^2)(1 - \omega^2 - \delta_1^2 - \delta_2^2 + 2\delta_1\delta_2\omega)]^{1/2}}, \quad (2.4)$$

$$\beta_2 = \frac{\delta_2 - \delta_1\omega}{[(1 - \omega^2)(1 - \omega^2 - \delta_1^2 - \delta_2^2 + 2\delta_1\delta_2\omega)]^{1/2}}. \quad (2.5)$$

Esta reparametrización, aunque no es uno–a–uno, ayuda a describir algunas características de la distribución. Por ejemplo, la construcción de la densidad de Z se hace bajo el supuesto de que $Z_1 \sim \mathcal{SN}(\lambda_1)$ y $Z_2 \sim \mathcal{SN}(\lambda_2)$, donde $\lambda_i = \delta_i/(1 - \delta_i^2)^{1/2}$, $\delta_i = \lambda_i/(1 + \lambda_i^2)^{1/2}$; $i = 1, 2$. Luego, si $X = \mu + SZ$, entonces $X_1 \sim \mathcal{SN}(\lambda_1, \mu_1, \sigma_1)$ y $X_2 \sim \mathcal{SN}(\lambda_2, \mu_2, \sigma_2)$. Y por (2.2) se tiene que,

$$\mathbb{E}(X_1) = \mu_1 + \sigma_1\sqrt{\frac{2}{\pi}}\delta_1, \quad \mathbb{E}(X_2) = \mu_2 + \sigma_2\sqrt{\frac{2}{\pi}}\delta_2. \quad (2.6)$$

El estudio de este tipo de distribuciones ofrece una extensa lista de alternativas para la modelación de datos. Lo anterior es una ventaja especialmente para el caso multivariado. Para el univariado no lo es tanto ya que, para este caso, las opciones de modelación ya son variadas.

Capítulo 3

Solución propuesta

Se propone medir la preferencia de una especie a las distintas zonas ambientales por medio de un modelo probabilístico. Se busca estimar una densidad para la cual las áreas más probables sean las más atractivas para la especie. Así, al dividir entre el valor más alto de esta función, se obtiene un “índice de preferencia” (score) entre cero y uno. Más aún, es posible interpretar la curva de algún nivel de esta función como una estimación del nicho fundamental, o al menos una cota inferior para éste. Además de la estimación puntual, se propone una herramienta gráfica que permite analizar la incertidumbre con respecto al índice y, por tanto, a la estructura del nicho.

Por la naturaleza de los datos, no necesariamente se observan más ocurrencias de la especie en los lugares de mayor preferencia para ésta. Por un lado, las observaciones de la especie están restringidas a cierto espacio disponible dados (t, G) . Por otro lado, la densidad de puntos en este espacio afecta la probabilidad de observar un espécimen. En otras palabras, no todas las regiones en el espacio ambiental son igualmente accesibles para la especie. Más aún, hay regiones que podrían pertenecer al nicho fundamental, pero que no están cubiertas por el espacio disponible en (t, G) . Un ejemplo de esto es el sesgo por muestreo, ya que el hecho de que haya zonas de difícil acceso (montañas, pantanosas, etc.) ocasiona que ciertos tipos de clima, muy específicos, no se puedan observar.

Con el objetivo de disminuir el efecto del problema anterior, primero se propone asignar ponderaciones a los datos de tal forma que “valgan por más” las ocurrencias en los conjuntos ambientales menos muestreados, permitiendo que la estimación del nicho abarque zonas con ausencia de datos. Así, con la nueva verosimilitud se obtiene una extrapolación del nicho realizado,

provocando que éste se acerque más al nicho fundamental.

Con el mismo objetivo, se sugiere adoptar un enfoque bayesiano. Las observaciones, sin ninguna modificación, proporcionan información del nicho realizado. En cambio, el conocimiento “previo” es información del nicho fundamental. Los rangos de tolerancia a las condiciones ambientales se interpretan como los extremos de intervalos de alta probabilidad de las marginales de la densidad asociada al nicho. Es decir, el cuadrado formado por estos rangos, no necesariamente queda cubierto por el nicho fundamental, aunque si representa una cota máxima. Para mezclar este conocimiento con la evidencia empírica, la inferencia de los parámetros del modelo se realiza bajo un enfoque bayesiano.

El desarrollo de la propuesta se presenta para dos dimensiones. Sin embargo, a excepción de la gráfica de incertidumbre, se puede generalizar a más dimensiones de manera análoga. Lo presentado por Azzalini y Valle (1996) acerca de la normal sesgada multivariada es suficiente para poder aplicar la metodología.

3.1. Supuestos

Durante el desarrollo de la metodología fue necesario hacer una serie de supuestos con respecto a la forma del nicho y la procedencia de los datos.

1. El N_F de cualquier especie es de forma convexa. El nicho fundamental se concibe como los límites del espacio ambiental en dónde puede crecer la especie. Entonces, si la especie soporta 2 combinaciones ambientales, es razonable suponer que también soporta combinaciones intermedias.
2. El único factor que reduce el nicho fundamental al realizado es la accesibilidad o disponibilidad al espacio ambiental. Es decir, el modelo no considera reducción del nicho existente por interacciones con otras especies u otros factores ambientales.
3. Las observaciones pertenecen a espacio disponible $\mathcal{E}(t; G)$. Suponiendo que la información de \mathcal{E} con la que se cuenta representa todas las combinaciones ambientales existentes que realmente estuvieron accesibles para la especie en el tiempo y región de observación, entonces las ocurrencias de las especies pertenecen a este espacio con probabilidad uno.

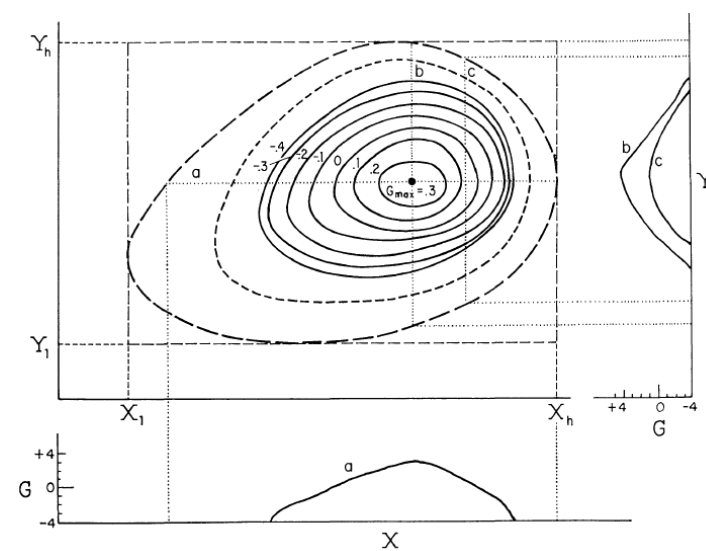


Figura 3.1: Estructura de un nicho como respuesta a las condiciones X y Y
 Imagen extraída de Maguire (1973).

3.2. Nichos no elipsoides

En un principio es razonable suponer que el nicho fundamental se puede representar por medio de una elipse, la cual cumple con ser convexa. Este supuesto, es tentador ya que se puede hacer uso del modelo normal bivariado (Christen y Soberón, 2016). Es común recurrir a este modelo en el caso multivariado debido a la poca variedad de modelos existentes para varias variables. Sin embargo, este supuesto es bastante fuerte para procesos ecológicos como el considerado en esta tesis. Algunos autores han supuesto estructuras bastantes complejas para describir la respuesta de una especie a los distintos factores ambientales. Maguire (1973), por ejemplo, se basa en el nicho de la figura 3.1 para analizar los procesos de competencia, crecimiento e interacción, entre otros, de una población.

Buscando más flexibilidad de la que ofrecen las elipses, se propone usar los contornos de la distribución normal sesgada como una aproximación a la estructura del nicho. Como se menciona en la sección 2.4, esta densidad cuenta con un parámetro que regula el sesgo. No obstante, para algunos valores de este parámetro, se consiguen densidades muy sesgadas, lo cual no es apropiado para representar los niveles de preferencia de una especie a los factores ambientales. Es decir, si la moda de la densidad corresponde a la combinación más favorable para la especie, se puede suponer que al alejarse de ésta, hacia cualquier dirección, no descende drásticamente el nivel de preferencia de la especie. Es por eso que, a priori, se restringen los valores

de este parámetro (sección 3.3).

Bajo el supuesto de convexidad, consideramos que la flexibilidad que aporta la densidad normal sesgada es suficiente para representar al nicho fundamental. Un nicho como el de la figura 3.1, por ejemplo, se puede obtener con esta densidad.

3.3. Inferencia

Sea $X \sim \mathcal{SN}(\mu, A, \beta)$ en \mathbb{R}^2 . La densidad de esta variable aleatoria está dada por la expresión:

$$f(x | \mu, \Sigma, \beta) \tag{3.1}$$

$$= \frac{1}{(2\pi) |\Sigma|^{-1/2}} \exp \left[-\frac{1}{2} (x_i - \mu)^T \Sigma^{-1} (x_i - \mu) \right] \Phi \left(\frac{x_{i1} - \mu_1}{\sigma_1} \beta_1 + \frac{x_{i2} - \mu_2}{\sigma_2} \beta_2 \right), \tag{3.2}$$

$$\text{donde } \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad A = \begin{pmatrix} \sigma_1^2 & \omega \sigma_1 \sigma_2 \\ \omega \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}^{-1} \text{ y } \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}.$$

Suponiendo que esta distribución, con los parámetros adecuados, describe la preferencia de alguna especie a las diferentes combinaciones de condiciones ambientales, entonces el valor $f(x | \mu, A, \beta)/q$, donde $q = \max_{x \in \mathcal{A}} f(x | \mu, A, \beta)$, es un “score” entre cero y uno que indica qué tan benéfica es la combinación $x \in \mathcal{A}$ para la especie. Así, la moda tiene un score igual a uno y corresponde a la combinación idónea para la especie. Se puede suponer que una estimación al nicho fundamental de la especie consiste en el conjunto de combinaciones cuyos scores son mayores a cierto valor pequeño (cercano a cero). El conjunto de puntos que cumplen con tener un score cercano a uno, forman una región muy angosta de muy alta preferencia para la especie. Este conjunto no sería adecuado para el nicho fundamental ya que éste consiste en todo lo que la especie puede soportar.

El enfoque que se propone para hacer inferencia de los parámetros de esta distribución es el bayesiano. Como se menciona en la sección 2.3, para mezclar la información proporcionada por los datos de presencia de la especie y el conocimiento previo de los rangos de tolerancia es necesario obtener la expresión analítica de la verosimilitud y establecer una distribución a priori adecuada para los parámetros los parámetros.

3.3.1. Verosimilitud

Sean $x_1, \dots, x_n \in \mathcal{A} \subset \mathbb{R}^2$ las observaciones (independientes) de la especie en el espacio ambiental. Para obtener la verosimilitud, se debe tomar en

cuenta que las ocurrencias provienen del conjunto discreto $\mathcal{E}(t; G)$ (Supuesto 3). Entonces, la aportación de cada punto a la verosimilitud es proporcional a la probabilidad de observarlo:

$$\mathbb{P}(x_i | \mu, \Sigma, \beta) = \frac{f(x_i | \mu, A, \beta)}{\sum_{y \in \mathcal{E}} f(y | \mu, A, \beta)}.$$

Con esto,

$$\mathcal{L}(\mu, A, \beta | D) \propto \prod_{i=1}^n \mathbb{P}(x_i | \mu, A, \beta).$$

Esta forma de proceder proporcionaría una estimación del nicho fundamental si el espacio disponible cubriera uniformemente toda la zona ambiental en donde se encuentra el nicho. Sin embargo, para el tiempo y la región de observación (t y G), hay áreas del espacio ambiental más abundantes que otras. Esta abundancia se advierte como si se hubiera realizado un muestreo no uniforme sobre el espacio ambiental, afectando la probabilidad de registrar una ocurrencia. Para corregir este “error de muestreo”, se propone ponderar $\mathbb{P}(x_i | \mu, \Sigma, \beta)$ según la densidad de puntos de la zona ambiental donde se encuentra la observación x_i , dándole menor peso a las combinaciones ambientales que se encuentran en conjuntos ambientales más abundantes, y viceversa. Para esto, es necesario encontrar una estimación de $\mathcal{E}(t; G)$ lo suficientemente flexible de tal forma que las regiones muy abundantes sean una moda de la estimación. Por esta razón, se hace una estimación por kernels (no paramétrica) de la densidad del espacio disponible. Otra razón para escoger un enfoque no paramétrico es que no se tiene conocimiento previo acerca de la forma de $\mathcal{E}(t; G)$. En contraste, para la estimación del N_f se escoge un modelo paramétrico (normal sesgada) después de hacer una serie de supuestos acerca de su forma.

Según el método por kernels, la estimación para cada punto $x = (x_1, x_2) \in \mathcal{A}$ es,

$$k_h(x) = \frac{1}{N} \sum_{i=1}^N \frac{1}{h^2} K\left(\frac{x_1 - y_{i1}}{h}\right) K\left(\frac{x_2 - y_{i2}}{h}\right),$$

donde N es la cardinalidad de $\mathcal{E}(t; G)$, $K(x)$ una función que integra uno y h el ancho de banda.

