Centro de Investigación en Matemáticas, A.C.

CIMAT

Descripción y caracterización de nichos ecológicos: una visión más cuantitativa del espacio ambiental

TESIS

que para obtener el grado de

Maestría en Ciencias con Especialidad en Probabilidad y Estadística

PRESENTA:

Miguel Angel López García

DIRECTOR DE TESIS:

Dr. Miguel Nakamura Savoy

Guanajuato, Gto, México *Septiembre del 2007*

Índice general

Agradecimientos

Quiero agradecer a mi director de Tesis el Dr. Miguel Nakamura Savoy, ya que sin su apoyo y paciencia, este trabajo no hubiera sido posible.

Al CIMAT por los recursos económicos proporcionados durante la realización de mis estudios de maestría. Así mismo, agradezco al CONACyT por la beca proporcionada para llevar a cabo dichos estudios. De igual forma, quiero agradecer al CONCyTEG por facilitar los recursos financieros para la elaboración de este proyecto de tesis.

A los profesores Dr. Fernando Ávila Murillo y Dr. Enrique Villa Diharce por haber aceptado ser mis sinodales.

Finalmente, quiero agradecer a: mi madre (por apoyarme en todo momento); a mis hermanos (por ser mis hermanos); a Edith (por ser una motivación para mi); a mis amigos (por su amistad); a todas aquellas personas que de manera indirecta hicieron posible el desarrollo y culminación de este trabajo.

Resumen

La concepción más elemental de nicho es que se trata del conjunto de características ambientales que necesita un organismo para sobrevivir. Existen diversas metodologías diseñadas para inferir nichos ecológicos. La mayor parte de estos métodos tienen poca formalidad matemática y consisten de algoritmos empíricos. En consecuencia, la interpretación que se le otorga al nicho es ambigua. En algunos casos se trata de un concepto determinístico, mientras que, en otros se intuye de manera informal alguna interpretación probabilística. Por otra parte, los métodos se limitan a representar resultados acerca de la distribución geográfica, dando poca atención en el concepto más elemental de nicho.

La motivación principal del presente trabajo fue una inquietud por explorar las propiedades de las especies en el espacio ambiental, a diferencia de prestar énfasis en su distribución geográfica. En esta tesis se presentan herramientas para la descripción y caracterización de los nichos de las especies. La descripción, conlleva el uso de técnicas de Estadística Multivariada, bajo la premisa de que el objetivo es meramente describir un subconjunto del espacio ambiental. La caracterización se originó en la búsqueda de una interpretación probabilística adecuada, donde se muestra que se puede utilizar una especificación probabilística para describir un nicho ecológico. Esto es quizás el resultado más relevante de esta tesis: que uno de los métodos diseñados para estimar nichos, puede ser explicado como el resultado de un mecanismo simple que tiene interpretación biológica directa, y no como el resultado de un complejo algoritmo computacional.

Capítulo 1

Antecedentes y conceptos biológicos.

1.1 La importancia de la distribución de las especies.

La biodiversidad comprende desde la variedad de los ecosistemas hasta las diferencias genéticas dentro de cada especie, permitiendo así, la combinación de múltiples formas de vida, cuyas interacciones fundamentan el sustento de la vida sobre el planeta. Al mismo tiempo que, los elementos que componen la naturaleza, conformados por diversos factores ambientales, aportan y aseguran muchos de los "servicios" básicos para la supervivencia de las especies.

En la naturaleza, de acuerdo a estudios realizados por los expertos en biología, las especies están conectadas de tal manera que un cambio en una de ellas tiene efectos casi inevitables en otras. Así, las interacciones entre las especies son diversas. Algunas de éstas establecen relaciones mutualistas, en la cual las especies obtienen algún beneficio. En este caso, la extinción de alguna puede provocar serios daños en el ecosistema. En cambio, otras especies establecen relaciones de tipo antagónico, lo cual significa que de la relación entre dos especies, sólo una logrará un beneficio directo. Incluso, puede suceder que la otra especie sea perjudicada a tal grado de provocar su extinción. En este caso, el efecto de no proteger esta especie provocaría una perdida en la biodiversidad.