Así, para cada punto $x \in \mathcal{A}$, $k_h(x)$ es un indicador de la cantidad de puntos “ceranos” a x . Buscando que los pesos para cada observación $x_i \in \mathcal{E}(t; G)$ sean decrecientes con respecto a la abundancia, Christen y Soberón

(2016) proponen las siguientes ponderaciones:

$$w_i = w(x_i) = \frac{\left(\frac{1}{k_h(y)}\right)^a}{\frac{1}{n} \sum_{z \in \mathcal{E}} \left(\frac{1}{k_h(z)}\right)^a}; \quad w_i > 0, a \geq 0,$$

cumpléndose que $\sum_i^n w_i = n$.

Con esto, la aportación de cada observación a la verosimilitud consiste en la probabilidad reponderada:

$$\mathbb{P}^*(x_i | \mu, \Sigma, \beta) = \left(\frac{f(x_i | \mu, \Sigma, \beta)}{\sum_{y \in \mathcal{E}} f(y | \mu, \Sigma, \beta)} \right)^{w_{x_i}}.$$

Así,

$$\mathcal{P}(D | \mu, A, \beta) = \mathcal{L}(\mu, A, \beta | D) = \prod_{i=1}^n \mathbb{P}^*(x_i | \mu, \Sigma, \beta) = \prod_{i=1}^n \left(\frac{f(x_i | \mu, \Sigma, \beta)}{\sum_{y \in \mathcal{E}} f(y | \mu, \Sigma, \beta)} \right)^{w_{x_i}} \quad (3.3)$$

$$\propto \prod_{i=1}^n \left(\frac{\left(\exp \left[-\frac{1}{2} (x_i - \mu)' A (x_i - \mu) \right] \Phi \left[\frac{x_{i1} - \mu_1}{\sigma_1} \beta_1 + \frac{x_{i2} - \mu_2}{\sigma_2} \beta_2 \right] \right)}{\sum_{y \in \mathcal{E}(t)} \left(\exp \left[-\frac{1}{2} (y - \mu)' A (y - \mu) \right] \Phi \left[\frac{y_1 - \mu_1}{\sigma_1} \beta_1 + \frac{y_2 - \mu_2}{\sigma_2} \beta_2 \right] \right)} \right)^{w_{x_i}}. \quad (3.4)$$

El parámetro a permite controlar qué tanta más importancia tienen las observaciones de zonas menos densas sobre el resto de las observaciones. Tomando $a = 0$ todas las observaciones tienen la misma ponderación. Conforme se eligen valores de a más grandes, la moda de la estimación del nicho se acerca más a las observaciones en zonas poco densas. Supóngase que x_i está en una zona más abundante que x_j , entonces $k_h(x_i) > k_h(x_j)$, $h \in \mathbb{R}$. Luego, $(1/k_h(x_i))^a \leq (1/k_h(x_j))^a$, para todo $a > 0$, lo que implica que $w_i \leq w_j$. La proporción $w_i/w_j = (k_h(x_j)/k_h(x_i))^a \leq 1$ tiende a cero conforme a crece y es uno cuando $a = 0$. Luego, con el parámetro a se cambia la proporción o el contraste en las ponderaciones de las observaciones. En la figura 3.2 se muestra este contraste para distintos valores de a y ancho de banda $h = (25, 150)$, los puntos más claros son los que se encuentran en zonas menos densas y viceversa.

Para ejemplificar la metodología en el capítulo 4 se fija $a = 0.4$. A falta de mejor criterio, la elección de este valor simplemente se basa en la búsqueda de un equilibrio entre nichos centrados en observaciones poco abundantes

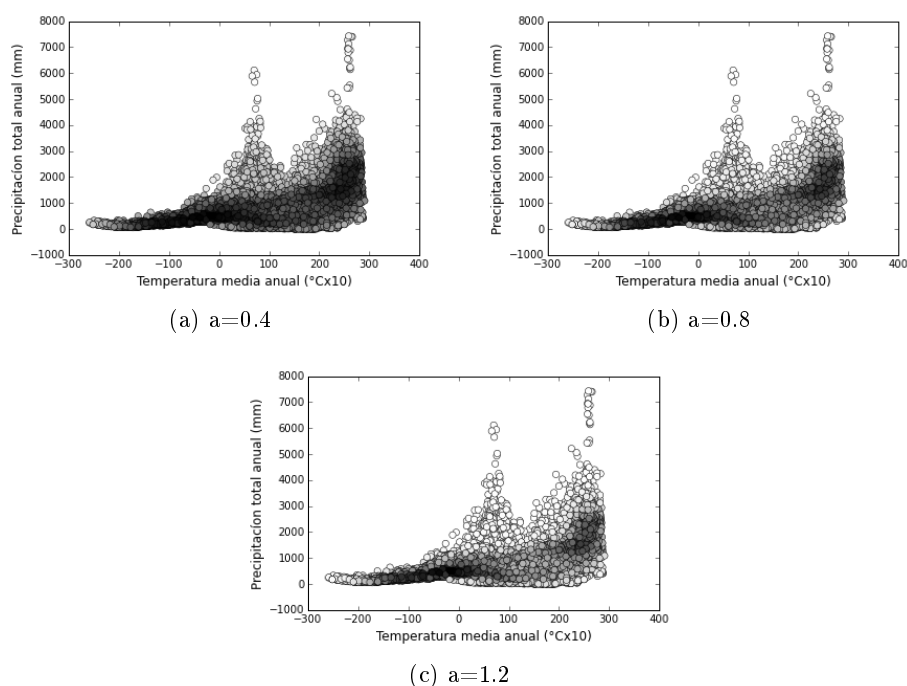


Figura 3.2: Efecto en las ponderaciones para distintos valores de a . Los puntos del espacio disponible se colorean en escala de grises según la ponderación. En negro el punto con menor ponderación y en blanco el punto de mayor.

y nichos que abarquen todos los datos y se centren conforme a éstos. Ciertamente, sigue en discusión cómo elegir este valor de manera adecuada. Al final del capítulo 4 se comparan los resultados con otros valores. Por otro lado, sin entrar en muchos detalles con respecto a la estimación no paramétrica de la densidad en $\mathcal{E}(t; G)$, se utilizan kernels normales y se estima el ancho de banda usando la “Rule of Thumb”. El resultado, según esta regla es $h = (25, 150)$. Otra forma de controlar el efecto de la densidad sobre las probabilidades sería mediante el ancho de banda h . Sin embargo, se decide fijar h y controlar el impacto en la ponderación mediante el parámetro a .

3.3.2. Distribución a priori

Se tiene que $X \in \mathbb{R}^2$ sigue una distribución normal-sesgada. Los parámetros que aparecen en la expresión (3.1) de la densidad de esta variable son (μ, A, β) . Sin embargo, para asignar una distribución a priori a los parámetros, será conveniente usar otra parametrización para el sesgo (δ en vez de β usando la ecuación 2.4). La asignación a priori de la distribución para

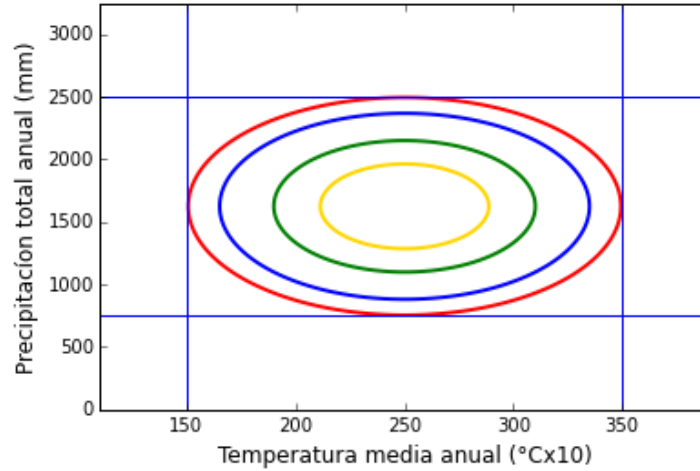


Figura 3.3: Valor esperado a priori de (μ, A, δ)
 Contornos de la densidad $f(x|\mu, A, \beta)$ para scores 0.15, 0.25, 0.5 y 0.75
 fijando μ , A y δ en sus valores esperados a priori.

(μ, A, δ) se realiza a partir los rangos de tolerancias a priori, de donde se puede extraer información de la media de X y de la varianza de las marginales, más no de la correlación ni del sesgo.

El valor esperado a priori que finalmente se asigna a los parámetros para ejemplificar la metodología en el capítulo 4 da como resultado la densidad de la figura 3.3. Cualquiera de las 2 curvas de nivel más grandes son buenas candidatas para ser consideradas como las estimaciones a priori del nicho fundamental.

3.3.2.1. Media del modelo sesgado

Para asignar una distribución a priori al parámetro μ , es conveniente primero establecer una distribución a priori para el valor esperado de X y luego considerar la relación entre μ y $\mathbb{E}(X)$ para encontrar la distribución de μ resultante. Se supondrá que, a priori, la media de X sigue una distribución normal centrada en el cuadrado a priori y con marginales independientes. Es decir, $\mathbb{E}(X) \sim \mathcal{N}(\nu, P)$, donde las coordenadas de ν son los promedios de los respectivos rangos de tolerancia y P es una matriz 2x2 diagonal. O bien, $\mathbb{E}(X_1) \sim \mathcal{N}(\nu_1, \rho_1^2)$ y $\mathbb{E}(X_2) \sim \mathcal{N}(\nu_2, \rho_2^2)$, donde las varianzas se asignan de acuerdo a la incertidumbre que se tenga con respecto a la media de X .

Para los ejemplos del capítulo 4, se fijó ρ_1^2 de manera que $\mathbb{E}(X_1)$ se encuentre entre los rangos a priori de precipitación con 99.99 % de probabilidad

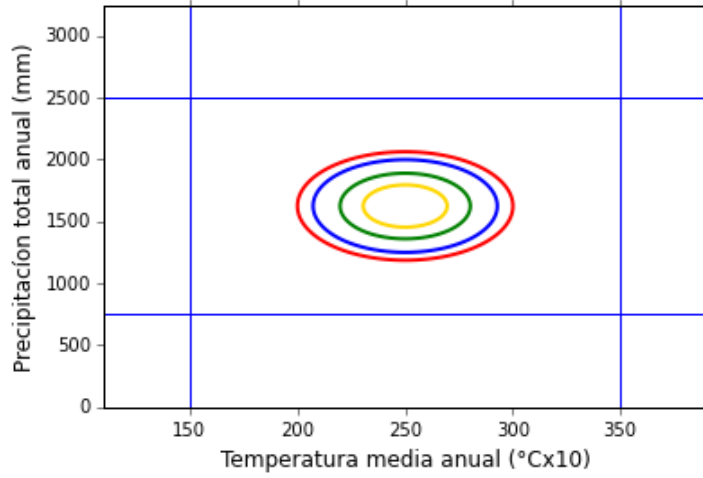


Figura 3.4: Distribución a priori para el valor esperado de $X \in \mathbb{R}^2$
Contornos de probabilidad 0.85, 0.75, 0.5 y 0.25 de $\mathcal{P}_{\mathbb{E}(X)}$.

y análogamente para ρ_2^2 . Aunque pareciera ser una probabilidad bastante alta, hay que recordar que se trata de la esperanza de X y no de X tal cual. En la figura 3.4 se muestran los contornos de la densidad normal resultante.

Ya establecidas las distribuciones para $\mathbb{E}(X_1)$ y $\mathbb{E}(X_2)$, se consideran las ecuaciones 2.6:

$$\mu_1 = \mathbb{E}(X_1) - \sigma_1 \sqrt{\frac{2}{\pi}} \delta_1, \quad \mu_2 = \mathbb{E}(X_2) - \sigma_2 \sqrt{\frac{2}{\pi}} \delta_2,$$

donde $\delta = \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}$ satisface las ecuaciones 2.4.

Luego, se cumple que

$$\begin{aligned} \mu_1 \mid \sigma_1, \delta_1 &\sim \mathcal{N}\left(\nu_1 - \sigma_1 \sqrt{\frac{2}{\pi}} \delta_1, \rho_1^2\right), \\ \mu_2 \mid \sigma_2, \delta_2 &\sim \mathcal{N}\left(\nu_2 - \sigma_2 \sqrt{\frac{2}{\pi}} \delta_2, \rho_2^2\right). \end{aligned}$$

O bien,

$$\mathcal{P}_{\mu|A,\delta}(\mu \mid A, \delta) = \frac{1}{2\pi \mid \Sigma_0 \mid^{-1/2}} \exp\left[-\frac{1}{2}(\mu - \mu_0)^T \Sigma_0^{-1}(\mu - \mu_0)\right], \quad (3.5)$$

donde $\mu_0 = \begin{pmatrix} \nu_1 - \sigma_1 \sqrt{\frac{2}{\pi}} \delta_1 \\ \nu_2 - \sigma_2 \sqrt{\frac{2}{\pi}} \delta_2 \end{pmatrix}$ y $\Sigma_0 = \begin{pmatrix} \rho_1 & 0 \\ 0 & \rho_2 \end{pmatrix}$.

Dado que la relación entre β y δ no es uno-a-uno, no es posible expresar lo anterior en términos de β . Esta es la razón por la cual se trabaja con la parametrización (μ, A, δ) .

3.3.2.2. Sesgo

Aunque el parámetro de sesgo con el que se decidió trabajar es $\delta \in (-1, 1)^2$, es más fácil entender el comportamiento del sesgo usando β . Conforme $\beta_1 \rightarrow \pm\infty$, aumenta el sesgo hacia la derecha o izquierda mientras que cuando $\beta_2 \rightarrow \pm\infty$ el sesgo hacia arriba o abajo aumenta (figura 3.5). Así mismo, es fácil entender el cambio en la forma de la distribución al cambiar β_1, β_2 simultáneamente, ya que β_1 siempre influye sobre el eje de las abscisas sin importar el valor de β_2 , y análogamente para β_2 . Sin embargo, esto no ocurre con el parámetro δ , haciendo la relación entre δ_1, δ_2 y el sesgo más complicada. Ciertamente, es posible asignar como distribución a priori para δ la uniforme en $(-1, 1)^2$, haciéndolo no informativo. No obstante, como se menciona en la sección 3.2, se desea que el sesgo esté restringido a valores pequeños. Esto, se puede lograr fácilmente asignando a β_1 y β_2 una distribución normal centrada en cero y con poca varianza, para después obtener, por medio de un cambio de variable, la distribución a priori de δ . En otras palabras,

$$\beta \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} s_1 & 0 \\ 0 & s_2 \end{pmatrix}\right),$$

donde s_1 y s_2 son lo suficientemente pequeños. Para hacerse una idea de las varianzas adecuadas, se explora visualmente un poco los contornos de la densidad normal-sesgada para distintos valores de β (figura 3.5). Se observa que para valores de β_i mayor o igual a 3, $i = 1, 2$, se obtienen estructuras poco apropiadas para el nicho fundamental de una especie. Así, para los ejemplos del capítulo 4, se elige s_i de tal forma que $\beta_i \in (-2, 2)$ con 90 % de probabilidad, $i = 1, 2$.

Una vez fija la distribución a priori de β , se realiza un cambio de variable utilizando la relación 2.4. Aunque el cambio de variable no es uno-a-uno, en las ecuaciones se tiene el despeje necesario para sustituir β en términos de δ :

$$\mathcal{P}_\beta(\beta) = \frac{1}{2\pi|U_0|^{-1/2}} \exp\left(-\frac{1}{2}\beta^T U_0^{-1}\beta\right), \text{ donde } U_0 = \begin{pmatrix} s_1 & 0 \\ 0 & s_2 \end{pmatrix}.$$

Entonces, aplicando el Teorema de Cambio de Variable

$$\mathcal{P}_\delta(\delta) = \mathcal{P}_\beta(\beta(\delta)) |J| = \frac{1}{2\pi|U_0|^{-1/2}} \exp\left(-\frac{1}{2}\beta(\delta)^T U_0^{-1}\beta(\delta)\right) |J|, \quad (3.6)$$

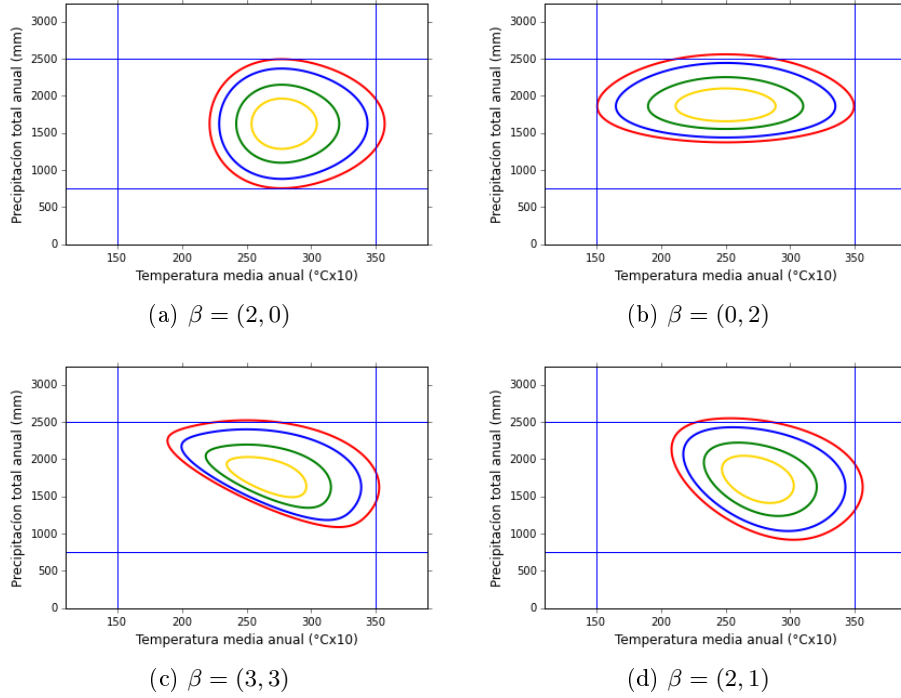


Figura 3.5: Sesgo para distintos valores de β
 Contornos de $\mathcal{SN}(\mu, A, \beta)$ con scores de 0.25, 0.5 y 0.75, fijando μ y A en sus valores esperados a priori (figura 3.3) y variando β .

donde

$$\beta(\delta) = \begin{pmatrix} \beta_1(\delta) \\ \beta_2(\delta) \end{pmatrix} = \begin{pmatrix} \frac{\delta_1 - \delta_2 \omega}{r^{1/2}} \\ \frac{\delta_2 - \delta_1 \omega}{r^{1/2}} \end{pmatrix}, \quad (3.7)$$

$$r = (1 - \omega^2)(1 - \omega^2 - \delta_1^2 - \delta_2^2 + 2\delta_1 \delta_2 \omega), \quad (3.8)$$

y $|J|$ es el jacobiano de la transformación. Después de un poco de álgebra (Apéndice B) se tiene que

$$|J| = r^{-1}(1 - \omega^2) - (1 + \omega)(1 - \omega^2)[(\delta_1 - \delta_2 \omega)^2 + (\delta_2 - \delta_1 \omega)^2].$$

3.3.2.3. Matriz de precisión del modelo normal sin sesgo

Para la matriz A se supondrá, a priori, que $A \sim \text{Wishart}(\alpha, W)$, con W una matriz 2×2 y $\alpha > 1$.

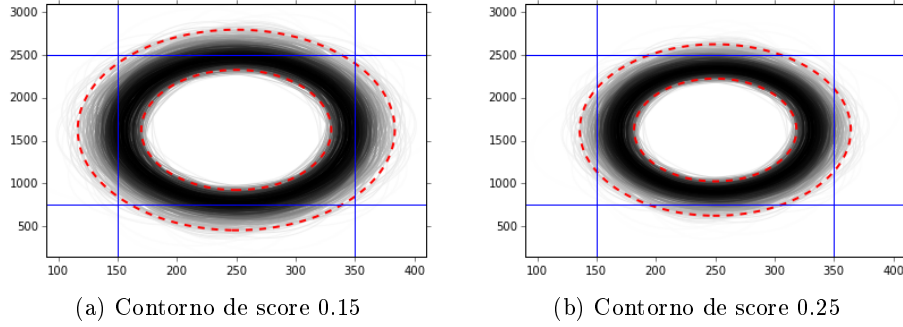


Figura 3.6: Variabilidad a priori de la matriz de precisión A . Contornos de $\mathcal{SN}(\mu, A, \beta)$ para scores de 0.15 y 0.25 asociados la densidad $\mathcal{SN}(\mu, A, \beta)$ fijando μ y β en sus valores esperados a priori y simulando $A \sim \text{Wishart}(W, \alpha = 30)$. Se utiliza una escala de grises donde negro representa la simulación con mayor probabilidad (\mathcal{P}_A) y blanco la de menor probabilidad. En rojo punteado contornos de del mismo score, con valores propios aproximadamente iguales a los cuantiles empíricos de los valores propios simulados (figura 3.7).

Entonces, se obtiene

$$\mathcal{P}_A(A) = \frac{|A|^{(\alpha-1)/2} \exp[-\frac{1}{2}\text{tr}(AW^{-1})]}{2^\alpha |W|^{\alpha/2} \pi^{1/2} \prod_{i=1}^d \Gamma(\frac{\alpha+1-i}{2})}. \quad (3.9)$$

Esta distribución cumple que $\mathbb{E}(A) = \alpha W$. Así, al fijar un valor esperado $A_0 \in \mathbb{R}^{2 \times 2}$, se tiene que $W = A_0/\alpha$. Fijando W , la incertidumbre de A depende del valor de α . El caso $\alpha = 2$ corresponde a la distribución propia menos informativa (más incertidumbre) y conforme aumenta α disminuye la variabilidad (Gelman et al., 2014, p. 582).

Dados los valores esperados del sesgo y la media, la estimación a priori del nicho corresponde a los contornos de una densidad normal no sesgada ($\beta_0 = (0, 0)^T$), centrada en el cuadrado a priori ($\mu = \mu_0$) y con matriz de precisión $A_0 = \alpha W$ (figura 3.3). En el desarrollo del capítulo 4 se fijó $\mathbb{E}(A) = A_0$ de forma que las marginales de la normal se encuentren entre el rango de tolerancia respectivo con 95 % de probabilidad. Por otro lado, para escoger un valor de α adecuado, se realizaron simulaciones A_1, \dots, A_M de A para varios valores de α y se analizó el comportamiento empírico. Dado un valor para α , se pueden visualizar los contornos asociados a la densidad $f(x | \mu_0, \beta_0, A_i)$, $i = 1, 2, \dots, M$, como en la figura 3.6.

La figura 3.6 corresponde al valor $\alpha = 30$, valor que se utiliza en los ejemplos. Al fijar los contornos máximos y mínimos (rojos punteados en

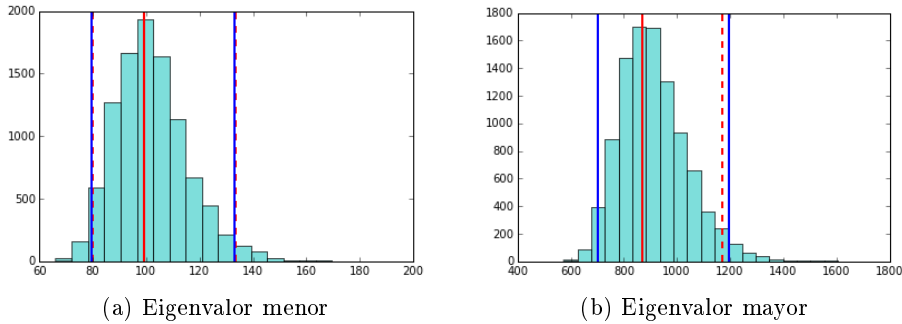


Figura 3.7: Histograma de eigenvalores para las elipses (score 0.15) simuladas Eigenvalores para $M=10000$ simulaciones de $A \sim \text{Wishart}(W, \alpha = 30)$. Las líneas azules corresponden a los cuantiles que abarcan el 95 % de probabilidad y las líneas rojas punteadas los valores propios de las elipses máximas y mínimas de la figura 3.6. En la gráfica de la derecha coinciden los mínimos.

la figura), el valor $\alpha = 30$ cumple con que los contornos simulados están contenidos entre éstos 2, con alta probabilidad. Para ver esto, primero se calculan los eigenvalores de las elipses correspondientes a cada simulación (figura 3.7), recordando que éstos están asociados al tamaño de cada elipse. Luego, se calculan los cuantiles empíricos del 2.5 % y 97.5 % de probabilidad de estos eigenvalores. Finalmente, se verifica que coincidan aproximadamente con los eigenvalores de los contornos máximos y mínimos fijados previamente (rojos punteados en la figura 3.6).

Como se menciona anteriormente, se busca que la estructura del nicho fundamental estimada no abarque un área que sólo contenga a los datos, si no un área más grande. Por esta razón se restringir a que el tamaño del nicho no sea muy distinta a la del cuadrado a priori (α grande).