Una forma de mantener un equilibrio entre las relaciones de las especies, es lograda mediante el conocimiento de su distribución geográfica. Cuando se conoce la región donde una especie puede vivir, es más sencillo diseñar estrategias para la conservación de las especies, y por lo tanto, de la biodiversidad. Existen distintas aplicaciones en las que se utiliza el conocimiento de la distribución geográfica de las especies. Entre ellas se encuentran el diseño de zonas de conservación para las especies en peligro de extinción, formulación de inventarios de especies, predicción de especies invasoras [20], epidemiología

y agricultura.

La determinación de la distribución de una especie se realiza a través de la estimación de su respectivo nicho ecológico, lo cual no es algo sencillo. Esto se debe a causas como la complicación que existe para determinar las ausencias de la especie, la variedad de factores ecológicos, la accesibilidad al lugar geográfico, factores intrínsecos de la especie, etc. Por otro lado, la dificultad conceptual del nicho, y la falta de información de éste en su espacio ecológico impiden avances en este campo de investigación científica.

1.2 Nicho ecológico y distribución.

1.2.1 Distribución de una especie.

Existen lugares, zonas geográficas donde se dan las condiciones apropiadas para el desarrollo, crecimiento y supervivencia de una especie. Esta región, donde la vida de la especie es viable, se denominará como la **distribución geográfica de una especie.** La distribución se puede clasificar en dos tipos, las regiones donde la especie existe actualmente, **distribución realizada**, y aquellas donde se dan las condiciones para la la vida de la especie, pero ésta no se necesariamente se encuentra en ese lugar, denominada como **distribución potencial**. El hecho de que en una región existan las condiciones para la vida de una especie y esta no se encuentre en ese lugar, puede deberse a distintas razones, tales como: evolución, competencia entre especies, accesibilidad, *etc.*

1.2.2 Nicho ecológico.

Todas las especies se encuentran en interacción directa con los factores ambientales (temperatura, humedad, presión atmosferica, *etc.*). Cada organismo posee para un determinado factor un margen de tolerancia fisiológica. Entonces, a través de las múltiples adaptaciones a diferentes rangos de tolerancia se forman áreas multidimensionales en las cuales los organismos realizan su desarrollo, se reproducen y permiten la preservación de la especie.

Actualmente, el concepto de nicho utilizado por la mayoria de los biólogos tiene sus raíces en el concepto propuesto por G. E. Hutchinson en 1957 [9]. "El **Nicho** es definido como la suma de todos los factores ambientales que actuan sobre un organismo; el nicho es una región sobre un espacio multidimensional...". Si se consideran p factores ambientales, entonces el nicho es una región sobre el espacio p-dimensional (espacio ecológico). En otras palabras, un nicho es el conjunto de características, variables ambientales o ecológicas, que describen los recursos precisos que necesita un organismo para sobrevivir. Por otro lado, el nicho no se debe considerar sólo como el espacio, sino como el subconjunto de R^p que contiene las propiedades del medio ambiente que permiten el cubrimiento de las necesidades genéticas de las especies.

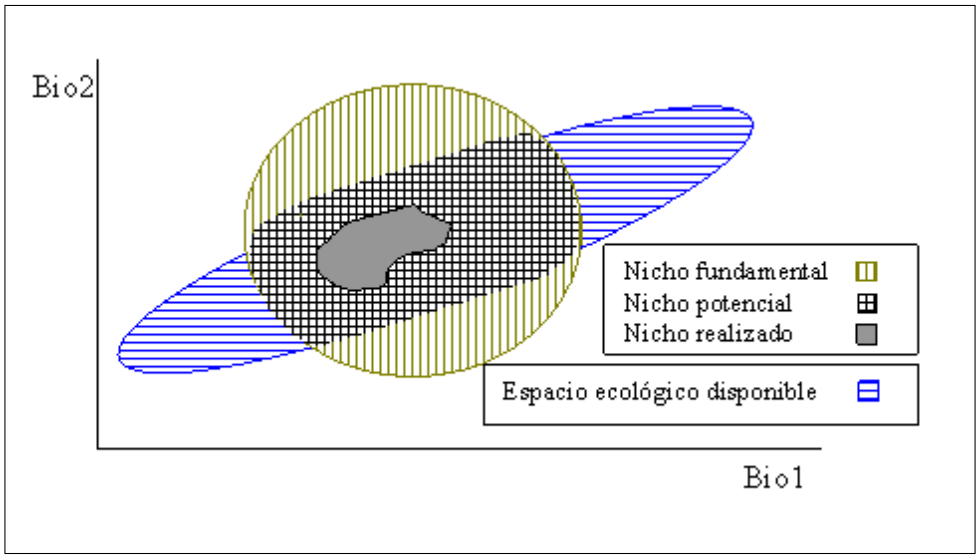


Figura 1.1: Nicho ecológico representado por dos variables ambientales.