3.3.3. Función Energía

La función energía es igual a menos el logaritmo de la distribución posterior de los parámetros dado los datos:

$$U(\mu, A, \delta) = -\log(\mathcal{P}(\mu, A, \delta|D)) + \text{cte.} = -\log(\mathcal{P}(D|\mu, A, \delta)) - \log(\mathcal{P}_{\mu, A, \delta}(\mu, A, \delta)).$$

Usando (3.3) y recordando la reparametrización dada por (3.7) se tiene que

$$\begin{aligned}
& -\log(\mathcal{L}(\mu, A, \delta|D)) \\
& \propto -\log \left(\prod_{i=1}^n \left(\frac{\left(\exp \left[-\frac{1}{2}(x_i - \mu)' A(x_i - \mu) \right] \Phi \left[\frac{x_{i1} - \mu_1}{\sigma_1} \beta_1(\delta) + \frac{x_{i2} - \mu_2}{\sigma_2} \beta_2(\delta) \right] \right)}{\sum_{y \in \mathcal{E}(t)} \left(\exp \left[-\frac{1}{2}(y - \mu)' A(y - \mu) \right] \Phi \left[\frac{y_1 - \mu_1}{\sigma_1} \beta_1(\delta) + \frac{y_2 - \mu_2}{\sigma_2} \beta_2(\delta) \right] \right)} \right)^{w_{x_i}} \right) \\
& = -\sum_{i=1}^n w_{x_i} \log \left(\exp \left[-\frac{1}{2}(x_i - \mu)' A(x_i - \mu) \right] \Phi \left[\frac{x_{i1} - \mu_1}{\sigma_1} \beta_1(\delta) + \frac{x_{i2} - \mu_2}{\sigma_2} \beta_2(\delta) \right] \right) \\
& + n \log \left(\sum_{y \in \mathcal{E}(t)} \exp \left[-\frac{1}{2}(y - \mu)' A(y - \mu) \right] \Phi \left[\frac{y_1 - \mu_1}{\sigma_1} \beta_1(\delta) + \frac{y_2 - \mu_2}{\sigma_2} \beta_2(\delta) \right] \right) \\
& = \sum_{i=1}^n w_{x_i} \left(\frac{1}{2}(x_i - \mu)' A(x_i - \mu) - \log \left(\Phi \left[\frac{x_{i1} - \mu_1}{\sigma_1} \beta_1(\delta) + \frac{x_{i2} - \mu_2}{\sigma_2} \beta_2(\delta) \right] \right) \right) \\
& + n \log \left(\sum_{y \in \mathcal{E}(t)} \exp \left[-\frac{1}{2}(y - \mu)' A(y - \mu) \right] \Phi \left[\frac{y_1 - \mu_1}{\sigma_1} \beta_1(\delta) + \frac{y_2 - \mu_2}{\sigma_2} \beta_2(\delta) \right] \right).
\end{aligned}$$

Y de (3.5),(3.6) y (3.9):

$$\begin{aligned}
\mathcal{P}_{\mu, A, \delta}(\mu, A, \beta) & = \left(\frac{1}{2\pi |\Sigma_0|^{-1/2}} \exp \left[-\frac{1}{2}(\mu - \mu_0)' \Sigma_0^{-1}(\mu - \mu_0) \right] \right) \\
& \times \left(\frac{|A|^{(\alpha-1)/2} \exp \left[-\frac{1}{2} \text{tr}(AW^{-1}) \right]}{2^\alpha |W|^{\alpha/2} \pi^{1/2} \prod_{i=1}^d \Gamma \left(\frac{\alpha+1-i}{2} \right)} \right) \\
& \times \left(\frac{1}{2\pi |U_0|^{-1/2}} \exp \left(-\frac{1}{2} \beta(\delta)' U_0^{-1} \beta(\delta) \right) |J| \right).
\end{aligned}$$

Quitando los términos que no dependen de μ , Σ o β , se tiene la siguiente expresión para la función energía:

$$\begin{aligned}
U(\mu, A, \delta) & = \sum_{i=1}^n w_{x_i} \left(\frac{1}{2}(x_i - \mu)' A(x_i - \mu) - \log \left(\Phi \left[\frac{x_{i1} - \mu_1}{\sigma_1} \beta_1(\delta) + \frac{x_{i2} - \mu_2}{\sigma_2} \beta_2(\delta) \right] \right) \right) \\
& + n \log \left(\sum_{y \in \mathcal{E}(t)} \exp \left[-\frac{1}{2}(y - \mu)' A(y - \mu) \right] \Phi \left[\frac{y_1 - \mu_1}{\sigma_1} \beta_1(\delta) + \frac{y_2 - \mu_2}{\sigma_2} \beta_2(\delta) \right] \right) \\
& + \frac{1}{2}(\mu - \mu_0)' \Sigma_0^{-1}(\mu - \mu_0) + \frac{\alpha-1}{2} \log(|A|) + \frac{1}{2} \text{tr}(AW^{-1}) \\
& + \frac{1}{2} \beta(\delta)' U_0^{-1} \beta(\delta) - \log(|J|).
\end{aligned}$$

3.4. Gráfica de incertidumbre

Una vez calculada la función energía, se aplica un método MCMC y se obtienen simulaciones de la distribución posterior $\mathcal{P}(\mu, A, \delta|D)$. Un estimador puntual para los parámetros que se obtiene fácilmente a partir de las simulaciones es la moda de la posterior, o mejor conocido como el MAP (valor de máxima posterior). No obstante, quedarse únicamente con este valor e ignorar las demás simulaciones no es lo más recomendable. En las simulaciones se encuentra una gran cantidad de información acerca del parámetro, sólo hay que saber extraerla.

En esta sección se explica una propuesta para aprovechar todas las simulaciones del parámetro obtenidas. Esta propuesta consiste en una representación gráfica de la variabilidad del parámetro en la estructura del nicho estimada. Con esto se logra un análisis más completo y profundo de la preferencia de la especie que se estudia a las combinaciones ambientales.

Supóngase que se tienen T simulaciones de los parámetros (μ_i, A_i, δ_i) , $i = 1, \dots, T$. Cada una se sustituye en la densidad normal-sesgada y se divide esta densidad entre su moda $f(x|\mu_i, A_i, \delta_i)/\max_{x \in \mathcal{A}} f(x | \mu_i, A_i, \delta_i)$. Así, es posible conocer los contornos para ciertos scores entre cero y uno (figura 3.3). Pretendiendo entender un poco la variabilidad, se podrían juntar las gráficas de los contornos correspondientes a todas las simulaciones (un score por gráfica), pero el resultado sería difícil de interpretar. Por ello, se procede de otra manera. Imagine, lector, que para cada simulación se rellenan 3 contornos en rojo, amarillo y blanco con intensidad decreciente con respecto al número de simulaciones, como en la figura 3.8. Ahora, imagínese que las gráficas para cada simulación se enciman de forma que las intensidades parecen “sumarse” y los colores se mezclan. Cuando la variabilidad es nula (simulaciones iguales) se obtiene una gráfica como la figura 3.8a. Por el contrario, un caso de alta variabilidad se aprecia en la figura 3.9. Considerando los scores 0.25, 0.5 y 0.75, un pixel es completamente blanco si el score en ese punto fue mayor a 0.75 para todas las simulaciones; es amarillo, con la intensidad de la figura 3.8a, si el score se encuentra entre 0.5 y 0.75 para todas las simulaciones; y análogamente para el color rojo.

Para lograr el efecto, se aprovechan las características de la codificación RGB de los colores. El rojo, amarillo y blanco tienen codificación (255,0,0), (255,255,0) y (255,255,255) respectivamente, donde el valor 255 ofrece la mayor intensidad de color. También se puede trabajar, como lo hacen algunos softwares, con la codificación estandarizada: (1,0,0), (1,1,0) y (1,1,1). La tercia (c, c, c) , $c \in (0, 1)$, ofrece los colores grises obteniéndose blanco cuando $c = 1$ y negro cuando $c = 0$. La combinación $(c, 0, 0)$ corresponde siempre al

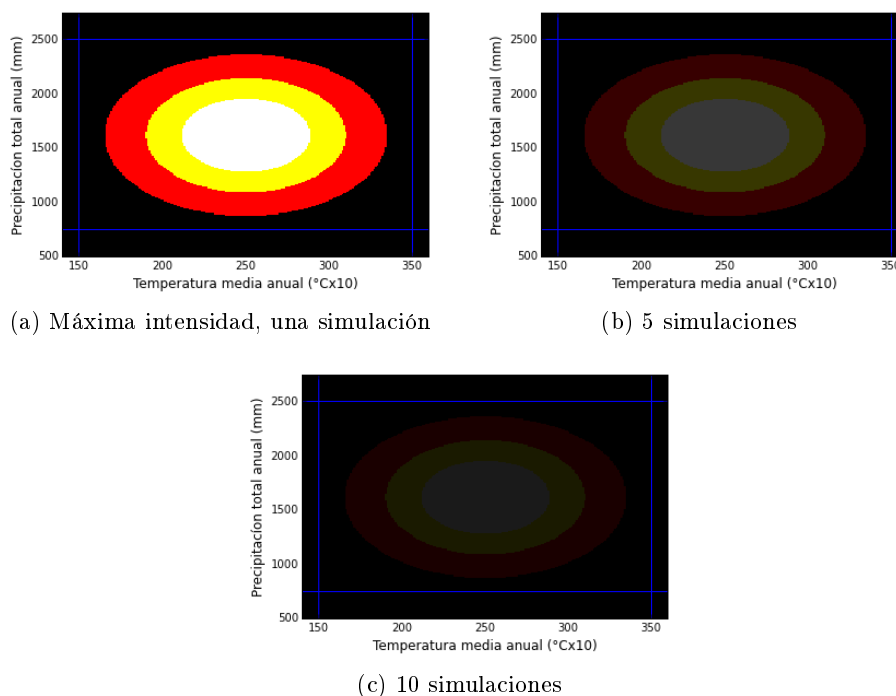


Figura 3.8: Intensidad del color según el número de simulaciones. Contornos de 0.25, 0.5 y 0.75 correspondientes al valor esperado a priori.

color rojo pero con mayor opacidad conforme c es cercano a cero. Lo mismo ocurre para el color amarillo. Al sumar los colores coordenada por coordenada se obtienen los colores intermedios, siempre y cuando todos los valores sean menores a uno. Una tonalidad naranja, por ejemplo, se consigue así: $(0.5, 0, 0) + (0.5, 0.5, 0) = (1, 0.5, 0)$. Entonces, para construir la gráfica de incertidumbre primero se toma una rejilla lo suficientemente fina para colorear pixel por pixel. El color de un pixel p es la suma de los colores sobre todas las simulaciones $c(p) = \sum_{i=1}^T c_i(p)$. Cada simulación aporta tonalidades de rojo, amarillo o blanco de acuerdo a sus respectivos contornos (figura 3.8). La aportación de cada una a la coloración tiene una intensidad de $1/T$. Es decir, $c_i(p) = (1/T, 0, 0)$ si para la i -ésima simulación el punto p tiene score 0.25, $c_i(p) = (1/T, 1/T, 0)$ si está entre 0.25 y 0.5, y $c_i(p) = (1/T, 1/T, 0)$ si es mayor a 0.75. Así, el color final de la imagen es una mezcla de los colores que cada simulación aporta según sus contornos.

Para interpretar la gráfica se debe recordar que conforme el número de simulaciones que coinciden con cierto score para un pixel aumenta, se obtienen las tonalidades más intensas (figura 3.8a). Un área completamente blanca corresponde a una probabilidad aproximada de 1 de pertenecer al contorno

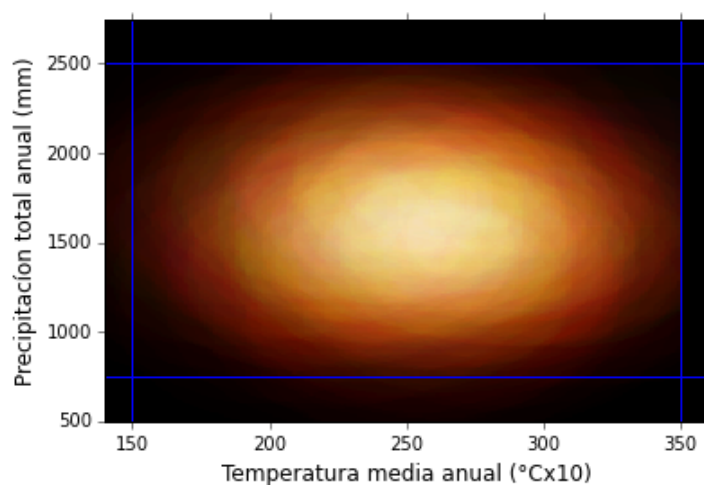


Figura 3.9: Ejemplo gráfica de incertidumbre
Variabilidad según la distribución a priori de los parámetros.

con score mayor a 0.75. Conforme el color del área se acerca a amarillo la probabilidad disminuye. Cuando el color es amarillo intenso la probabilidad de pertenecer al contorno más pequeño es cero, pero la probabilidad de pertenecer al contorno de 0.5 es uno. Conforme el amarillo se vuelve naranja, nuevamente esta probabilidad (aproximada) disminuye. Las zonas naranjas tiene probabilidad alrededor de 0.5 tener un score mayor a 0.5. Finalmente, las zonas de color rojo intenso, tienen probabilidad aproximada 1 de tener un score mayor a 0.25, y conforme el rojo se vuelve opaco la probabilidad disminuye.

La gráfica 3.9 corresponde a la variabilidad que se obtiene a priori. Es decir, en vez de simulaciones de la posterior, se utilizan simulaciones de $\mathcal{P}_{\mu,A,\delta}(\mu, A, \delta)$, donde los parámetros de (μ, A, δ) se asignan como se explica en la sección 3.3.2. Se puede percibir que, a priori, la variabilidad es alta aunque restringida al cuadrado a priori. En ningún punto el color es rojo, amarillo ni blanco con la máxima intensidad. El cambio de tonalidades es bastante difuso, lo que significa que a priori no se está suponiendo ninguna forma en el nicho. No obstante, se puede concluir que las zona centrales del cuadrado tiene mayor probabilidad (a priori) de tener un score alto.

3.5. Programación

Para implementar la metodología, el lenguaje de programación recomendado es Python, principalmente por su velocidad. La rapidez es un factor importante sobretodo por la discretización que aparece en el modelo ya que

implica una evaluación en 20,000 puntos (cardinalidad de $\mathcal{E}(t; G)$), evaluación que se realiza en cada iteración del MCMC.