Dos especies no pueden ocupar exactamente el mismo nicho; ya que sus requerimientos ambientales y tolerancias fisiológicas no son exactamente los mismos. Los diferentes tipos de bosques brindan hábitats marcadamente distintos y la comunidad que sostiene es diferente. Aunque diferentes especies puedan ocupar el mismo hábitat la competencia puede ser ligera o inexistente para la mayor parte de las poblaciones que conviven. Aún así el nicho considera, además de los factores que defienen un hábitat, factores que cada especie posee como qué come, dónde se anida, *etc.* En otras palabras, cuando se incluyen factores bióticos en la definición de nicho, éstos distinguen entre posibles especies con hábitats similares.

Por otro lado, el nicho ecológico es restringido por las condiciones ambientales que existen en el planeta. Además, se pueden diferenciar en el espacio ecológico regiones ambientales con presencia de aquellos ambientes que sólo son potenciales para la vida de la especie. De esta manera es como surgen las siguientes definiciones. El **nicho fundamental** es definido como las condiciones ambientales bióticas y abióticas bajo las cuales una especie tiene la capacidad de subsistir; mientras que el **nicho potencial** se define como la parte del nicho fundamental cuyos factores ambientales ocurren en el planeta. En otras palabras, el nicho potencial es la intersección del nicho fundamental con el espacio ecológico disponible sobre el planeta Tierra. Por último, el **nicho realizado** [15] es definido como la parte del espacio ecológico donde existe la especie. Equivalentemente, el nicho realizado es un subconjunto del nicho potencial donde se encuentra presente la especie. En la Figura 1.1 se

ilustran los tipos de nicho ecológico considerando dos variables ambientales.

1.2.3 Relación entre nicho y distribución.

En la Sección 1.2.1 se definió a la distribución como la zona geográfica donde una especie es viable. En dicha zona geográfica existen factores ambientales que afectan directamente a la especie. Entonces, es obvia la relación que existe entre nicho y distribución: al identificar las variables ambientales que están presentes en la región geográfica que define la distribución de una especie se está determinando su correspondiente nicho ecológico. En ambos, nicho y distribución se habla de la factibilidad de la vida de la especie; la diferencia está en los espacios: la distribución se define sobre el espacio geográfico, mientras que, el nicho en un espacio ecológico. Se denota al espacio geográfico como G ($\subset R^2$), y al espacio ambiental o ecológico como E (R^p, cuando se consideran p variables ambientales).

Una manera de realizar la conexión entre los espacios es mediante los Sistemas de Información Geográfica (SIG). Un SIG es un sistema integrado compuesto por hardware, software, personal, información espacial y procedimientos computarizados, que permite la representación de datos espaciales. En este caso, los datos espaciales refieren a información ambiental sobre el planeta.

Aunque en teoría el espacio geográfico es continuo, en la práctica se requiere de una discretización de éste para llevar al cabo las mediciones. En este caso, se crea una rejilla regular compuesta por m celdas geográficas. Una celda es un cuadro de cierta longitud de lado, y el que la rejilla sea regular significa que todas las celdas son del mismo tamaño. Por cada celda, se obtiene un vector (configuración ambiental) que contiene las mediciones de p factores ambientales que caracterizan a la región contenida por la celda. Notar que el tamaño de ésta determina la resolución de la retícula, es decir, el tamaño del lado de la celda determina de nivel de resolución de la rejilla. Por último, como consecuencia de esta rejilla se obtiene una aproximación del nicho ecológico mediante un conjunto de vectores en R^p.

A continuación se presenta una función que será de utilidad para definir, de una manera más formal, la relación entre el nicho y la distribución. Se define una función φ que va del espacio geográfico al espacio ecológico, $\varphi : G \to E$. Dicha función se utiliza cuando se desea encontrar la configuración ambiental asociada a una región geográfica. Notar que, no se puede definir la función φ^{-1}, en un sentido matemático, pues a una misma configuración ambiental le pueden corresponder distintas regiones geográficas. Esto es debido a que en el planeta pueden existir distintos lugares con características ambientales similares.