Por otro lado, el método MCMC que se propone usar es el T-walk, desarrollado por Christen y Fox (2010). Una de las ventajas de este método es que los resultados son satisfactorios para una gran variedad de distribuciones objetivo. Así, no se pierde tiempo preocupándose por los detalles del método MCMC, permitiendo enfocarse en la modelación estadística e interpretación de resultados.

Otro punto importante acerca de la programación son los errores computacionales que pueden surgir en la implementación. Estos pueden llevar a conclusiones erróneas por lo que se debe ser cuidadoso para evitar errores numéricos.

3.5.1. T-walk en Python

Al realizar inferencia desde un enfoque bayesiano, generalmente se tiene que recurrir a un método MCMC. En particular, el Metrópolis–Hasting es comúnmente usado. Para aplicarlo, se propone una distribución instrumental, la cual influye fuertemente en la eficiencia del método. Para especificar una distribución adecuada, es necesario entender un poco el comportamiento de la posterior. Esto en muchos casos no es posible o requiere de mucho esfuerzo y tiempo. El t-walk (Christen y Fox, 2010) es un método—tipo M–H— autoajutable que conserva la propiedad de convergencia de la cadena de markov resultante (bajo las condiciones usuales).

El t-walk está disponible en R, Python, Matlab y C⁺⁺. Para usarlo únicamente es necesario introducir la función energía, el soporte de la posterior y dos puntos iniciales. Aunque en vez de usar puntos iniciales cualesquiera, es recomendable simular de la función a priori.

Existen otras modalidades M–H que son igualmente sencillas de aplicar. El Random Walk M–H (RWMH), por ejemplo, es frecuentemente usado. El RWMH se caracteriza por el uso de cierto tipo de distribuciones instrumentales, entre las cuales se encuentran las densidades normales $\mathcal{N}(\cdot|\kappa^2)$. Éstas son comúnmente usadas ya que fácilmente se simula de ellas. Sin embargo, ofrece resultados satisfactorios sólo para distribuciones objetivo muy sencillas. Además, en contraste con el t-walk, es necesario ajustar los parámetros, en este caso, la varianza κ^2 .

En la figura 3.10 se presenta una breve comparación de los 2 métodos. Para favorecer al RWMH, se eligieron varianzas (κ^2) iguales a las varianzas

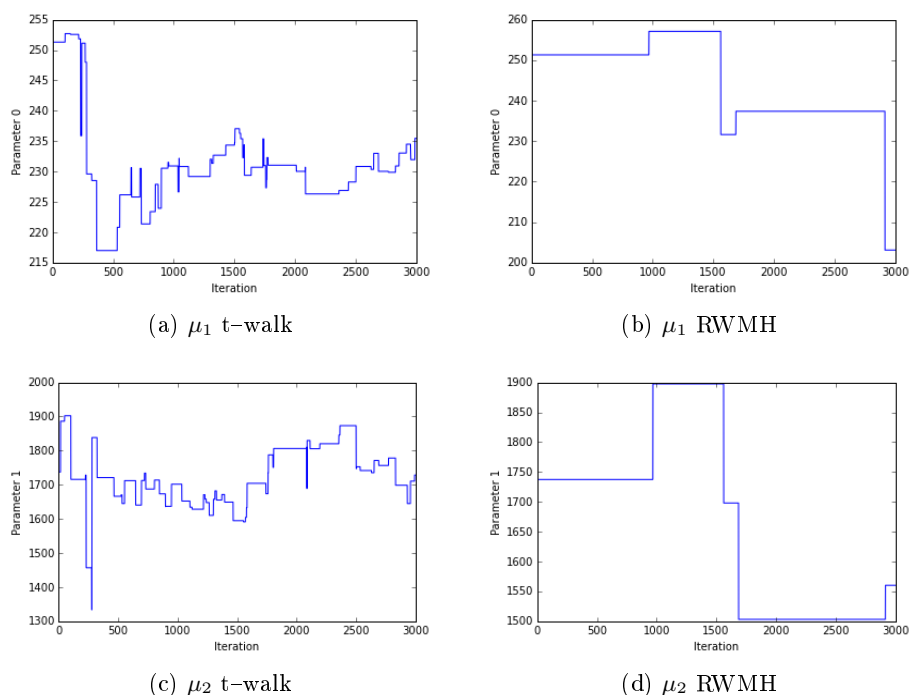


Figura 3.10: Comparación de resultados: t-walk vs. RWMH

Series de tiempo para el parámetro μ , $T = 3000$ iteraciones y misma semilla (mismos valor inicial) para ambos métodos.

muestrales resultantes de haber simulado previamente con el t-walk. A pesar de esto, los resultados favorecen al t-walk ya que, como se muestra en el figura, el método RWMH tarda bastante en recorrer el soporte de la función lo que implica que se deben tomar rezagos más grandes que con el t-walk.

3.5.2. Comparación de velocidad: Python vs. R

R es un lenguaje muy amigable para analizar datos. Casi cualquier herramienta estadística se encuentra ya disponible en este lenguaje y, además, son fáciles de usar. Por su parte, Python, aunque cuenta con capacidades estadísticas, se enfoca más en la productividad y legibilidad de códigos. Por ello, ha sido mayormente adoptado por programadores y desarrolladores de software que desean introducirse en la ciencia de datos. Sin embargo, la principal ventaja de Python sobre R consiste en la rapidez.

En lo que concierne a esta tesis, el optimizar tiempos es importante principalmente por la expresión de la discretización sobre el espacio disponible. Recuérdese que en cada iteración del método MCMC, las simulaciones se evalúan en la posterior. Según la expresión de la verosimilitud, el evaluar

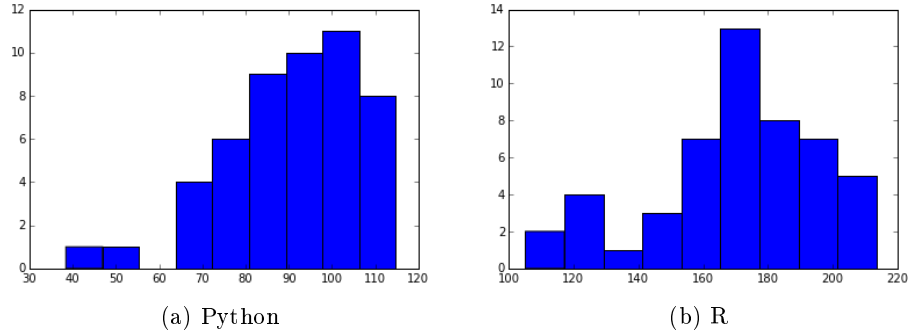


Figura 3.11: Comparación de velocidad con R y Python
 Histogramas con 50 valores del tiempo (segundos) que se tarda cada
 lenguaje de programación para $T = 200$ iteraciones.

una simulación en la posterior implica evaluar la densidad $f(y|\mu, A, \delta)$ para toda $y \in \mathcal{E}(t; G)$, donde éste espacio consta de 20,000 elementos. En la figura 3.11 se comparan los tiempos de R y Python para un mismo número de iteraciones. Para 200 iteraciones del t-walk, el software Python tarda en promedio (50 veces) 90 segundos mientras que R 170 segundos, casi el doble de tiempo. El mínimo para Python fue de 38 segundos, considerablemente más bajo. El tiempo del software R no bajó de los 105 segundos. Esto se ve reflejado en las medianas para las 50 pruebas: 90 segundos para Python y 173 para R.

3.5.3. Problemas computacionales

Para evitar errores computacionales durante el MCMC, se debe de ser cuidadoso con especificar bien el soporte de la distribución posterior. En este caso, es importante restringir los parámetros $\delta \in (-1, 1)$ y A una matriz positiva definida. Además de estas restricciones, es necesario incluir una última con respecto al término de la función energía siguiente,

$$n \log \left(\sum_{y \in \mathcal{E}(t)} \exp \left[-\frac{1}{2} (y - \mu)' A (y - \mu) \right] \Phi \left[\frac{y_1 - \mu_1}{\sigma_1} \beta_1(\delta) + \frac{y_2 - \mu_2}{\sigma_2} \beta_2(\delta) \right] \right). \quad (3.10)$$

Para ciertos valores $\theta = (\mu, A, \delta)$ el valor $f(y|\mu, A, \delta)$ puede ser prácticamente cero para todo $y \in \mathcal{E}(t; G)$, generalmente por que la precisión A es muy grande y la media está alejada de todos los puntos (contornos demasiado pequeños y con centros fuera del espacio disponible). En estos casos, la suma que aparece en la expresión 3.10 se vuelve cero computacionalmente, provocando que la función energía se vaya a menos infinito. Así, la cadena de Markov en el MCMC se estanca en este valor, obteniéndose nichos nada

razonables.

El error es fácilmente evitable una vez detectado. Sólo es necesario incluir la restricción,

$$\sum_{y \in \mathcal{E}(t)} \exp \left[-\frac{1}{2} (y - \mu)' A (y - \mu) \right] \Phi \left[\frac{y_1 - \mu_1}{\sigma_1} \beta_1(\delta) + \frac{y_2 - \mu_2}{\sigma_2} \beta_2(\delta) \right] \neq 0$$

en la función soporte que se introduce al t-walk.

Capítulo 4

Ejemplos

En este capítulo se presentan los resultados de aplicar la metodología a 3 géneros de plantas. La distribución a priori y las ponderaciones se fijan de acuerdo a lo mencionado en la sección 3.3. Se interpretan las estimaciones puntuales de los nichos (MAP) y la gráfica de variabilidad. Para la gráfica de incertidumbre, se consideran los contornos correspondientes a los scores 0.25, 0.5 y 0.75. Para el MAP, se consideran estos 3 valores y se aumentan el contorno asociado al score 0.03 ya que proporciona una estimación razonable para el nicho fundamental. Al final del capítulo, se comparan las estimaciones puntuales de otros experimentos interesantes.

4.1. Género *Arachis*

Arachis es un género que comprende alrededor de 70 especies. Una de las especies, el cacahuete (*Arachis hypogaea*), es muy estudiada por su consumo global. Esta especie es nativa de América del Sur. De esta región, en específico de Brasil, provienen las ocurrencias que se analizan (figura 4.1a). En la imagen 4.1b, se observa que los datos están concentrados en una zona que no aparenta tener una densidad de puntos muy desequilibrada.

Los resultados para este género se presentan en la figura 4.2. La estimación puntual del nicho fundamental (MAP) se representa por el contorno más grande (score de 0.03). En la figura 4.2b se observa que algunas observaciones, los de mayor precipitación, están en una zona menos densa que el resto. Esta es la razón por la que la estimación del nicho fundamental se expande hacia estas zonas, a pesar de que no haya datos. En la gráfica de variabilidad se observa que la forma del MAP se conserva aunque la zona de mayor in-

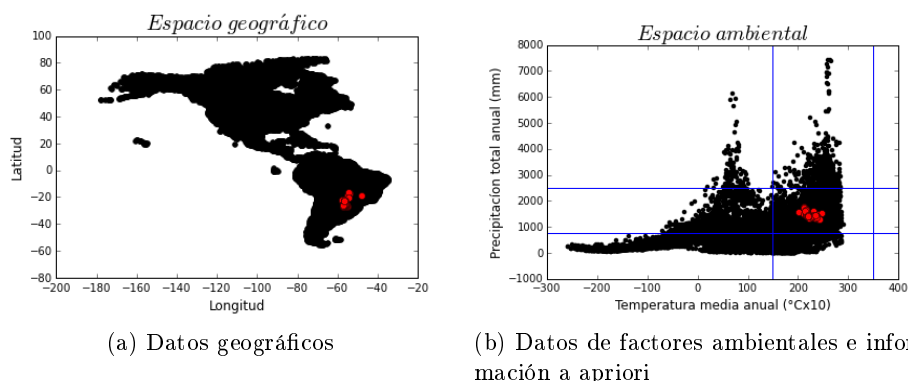


Figura 4.1: Datos género *Arachis*

certidumbre es, precisamente, la que corresponde a la zona menos densa. A pesar de esto, se puede concluir que la especie soporta ambientes de menor temperatura y mayor precipitación con respecto a su media (cola más pesada). Hacia la otra dirección (mayor temperatura y menor precipitación) hay bastantes observaciones aunque con poca ponderación ya que se encuentran en una zona muy densa. Es probable que haya tantas observaciones en esa región, no por que la especie la prefiera, sino porque se muestrean muchos puntos con esas condiciones. Además, el MAP y la gráfica de incertidumbre concuerdan con que la especie presenta poca inclinación hacia ambientes muy húmedos (mucho precipitación y altas temperaturas).

4.2. Género *Chenopodium*

Chenopodium es un género que comprende alrededor de 150 especies. Son comúnmente conocidos como cenizos o quinoas. Los datos geográficos de la especie se encuentran esparcidos en varios países de Sur América, principalmente cerca de las costas (figura 4.3a). Estas ocurrencias en el espacio ambiental (figura 4.3b) abarcan gran parte del espacio disponible. Se observan algunos puntos de zonas muy poco densas. Además, este es un caso en el que la información a priori y los datos comprenden rangos ambientales distintos.

Los resultados para este género se presentan en la figura 4.4. Para esta especie se encontraron 2 modas en la posterior, dando lugar a dos estimaciones puntuales para el nicho. La primera (figura 4.4b) está más acorde a la información a priori, mientras que la segunda (figura 4.4d) representa preferencia hacia temperaturas más bajas. Ambas modas se centran en precipitaciones mayores a la del promedio de los datos. Esto se debe a que las observaciones de mayor ponderación son las que corresponden a precipitaciones altas. A

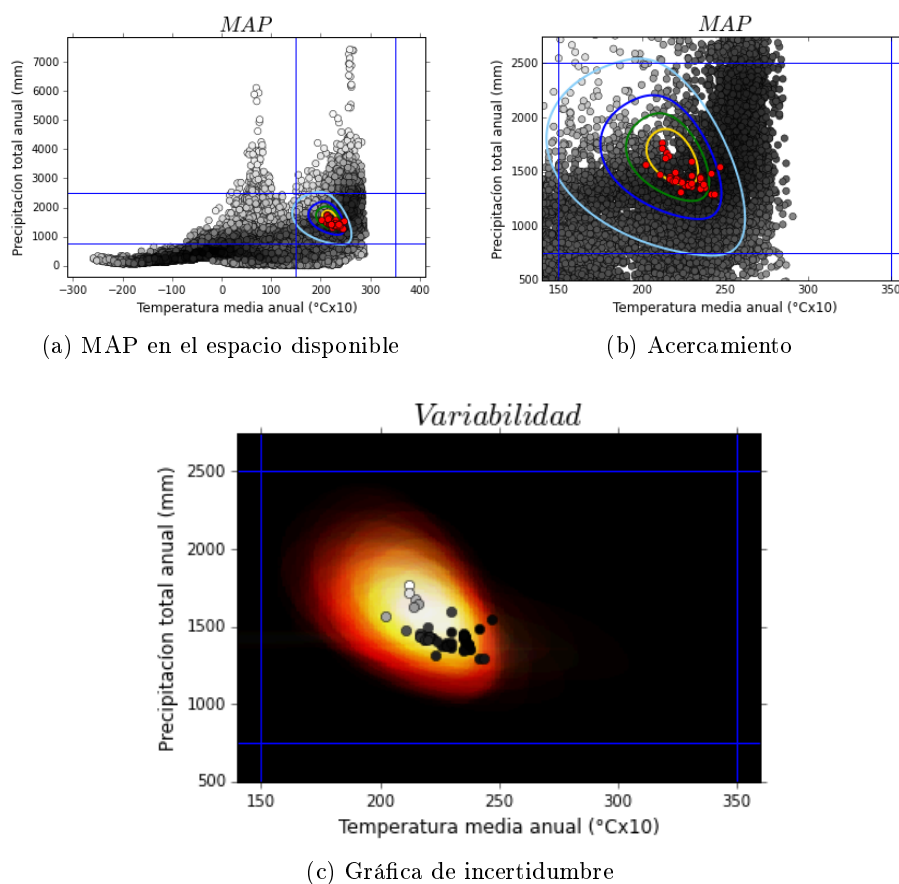


Figura 4.2: Resultado género *Arachis* MAP y variabilidad (50 simulaciones). Escala de grises proporcional a la ponderación, donde negro es el punto con mayor densidad y blanco el de menor.

pesar de la falta de información en zonas de alta precipitación, se puede concluir que si climas aceptables para la especie. Las dos modas se identifican al analizar las series de tiempo de los parámetros en el MCMC (figura 4.5). En ellas, se observa que la cadena pasa aproximadamente el mismo tiempo en las dos modas, lo que también se refleja en la gráfica de variabilidad (figura 4.4e). Con esta gráfica se puede concluir que el rango de temperatura que soporta la especie es más reducido de lo que se esperaba al observar las modas. El nicho de esta especie abarca un área bastante amplia en el espacio ambiental, presentado preferencia hacia temperaturas templadas y soportando bien poca y mucha precipitación.

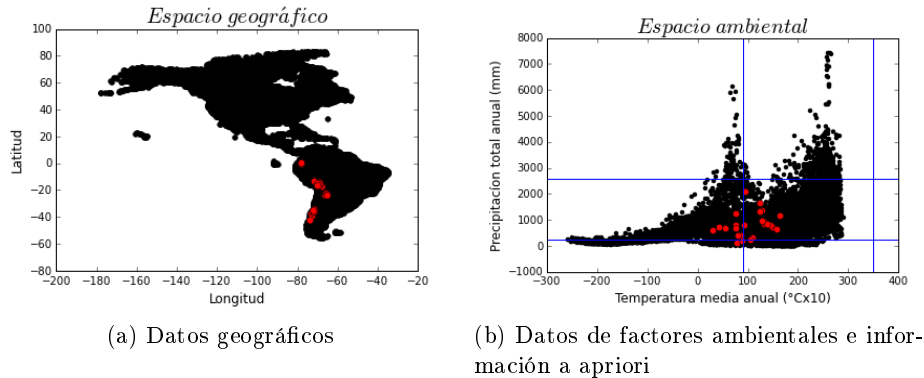


Figura 4.3: Datos género *Chenopodium*

4.3. Género *Panicum*

El género *Panicum* comprende alrededor 450 especies. Tanto la información a priori como las observaciones ambientales indican que la especie crece en zonas de poca precipitación, característica principal de los desiertos.

Los resultados para este género se presentan en la figura 4.7. La estimación del nicho fundamental de esta especie (figura 4.7b) abarca amplios rangos de temperatura mas no de precipitación, con preferencia a precipitaciones bajas. La zona de alta preferencia de la especie consiste en una pequeña zona de temperaturas altas y precipitaciones bajas (contorno de score 0.25). Aunque en la representación de la variabilidad se pierde el sesgo (figura 4.7c), las conclusiones son muy similares a las obtenidas con la estimación puntual.

4.4. Otros experimentos

4.4.1. Ponderaciones más contrastantes

Usando una ponderación más contrastante ($a = 1$), las estimaciones puntuales de la estructura del nicho fundamental para cada especie se presentan en la figura 4.8. Comparando con los resultados anteriores, la estimación puntual del nicho de *Panicum* no cambia considerablemente. Los nichos para las otras dos especies claramente son distintos. Para el caso de *Arachis*, el nicho se expande más hacia las zonas menos densas. El nicho de la especie *Chenopodium* también presenta mayor preferencia a estas zonas, sólo que en vez de expandirse, se desplaza, provocando que la estimación del nicho no contenga a todas las observaciones. Esto se debe a que algunas observaciones, las que no están contenidas en el nicho, se encuentran en zonas demasiado

densas.

4.4.2. Sin ponderaciones

Tomando $a = 0$ para definir las ponderaciones, es equivalente a tomar pesos iguales. Los resultados de este experimento se muestran en la figura 4.9. La estructura del nicho de *Arachis*, abarca un poco las zonas menos densas. Esto es de esperarse ya que con sólo la discretización, el modelo ya considera (hasta cierto punto) la densidad de puntos en el espacio disponible. Sin embargo, para esta especie, la forma del nicho cambia radicalmente, moldeándose según la forma de los datos. Para la especie *Chenopodium*, el nicho resultante se percibe desplazado hacia precipitaciones más bajas, en comparación con los otros dos casos, aunque si abarca zonas poco densas. Para esta especie, la forma del nicho se conserva. Para la especie *Panicum*, el considerar todos los puntos por igual da pie a una estructura poco razonable, principalmente porque los contornos abarcan gran parte de precipitaciones negativas.

4.4.3. A priori con mayor incertidumbre

Un experimento interesante consiste en medir el efecto de la información a priori. En los casos anteriores, se fijaban las distribuciones a priori con poca incertidumbre con respecto a la información previa. Un caso de mucha incertidumbre se encuentra al tomar otros parámetros para la matriz A y para μ . La variabilidad de la matriz A se controla con el parámetro α . Tomando $\alpha = 5$, una densidad normal con mayor varianza para μ (marginales entre los rangos de tolerancia con 50 % de probabilidad) y $a = 0.4$ para las ponderaciones, se obtienen los resultados de la figura 4.10. El nicho para *Arachis* es una estructura muy pequeña, abarcando solamente el área donde se encuentran los datos. El de *Chenopodium* abarca menor rango de temperatura aunque alcanza precipitaciones mucho más altas. Para *Panicum* se reducen ambos rangos dejando fuera dos observaciones.

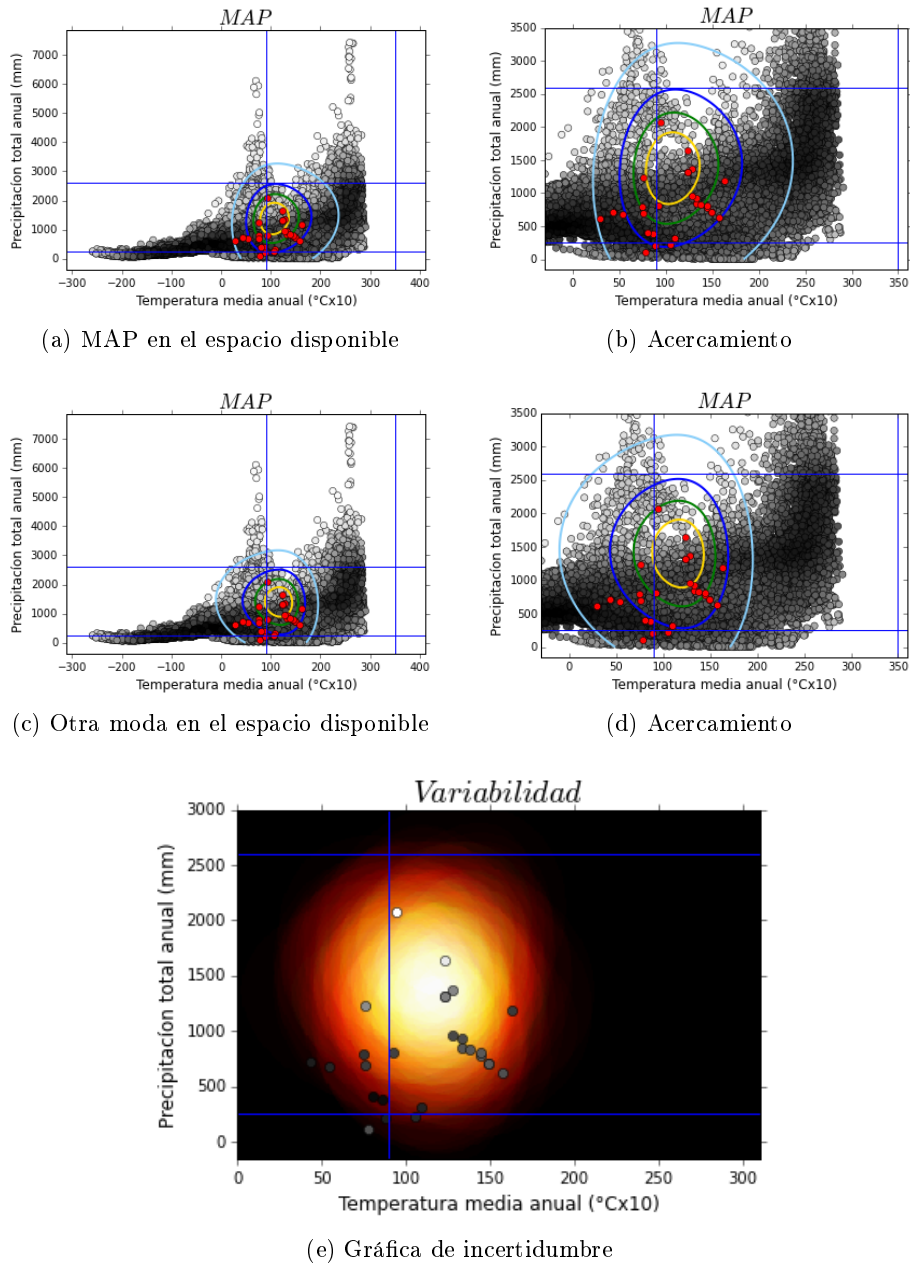


Figura 4.4: Resultado género *Chenopodium* MAP y variabilidad (50 simulaciones). Escala de grises proporcional a la ponderación, donde negro es el punto con mayor densidad y blanco el de menor.

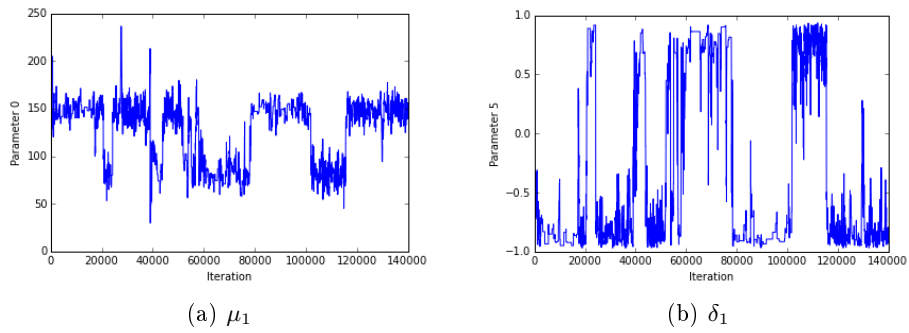


Figura 4.5: Modas de la posterior para la especie *Chenopodium*
Cadenas del MCMC para los parámetros μ y δ .