La función φ asigna a cada punto de la rejilla en G un único vector de variables ambientales en E, es decir, para $g \in G$ y $e \in E$, se tiene que $\varphi(g) = e$.

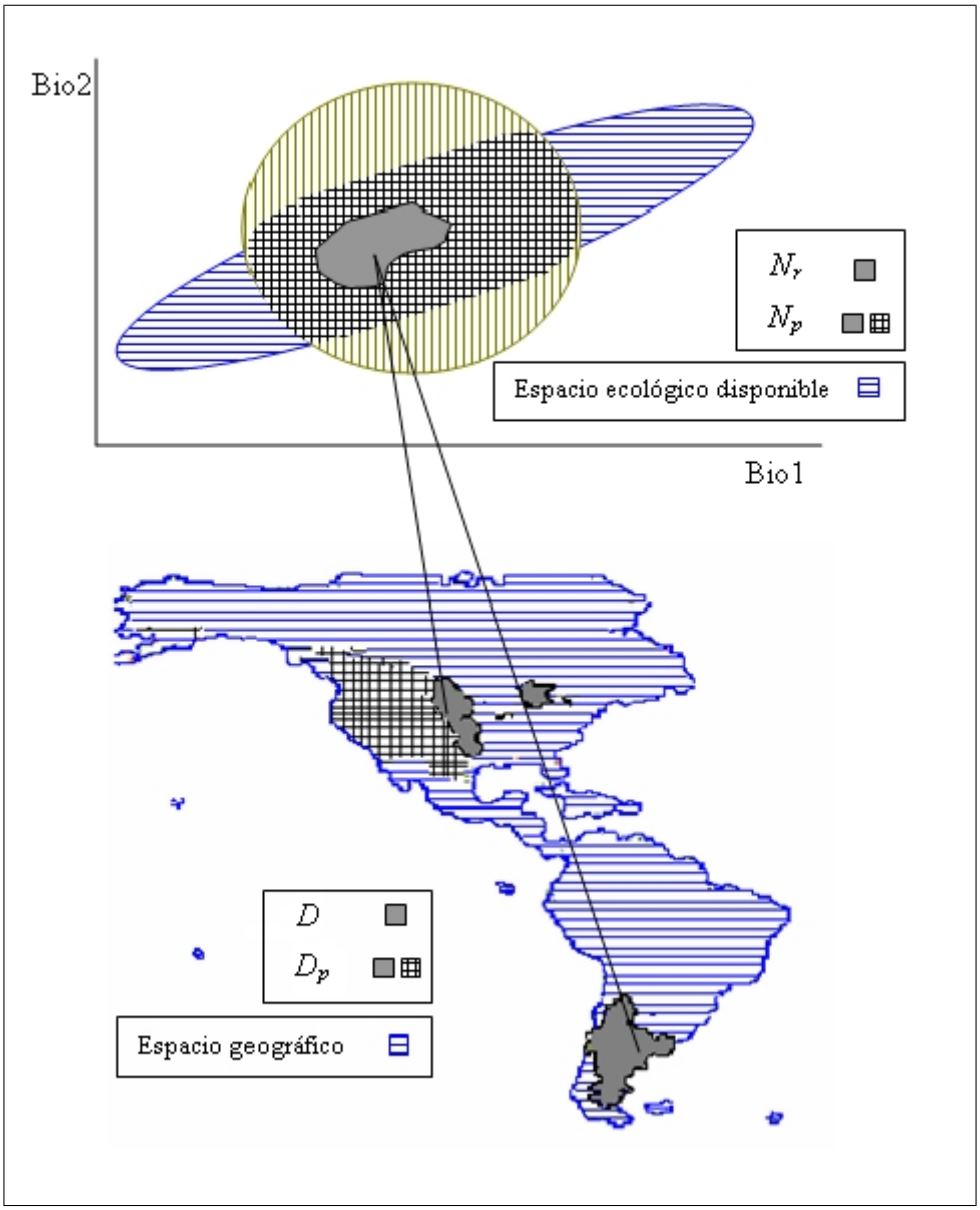


Figura 1.2: Relación entre nicho ecológico y distribución: $N_r = \varphi(D)$ (la imagen directa de la distribución es el nicho realizado); $D_p = \varphi^{-1}(N_p)$ (la imagen inversa del nicho potencial es la distribución potencial).