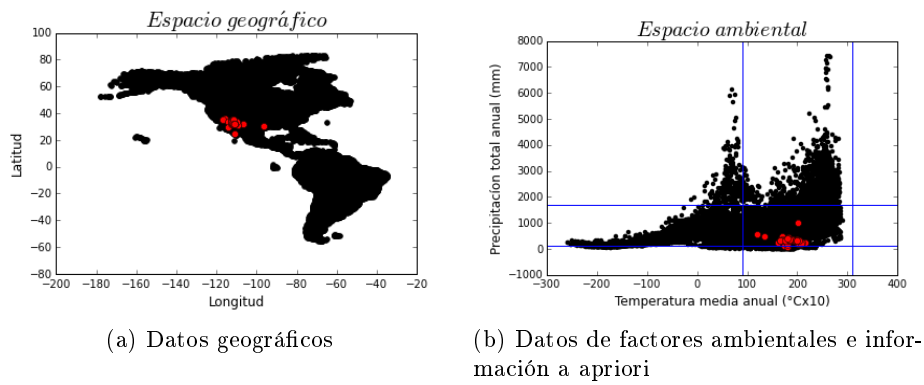


Figura 4.6: Datos género *Panicum*

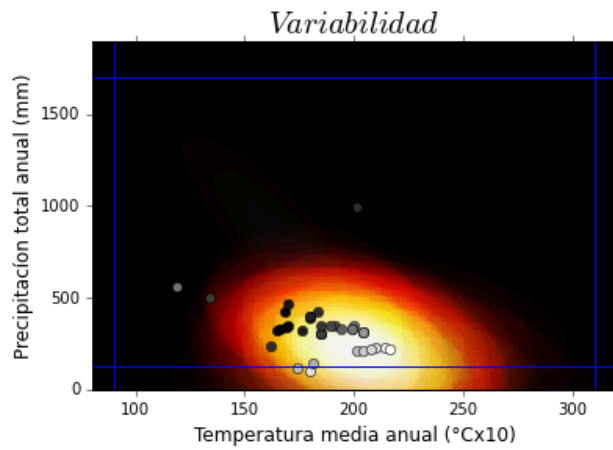
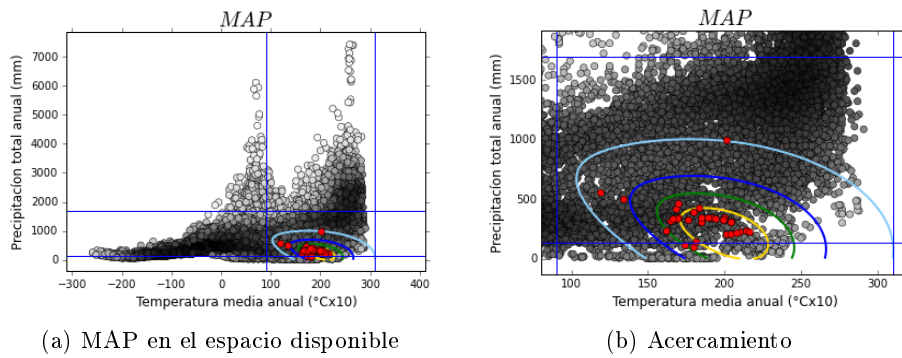


Figura 4.7: Resultado género *Panicum* MAP y variabilidad (40 simulaciones). Escala de grises proporcional a la ponderación, donde negro es el punto con mayor densidad y blanco el de menor.

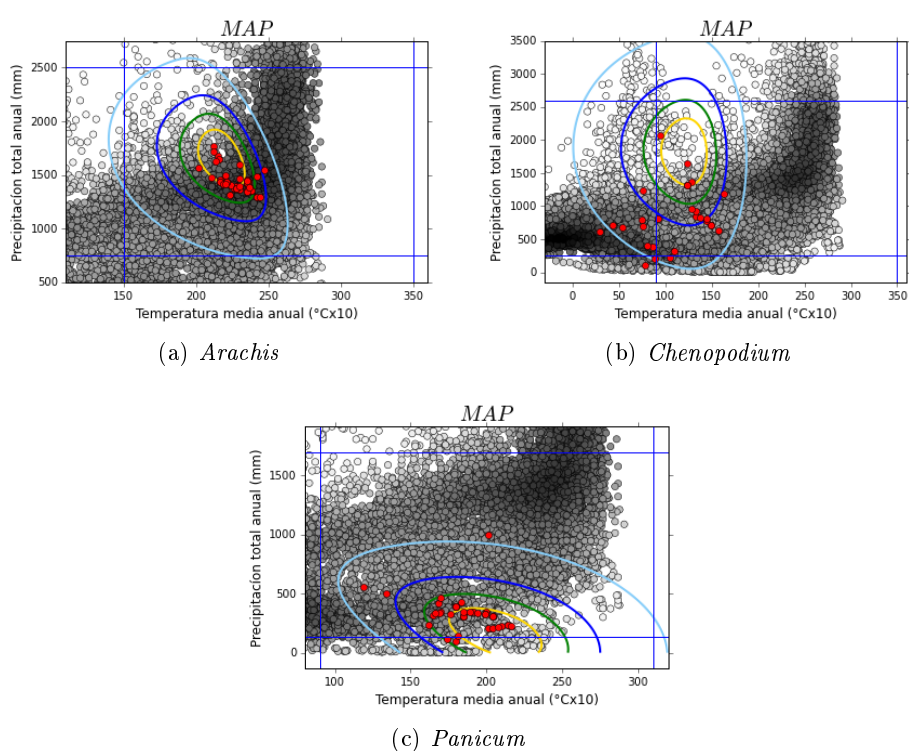


Figura 4.8: Estimaciones para $a = 1$ MAP de cada especie. Escala de grises proporcional a la ponderación, donde negro es el punto con mayor densidad y blanco el de menor.

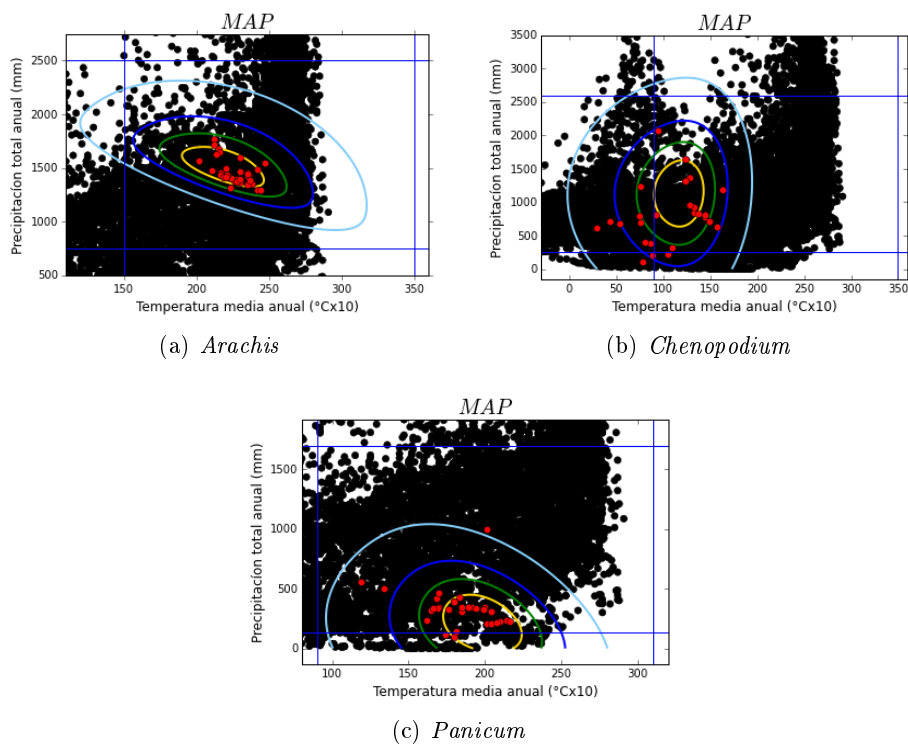


Figura 4.9: Estimaciones para $a = 0$
MAP de cada especie. Escala de grises proporcional a la ponderación,
donde negro es el punto con mayor densidad y blanco el de menor.

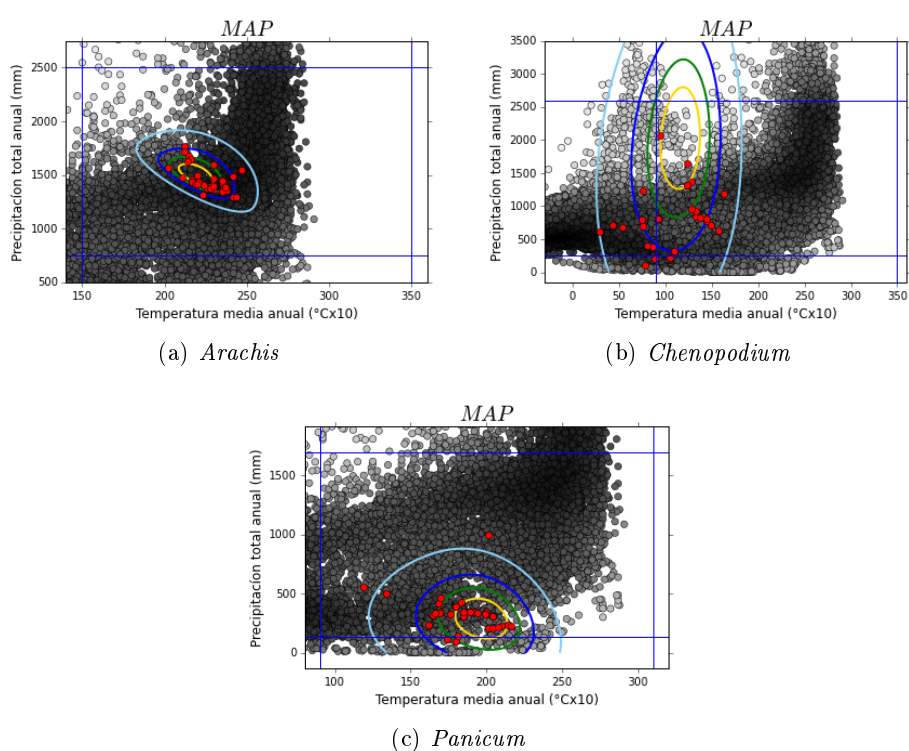


Figura 4.10: Estimaciones con mayor varianza a priori para μ y A MAP de cada especie. Escala de grises proporcional a la ponderación, donde negro es el punto con mayor densidad y blanco el de menor.

Capítulo 5

Discusión y conclusiones

Para estimar el nicho fundamental de una especie es de suma importancia introducir en el modelo la información del espacio disponible. De otra forma, las estimaciones corresponderían a estructuras de otros nichos, como el realizado o el potencial. Del espacio disponible se debe considerar tanto la forma como la densidad o abundancia. La forma puede ser considerada al discretizar sobre este espacio. La densidad del espacio afecta la probabilidad de observar una ocurrencia. Para contrarrestar el efecto se asignan ponderaciones a cada observación. Ciertamente, la manera de asignar estas ponderaciones de forma adecuada es un problema que sigue en discusión. Se resolverá conforme se conteste la pregunta: ¿qué tanto afecta la accesibilidad de la especie a las distintas combinaciones ambientales en la probabilidad de observarla? Aunque al discretizar sobre $\mathcal{E}(t; G)$ el modelo ya considera la densidad de puntos sobre este espacio, asigna la mínima importancia posible a las observaciones en zonas poco densas. La forma en que estos puntos van adquiriendo importancia se controla con el parámetro a , fijando el ancho de banda en $h = (25, 150)$. Cuando el parámetro es muy grande (arriba de uno) el contraste entre las diferentes zonas es muy notorio y las zonas poco densas adquieren demasiada importancia. En estos casos, el nicho puede centrarse en pocos puntos (los menos densos) e incluso calificar con un score bajo una gran cantidad de observaciones, ya sea porque el contorno estimado resulta muy pequeño, o bien porque se desplaza hacia las zonas menos densas. Para valores de a entre cero y uno se obtienen resultados aceptables y bastantes interpretables, sin embargo la variedad de densidades que ofrece este rango sugiere seguir buscando un mejor argumento teórico para la selección de este valor.

Por otro lado, incluir información a priori de los rangos de tolerancia influye en la forma del nicho y en su tamaño. Al aumentar la incertidumbre en la distribución a priori, la forma del nicho queda determinada por lo que señalan las observaciones y las ponderaciones, obteniéndose, en algunos casos, nichos muy pequeños o angostos sobre un eje. Esto es precisamente lo que se intenta evitar ya que el nicho fundamental se concibe como un conjunto amplio en proporción al área que abarca los datos pues las observaciones son resultado de muchos factores que afectan la accesibilidad de la especie a las distintas combinaciones ambientales.

Ya que en la base de nuestra formulación se encuentra una estimación de densidad por kernel, tal vez sea necesario considerar que el soporte de la densidad subyacente es acotado. En este caso, se ha sugerido el uso de correcciones en los kernels usados. Ésto con la finalidad de mantener el soporte de manera sistemática, y no generar sesgos o estimaciones incoherentes. Estas correcciones son conocidas como Edge Correction o Boundary Correction. Se puede consultar Karunamuni y Alberts (2005) para este tema.