Para los conjuntos $A \subset G$ y $B \subset E$ se consideran las siguientes definiciones. Se define la imagen del conjunto A como $\varphi(A) = \{\varphi(a) \mid a \in A\}$. La imagen inversa de B se define como $\varphi^{-1}(B) = \{g \in G \mid \varphi(g) \in B\}$.

La distribución realizada se obtiene mediante la observación de presencias de la especie en el área geográfica. Por otro lado, el nicho fundamental (y por lo tanto el nicho potencial), se determina por las caracteristicas genéticas de la especie en el espacio ecológico. De esta manera se tiene lo siguiente: la imagen directa de φ de la distribución es el nicho realizado, mientras que, la imagen inversa del nicho potencial es la distribución potencial. En notación matemática, se definen $D(\subset G)$, $N_p(\subset E)$ como la distribución realizada y el nicho potencial, entonces, $N_r = \varphi(D)$, $D_p = \varphi^{-1}(N_p)$ son el nicho realizado y la distribución potencial de la especie, respectivamente. En la Figura 1.2 se ilustra esta relación entre el nicho y la distribución de una especie.

Rejilla y su resolución. Como se mencionó anteriomente, en el espacio geográfico se construye una rejilla donde por a cada celda se miden los factores ambientales correspondientes al lugar, creando así vectores cuyas entradas son las p variables ambientales. Esta rejilla tiene una cierta resolución, y un refinamiento en la misma proporciona mayor información, en el sentido de que se logra una mejor aproximación de la verdadera forma del nicho ecológico al obtener más configuraciones ambientales. Sin embargo, esta mejora tiene un precio, pues un refinamiento reclama mayor cantidad de mediciones, lo cual es más costoso o imposible de obtener.

1.3 Métodos para la obtención de nichos ecológicos.

1.3.1 Estimación de Nichos ecológicos.

La determinación de la distribución geográfica de una especie se lleva al cabo a través de la estimación de su respectivo nicho ecológico. Existen diversos métodos para la estimación de nichos ecológicos (MENE). Éstos tienen en común que, como ingredientes de entrada, utilizan datos reportados de presencias de la especie y las características ambientales del área de estudio. Entre los más utilizados se encuentran Bioclim [7], Domain [5], Floramap [8], GARP [16], y Maxent [17] (éste estima directamente la distribución). Otro método no tan utilizado como los anteriores es Biop [4]. Estos métodos son procedimientos que extrapolan, a partir de un conjunto de puntos, áreas que identifican a los nichos ecológicos. Finalmente, para encontrar la distribución de una especie se estima el nicho ecológico y despues con la ayuda de un SIG se obtiene la distribución asociada. A continuación se presenta una descripción más detallada de los métodos mencionados.

Bioclim. Este método se basa en un algoritmo de "envolventes bioclimáticas". La idea básica es encontrar una regla sencilla que identifique todas las áreas con características similares a las zonas de ocurrencia de la especie. A partir de los puntos ambientales generados por las presencias, se determina la media y desviación típica por cada variable. Asumiendo una distribución normal, se definen máximos y mínimos por cada variable que incluyen un percentil alto de la especie (Por ejemplo, el valor de la media menos 3 veces la desviación se utiliza como un mínimo que considera un percentil del 99%). De esta manera se genera un hipercubo en el espacio ambiental, y se define el nicho como el conjunto de ambientes contenidos dentro de este hipercubo. Una desventaja del método es que no toma en cuenta las posibles relaciones entre las variables.

Domain se basa en una medida de disimilaridad entre ambientes. Su idea intuitiva es que compara el ambiente en un sitio arbitrario con el conjunto de ambientes de los sitios de presencia usando cierta métrica. El valor de disimilaridad puede representarse directamente en un mapa en una escala continua, y para especificar un nicho se consideran aquellos puntos que no son disimilares más allá de un valor umbral arbitrario.

Floramap es un sistema basado en el cálculo de la probabilidad de que un ambiente arbitrario pertenezca a una distribución normal multivariada descrita por los ambientes con presencia de la especie. El método utiliza técnicas de componentes principales para construir una regla de clasificación para cada ambiente. Tiene algunas suposiciones distribucionales en su justificación, en el sentido de que se pretende que las reglas de clasificación sean óptimas. El punto principal es que para identificar un nicho, Floramap realiza una clasificación nodo por nodo, tras calcular los componentes principales.

GARP(Genetic Algorithm for Rule-set Production). Un algoritmo genético es un método para buscar la solución a un problema que asemeja dicha busqueda a la evolución biológica. GARP genera una serie de reglas: en la primera iteración genera la primera regla y evalúa los errores de omisión y comisión del modelo, en las siguientes iteraciones genera más reglas que son inluidas o excluidas del conjunto según el grado de ajuste del modelo. Finalmente, el algoritmo se para cuando ya no se pueden crear mejores modelos o se alcanza el número máximo de iteraciones. Estas técnicas computacionales intensivas generan interpolaciones las presencias de la especie como función de las variables ambientales. GARP da como resultado una predicción distinta en cada corrida, por lo que en su implementación se utilizan distintos criterios para conseguir un nicho. Entre los más usados se encuentra el "Best Sub-sets" [2], el cual consiste en llevar a cabo un cierto número de corridas seleccionando las

más importantes. De esta manera se realiza la suma de aquellas corridas que caen dentro de la categoria de "Best Sub-sets".

Maxent trabaja directamente sobre el espacio geográfico y utilizando, además de las presencias, funciones de las características ambientales. Dichas funciones ambientales son definidas sobre este espacio geográfico. Se trata de un método genérico para hacer estimaciones o inferencias a partir de información incompleta. La idea de Maxent es estimar la distribución objetivo, distribución potencial de la especie, encontrando la distribución de máxima entropía (la más cercana a la uniforme), sujeta a la restricción de que el valor esperado de cada función del ambiente bajo esta distribución estimada es igual al esperado de la función bajo la distribución empírica. Un supuesto del método es que la información de presencias es tomada de manera independiente de acuerdo a una distribución de probabilidad, lo cual no es totalmente cierto debido a que existe sesgo en el muestreo.

BioP determina un mapa de establecimiento potencial de una especie de interés, mediante un enfoque bayesiano. Para encontrar la probabilidad de presencia formula un modelo que postula que un registro de presencia es producto de tres factores a la vez: la presencia de la especie en la celda geográfica, la visita a la celda y la detección de la especie. Aunque BioP no se encuentra entre los más utilizados se pueden mencionar algunas ventajas sobre los otros métodos: produce una medida que califica la certidumbre que se tiene en el resultado que proporciona, toma en cuenta el sesgo espacial que existe al momento de tomar las mediciones, y utiliza de manera transparente el conocimiento de un experto a priori.

1.3.2 Salidas de los métodos de estimación de nichos.

La salidas de los métodos son mapas con la localización de los nichos ecológicos. Dichos mapas, dependiendo del método, están basados en diferentes nociones como pueden ser: aseveraciones binarias, medidas de similitud, distancias o proporciones. Algunos de éstos métodos, independientemente de su noción, convierten sus salidas en aseveraciones binarias, dividiendo la región en dos partes: nicho y no nicho. En general, los resultados obtenidos no siempre tienen interpretaciones transparentes. Esto es particularmente cierto de GARP, por su estructura computacional intrínseca.

Como se vió anteriormente, los MENE son procedimientos que extrapolan, a partir de una nube de puntos, áreas que identifican a los nichos ecológicos, Maxent lo hace directamente en el espacio geográfico. Estos procedimientos se componen primordialmente de modelos de tipo empírico que, por la complejidad de los factores mencionados en la primera sección, no incorporan en su

definición estructura biológica explícita o bien la simplifican. Esta naturaleza empírica de las soluciones provoca que no se tenga una interpretación clara de los resultados de cada método, ni de su significado. Existe una discusión entre los expertos en biología, en cuanto a la naturaleza de las soluciones de los MENE; algunos afirman que se obtiene como resultado el nicho realizado, mientras otros defienden que se trata del nicho fundamental. En lo que se refiere a esta interpretación, de aquí en adelante en el presente trabajo, se entenderá que se trata del nicho potencial.

1.4 Espacios y objetivo de la tesis.

1.4.1 Espacio geográfico vs espacio ecológico.

A partir de la segunda sección se distinguieron dos escenarios: espacio geográfico (G) y espacio ecológico (E). El primero, representado por coordenadas longitud-latitud, mientras que, el segundo es conformado por configuraciones ambientales. Es obvio, que estos espacios son distintos, y por lo tanto, ofrecen diferentes resultados cuando se estudian a las especies.