Para la interpretación de resultados, representó una ventaja el poder visualizar la variabilidad de las estimaciones del nicho. En algunos casos, con las estimaciones puntuales se llegaban a conclusiones erróneas debido a la alta variabilidad. Con esta herramienta fue posible medir con mayor precisión la preferencia de la especie a las combinaciones ambientales. Esta gráfica es fácil de construir y de interpretar.

Finalmente, con respecto a la programación, el implementar la metodología en Python permitió que fuera posible realizar varios experimentos, variando los parámetros e incluso el modelo, sin que el tiempo fuese un impedimento. En comparación con Python, el software R tarda casi el doble de tiempo. Aunque el tiempo parece no ser un factor tan relevante cuando se tratan de $T = 200$ iteraciones, lo es cuando son necesarias $T = 500,000$ iteraciones para analizar la variabilidad. Este número de iteraciones fue necesario en los casos donde los rezagos eran hasta de $T = 5000$ iteraciones. En cuanto al método MCMC, dada que la función objetivo es compleja, hubiese implicado mucho esfuerzo el buscar una distribución instrumental adecuada. Con el t-walk esto no fue necesario y, en general, los resultados fueron satisfactorios, siendo suficientes, en la mayoría de los experimentos, rezagos de $T = 1000$ iteraciones.

Bibliografía

*La paciencia es la compañera de la
sabiduría.*

San Agustín

- AZZALINI, A. The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics*, vol. 32, páginas 159–188, 2005.
- AZZALINI, A. y VALLE, A. D. The multivariate skew-normal distribution. *Biometrika*, vol. 83(4), páginas 715–726, 1996.
- CHRISTEN, J. A. y FOX, C. A general purpose sampling algorithm for continuous distributions (the t-walk). *Internactional Society for Bayesian Analysis*, vol. 5(2), páginas 263–281, 2010.
- CHRISTEN, J. A. y SOBERÓN, J. On the problem of estimating a fundamental niche from ocurrence data. 2016.
- ERDELY, A. y GUTIÉRREZ, E. Monografía de estadística bayesiana. 2007.
- GAMERMAN, D. y LOPES, H. F. *Markov Chain Monte Carlo, 2th Edition*. Chapman & Hall. CRC Press, 2006.
- GELMAN, A., JOHN B. CARLIN, H. S. S., DUNSON, D. B., VEHTARI, A. y RUBIN, D. B. *Bayesian Data Analysis, 3th Edition*. Chapman & Hall. CRC Press, 2014.
- HIJMANS, R. J., CAMERON, S. E., PARRA, J. L., JONES, P. G. y JARVIS, A. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 2005.
- HUTCHINSON, G. Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, 1957.
- KARUNAMUNI, R. y ALBERTS, T. On boundary correction in kernel density estimation. *Statistical Methodology*, 2005.

MAGUIRE, B. Niche response structure and the analytical potencial of its relationship to the habitat. *The American Naturalist*, vol. 107(954), páginas 213–246, 1973.

PETERSON, A. T., SOBERÓN, J., PEARSON, R. G., ANDERSON, R. P., MARTÍNEZ-MEYER, A., NAKAMURA, M. y ARAÚJO, M. B. *Ecological Niches and Geographic Distributions*. Princeton University Press, 2011.

WHITTAKER, R. H. *Communities and Ecosystems*. Macmillan USA, 1970.

Apéndice A

Densidad \mathcal{SN} bivariada

Sea $Z \sim \mathcal{SN}(\omega, \beta)$ y $X_i = \sigma Z_i + \mu_i$, $i = 1, 2$. Para obtener la expresión de la densidad de $X = (X_1, X_2)$, se aplica el Teorema de Cambio de Variable:

$X = \mu + SZ = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} \mu_1 + \sigma_1 z_1 \\ \mu_2 + \sigma_2 z_2 \end{pmatrix}$ es una transformación uno-a-uno, donde $S = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}$.

Despejando Z ,

$$Z = S^{-1}(X - \mu) = \begin{pmatrix} 1/\sigma_1 & 0 \\ 0 & 1/\sigma_2 \end{pmatrix} \begin{pmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{pmatrix} = \begin{pmatrix} \frac{X_1 - \mu_1}{\sigma_1} \\ \frac{X_2 - \mu_2}{\sigma_2} \end{pmatrix}.$$

Luego, se calcula el jacobiano

$$\left| \frac{\partial(Z_1, Z_2)}{\partial(X_1, X_2)} \right| = \begin{vmatrix} \frac{\partial Z_1}{\partial X_1} & \frac{\partial Z_1}{\partial X_2} \\ \frac{\partial Z_2}{\partial X_1} & \frac{\partial Z_2}{\partial X_2} \end{vmatrix} = \begin{vmatrix} \frac{1}{\sigma_1} & 0 \\ 0 & \frac{1}{\sigma_2} \end{vmatrix} = \frac{1}{\sigma_1 \sigma_2}.$$

Entonces,

$$\begin{aligned} f_X(x; \mu, \sigma_1, \sigma_2, \omega, \beta) &= \frac{1}{\sigma_1 \sigma_2} f_Z(S^{-1}(x - \mu); \omega, \beta) \\ &= \frac{2}{\sigma_1 \sigma_2} \phi_2(S^{-1}(x - \mu); \omega, \beta) \Phi\left(\beta_1 \frac{x_1 - \mu_1}{\sigma_1} + \beta_2 \frac{x_2 - \mu_2}{\sigma_2}\right) \\ &= \frac{2}{2\pi \sigma_1 \sigma_2 |\Omega|^{1/2}} \exp\left(-\frac{1}{2}(X - \mu)^T S^{-1} \Omega^{-1} S^{-1} (X - \mu)\right) \Phi\left(\beta_1 \frac{x_1 - \mu_1}{\sigma_1} + \beta_2 \frac{x_2 - \mu_2}{\sigma_2}\right) \end{aligned}$$

Desarrollando algunos términos por separado,

$$\begin{aligned}\sigma_1\sigma_2|\Omega|^{1/2} &= \sigma_1\sigma_2 \begin{vmatrix} 1 & \omega \\ \omega & 1 \end{vmatrix}^{1/2} = \sigma_1\sigma_2(1-\omega^2)^{1/2} = (\sigma_1^2\sigma_2^2(1-\omega^2))^{1/2} = (\sigma_1^2\sigma_2^2 - \omega^2\sigma_1^2\sigma_2^2)^{1/2} \\ &= \begin{vmatrix} \sigma_1^2 & \omega\sigma_1\sigma_2 \\ \omega\sigma_1\sigma_2 & \sigma_2^2 \end{vmatrix}^{1/2} = |\Sigma|^{1/2}.\end{aligned}$$

$$\begin{aligned}S^{-1}\Omega^{-1}S^{-1} &= \begin{pmatrix} 1/\sigma_1 & 0 \\ 0 & 1/\sigma_2 \end{pmatrix} \begin{pmatrix} 1 & \omega \\ \omega & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1/\sigma_1 & 0 \\ 0 & 1/\sigma_2 \end{pmatrix} \\ &= \frac{1}{1-\omega^2} \begin{pmatrix} 1/\sigma_1 & 0 \\ 0 & 1/\sigma_2 \end{pmatrix} \begin{pmatrix} 1 & -\omega \\ -\omega & 1 \end{pmatrix} \begin{pmatrix} 1/\sigma_1 & 0 \\ 0 & 1/\sigma_2 \end{pmatrix} \\ &= \frac{1}{1-\omega^2} \begin{pmatrix} 1/\sigma_1 & -\omega/\sigma_1 \\ -\omega/\sigma_2 & 1/\sigma_2 \end{pmatrix} \begin{pmatrix} 1/\sigma_1 & 0 \\ 0 & 1/\sigma_2 \end{pmatrix} \\ &= \frac{1}{1-\omega^2} \begin{pmatrix} 1/\sigma_1^2 & -\omega/\sigma_1\sigma_2 \\ -\omega/\sigma_1\sigma_2 & 1/\sigma_2^2 \end{pmatrix} = \frac{1}{(1-\omega^2)\sigma_1^2\sigma_2^2} \begin{pmatrix} \sigma_2^2 & -\omega\sigma_1\sigma_2 \\ -\omega\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} = \Sigma^{-1}.\end{aligned}$$

donde $\Sigma = \begin{pmatrix} \sigma_1^2 & \omega\sigma_1\sigma_2 \\ \omega\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$.

Por lo tanto, la densidad de X es

$$\begin{aligned}f_X(x; \mu, \Sigma, \beta) &= \frac{2}{2\pi|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(X-\mu)^T \Sigma^{-1}(X-\mu)\right) \Phi\left(\beta_1 \frac{x_1 - \mu_1}{\sigma_1} + \beta_2 \frac{x_2 - \mu_2}{\sigma_2}\right) \\ &= 2\psi(x; \mu, \Sigma) \Phi\left(\beta_1 \frac{x_1 - \mu_1}{\sigma_1} + \beta_2 \frac{x_2 - \mu_2}{\sigma_2}\right), \quad z \in \mathbb{R}^2.\end{aligned}$$

donde $\psi(x; \mu, \Sigma)$ es la densidad Normal bivariada con media μ y matriz de covarianza Σ .

Apéndice B

Jacobiano de la transformación para el parámetro de sesgo.

En este apartado, se obtiene la expresión para $|J| = \begin{vmatrix} \frac{\partial \beta_1}{\partial \delta_1} & \frac{\partial \beta_1}{\partial \delta_2} \\ \frac{\partial \beta_2}{\partial \delta_1} & \frac{\partial \beta_2}{\partial \delta_2} \end{vmatrix}$, donde $\beta_1 = \frac{\delta_1 - \delta_2 \omega}{r^{1/2}}$, $\beta_2 = \frac{\delta_2 - \delta_1 \omega}{r^{1/2}}$ y $r = (1 - \omega^2)(1 - \omega^2 - \delta_1^2 - \delta_2^2 + 2\delta_1 \delta_2 \omega)$.

Primero, se calculan las derivadas parciales:

$$\begin{aligned} \frac{\partial \beta_1}{\partial \delta_1} &= \frac{r^{1/2} + \frac{1}{2}(\delta_1 - \delta_2 \omega)r^{-1/2}(1 - \omega^2)(-2\delta_1 + 2\delta_2 \omega)}{r} \\ &= r^{-1/2} - r^{1/2}(\delta_1 - \delta_2 \omega)^2(1 - \omega^2), \\ \frac{\partial \beta_2}{\partial \delta_2} &= r^{-1/2} - r^{1/2}(\delta_2 - \delta_1 \omega)^2(1 - \omega^2), \\ \frac{\partial \beta_1}{\partial \delta_2} &= \frac{-\omega r^{1/2} + \frac{1}{2}r^{1/2}(1 - \omega^2)(-2\delta_2 + 2\delta_1 \omega)(\delta_2 - \delta_1 \omega)}{r} \\ &= -\omega r^{-1/2} - r^{1/2}(1 - \omega^2)(\delta_2 - \delta_1 \omega)^2, \\ \frac{\partial \beta_2}{\partial \delta_1} &= -\omega r^{-1/2} - r^{1/2}(1 - \omega^2)(\delta_1 - \delta_2 \omega)^2. \end{aligned}$$

Ahora,

$$|J| = \left(\frac{\partial \beta_1}{\partial \delta_1} \right) \left(\frac{\partial \beta_2}{\partial \delta_2} \right) - \left(\frac{\partial \beta_1}{\partial \delta_2} \right) \left(\frac{\partial \beta_2}{\partial \delta_1} \right),$$

donde

$$\left(\frac{\partial \beta_1}{\partial \delta_1} \right) \left(\frac{\partial \beta_2}{\partial \delta_2} \right) = r^{-1} - (\delta_1 - \delta_2 \omega)^2(1 - \omega^2) - (\delta_2 - \delta_1 \omega)^2(1 - \omega^2) + c(\delta_1 - \delta_2 \omega)^2(\delta_2 - \delta_1 \omega)^2(1 - \omega^2)^2$$

y

$$\left(\frac{\partial \beta_1}{\partial \delta_2} \right) \left(\frac{\partial \beta_2}{\partial \delta_1} \right) = \omega^2 r^{-1} + \omega(1 - \omega^2)(\delta_2 - \delta_1 \omega)^2 + \omega(1 - \omega^2)(\delta_1 - \delta_2 \omega)^2 + c(1 - \omega^2)^2(\delta_1 - \delta_2 \omega)^2(\delta_2 - \delta_1 \omega)^2.$$

Por lo tanto,

$$\begin{aligned} |J| &= r^{-1} - (\delta_1 - \delta_2\omega)^2(1 - \omega^2) - (\delta_2 - \delta_1\omega)^2(1 - \omega^2) - \omega^2 r^{-1} \\ &\quad - \omega(1 - \omega^2)(\delta_2 - \delta_1\omega)^2 - \omega(1 - \omega^2)(\delta_1 - \delta_2\omega)^2 \\ &= r^{-1}(1 - \omega^2) - (1 + \omega)(1 - \omega^2)[(\delta_1 - \delta_2\omega)^2 + (\delta_2 - \delta_1\omega)^2]. \end{aligned}$$

*—¿Qué te parece desto, Sancho? — Dijo Don Quijote —
Bien podrán los encantadores quitarme la ventura,
pero el esfuerzo y el ánimo, será imposible.*

*Segunda parte del Ingenioso Caballero
Don Quijote de la Mancha
Miguel de Cervantes*