Para ilustrar, considerar dos especies cuyas distribuciones se encuentran en Alaska y la Patagonia, respectivamente, y por otro lado, otras dos especies, diferentes a las primeras, cuyas distribuciones se encuentran en distintos lugares de México. En el primer caso, las especies se encuentran muy distantes entre sí, mientras que, en el segundo, las especies se encuentran cercanas. Ahora, si se le pone especial atención a los nichos de las especies anteriores y se estableciera una medida de cercanía en E, probablemente se llegaría a la conclusión de que dicha medida es más pequeña para las primeras especies que para las segundas. Quizá puede deberse a que México es uno de los países con mayor diversidad en el mundo, y por lo tanto, su ambiente puede variar mucho en distintas regiones, y que en los polos del planeta los ambientes posiblemente son muy parecidos.

Por otra parte, una medida de volumen en G de una región geográfica pequeña, pero con gran diversidad biológica, deberá ser chica, mientras que, el volumen (asociado a la región) en E puede ser muy grande. Por el contrario, posiblemente existen grandes regiones con poca biodiversidad. Dichas regiones tendrán asociado un volumen grande en G y un volumen pequeño en E.

Los ejemplos anteriores ilustran que existen diferencias en los espacios, incluso que se pueden llegar a conclusiones contradictorias al trabajar en uno u otro espacio. Por otro lado, en E se más sencillo identificar las condiciones que satisfacen las necesidades genéticas de cada especie. Entonces, establecer medidas que caractericen a las especies en G provocaría perdida de información que se encuentra en los factores ambientales. De este modo, parece adecuado recomendar que cualquier tipo de estudio sobre las especies deba basarse en

el espacio ecológico.

Cabe notar que en el espacio ecológico, por su dimensión, es más difícil encontrar las relaciones entre las especies. Generalmente, este espacio no es representable gráficamente de forma directa, ya que el número de variables utilizadas para construir una configuración ambiental suele ser mayor que tres. Además, la falta de conceptos para trabajar en dimensiones altas provoca que sea más complicado el establecer una medida razonable de cercanía entre las especies.

El presente trabajo pretende describir y formalizar las relaciones básicas entre distintos subconjuntos del espacio ecológico, en esencia, se describirán distintas especies del espacio geográfico, mediante sus correspondientes imágenes en el espacio ecológico.

1.4.2 Objetivos: descripción bajo dos distintos enfoques.

Como se mencionó en la sección anterior, en general los resultados que se obtienen con los MENE no siempre tienen interpretaciones transparentes. La mayor parte de los MENE tienen poca formalidad matemática o consisten de complejos algoritmos empíricos. En consecuencia, la interpretación que se le otorga al nicho, es por consiguiente ambigua. En algunos casos se trata de un concepto determinístico, mientras que, en otros se intuye de manera informal alguna interpretación probabilística. Por otro lado, hoy en dia, no existen definiciones de medidas del nicho de Hutchinson, ni teoría explícita sobre la dimensión de estos conjuntos de configuraciones ambientales, y distancias entre ellos [21]. Es necesario definir conceptos más formales, cuantitativos y operacionales para explicar complicaciones como el significado de los distintos tipos de nichos (fundamental, potencial y realizado). Quizás, un MENE pueda ser explicado como el resultado de un mecanismo simple, que tiene interpretación biológica, y no como el resultado de un complejo algoritmo tal como muchos MENE toman en cuenta. Lo cierto es que, la falta de conceptos formales en la teoría de nichos impedirá posibles avances en este campo de investigación.

El objetivo del presente trabajo es proporcionar herramientas formales para la descripción de los nichos obtenidos por los MENE. Dicho objetivo se puede ver desde dos puntos de vista distintos equivalentemente a subdividirlo en dos objetivos contrastantes entre sí. El primer objetivo se origina por una inquietud [21] por explorar las propiedades de las especies en el espacio ecológico. Dicho objetivo es llevado al cabo con la identificación de técnicas de Estadística Multivariada para describir conjuntos de configuraciones ambientales en R^p. Lo que se desea con este primer objetivo es compilar un catalogo de herramientas para medir, representar y comparar nichos ecológicos. Dichas herramientas se presentan en el Capítulo 2. Posteriormente, surge una nueva inquietud la cual consiste en conocer el significado de los conjuntos que se obtienen con los MENE. De esta manera, en un segundo objetivo se desea

G		E				Especie			
c_1	c_2	Bio1	Bio2	...	Bio9	pust	nigr	...	abei
$v_{1,1}$	$v_{1,2}$	$x_{1,1}$	$x_{1,2}$	...	$x_{1,9}$	$y_{1,1}$	$y_{1,2}$	...	$y_{1,8}$
$v_{2,1}$	$v_{2,2}$	$x_{2,1}$	$x_{2,2}$	...	$x_{2,9}$	$y_{2,1}$	$y_{2,2}$	...	$y_{2,8}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$		$\vdots$
$v_{m,1}$	$v_{m,2}$	$x_{m,1}$	$x_{m,2}$	...	$x_{m,9}$	$y_{m,1}$	$y_{m,2}$	...	$y_{m,8}$

Tabla 1.1: Nichos estimados por GARP, donde, m es el número de celdas (número de configuraciones ambientales); $v_{i,j}$ es el valor de la coordenada j sobre la celda i, donde $i = 1, 2, ..., m$ y $j = 1, 2$; $x_{i,j}$ es la medición de la variable ambiental j de la configuración ambiental i, donde $i = 1, 2, ..., m$ y $j = 1, 2, ...9$; $y_{i,j}$ es 1 si la configuración i pertenece al nicho de la especie j, donde $i = 1, 2, ..., m$ y $j = 1, 2, ...8$.

proveer de una interpretación probabilística de los resultados de los MENE. Dicho objetivo se realiza con la determinación de una especificación probabilística, a través de una idealización del comportamiento de la especie en el espacio ecológico completo. Aquí, lo que se obtiene es un objeto probabilístico denominado como "función de preferencia de la especie" el cual caracteriza a los nichos ecológicos. Los conceptos propuestos y una métodología encaminada a este segundo objetivo son presentados en el Capítulo 3.

Para ilustrar las descripciones mediante los dos enfoques anteriores se utilizarán los datos proporcionados por el Dr. Jorge Soberón que corresponden a 8 especies de aves (orioles), pertenecientes a la familia ictearidae, con 9 variables ambientales *bioclimáticas* (BIO1, BIO4, BIO5, BIO6, BIO7, BIO12, BIO13 BIO14, BIO15). Entonces, se trabajará con 8 nichos que se encuentran en un espacio 9-dimensional. Dichos nichos corresponden a las especies: *icterus pustulatus, icterus nigrogularis, icterus leucopteryx, icterus gularis, icterus galbula, icterus bullockii, icterus auratus e icterus abeillei.* Estos datos son salidas binarias en el Continente Americano obtenidas utilizando el método GARP. En la Tabla 1.1 se ilustra la forma de las salidas de este método.

Las *variables ambientales bioclimáticas* son derivadas de los valores de la temperatura y precipitación mensual a fin de generar variables más significativas biológicamente. La descripción de estas variables se presenta en la Tabla 1.2. Es importante comentar que se trata de variables que miden el clima y no el estado del tiempo. Con frecuencia se confunde el estado del tiempo con el clima de un lugar. El tiempo atmosférico a una hora fija es determinado por la temperatura, presión atmosférica, humedad, *etc.*, registrados en el instante que se considera. Se entiende que el estado de tiempo cambia constantemente. Por otro lado, puede decirse que dos lugares geográficamente distantes tienen el mismo tiempo en un momento dado (por ejemplo, un día con lluvia en

Variable	descripción
BIO1	temperatura media anual
BIO2	medias mensuales (máxima temperatura - mínima temperatura)
BIO3	isothermality
BIO4	temperatura estacional (desviación estándar $\times$ 100)
BIO5	máxima temperatura del mes más cálido
BIO6	mínima temperatura del mes más frío
BIO7	rango de temperatura anual
BIO8	temperatura media del trimestre más húmedo
BIO9	temperatura media del trimestre más seco
BIO10	temperatura media del trimestre más caluroso
BIO11	temperatura media del trimestre más frío
BIO12	precipitación anual
BIO13	precipitación del mes más húmedo
BIO14	precipitación del mes más seco
BIO15	precipitación estacional (coeficiente de variación)
BIO16	precipitación del trimestre más húmedo
BIO17	precipitación del trimestre más seco
BIO18	precipitación del trimestre más caluroso
BIO19	precipitación del trimestre más frío

Tabla 1.2: Variables ambientales bioclimáticas

ambos sitios). Sin embargo, es evidente que estos lugares no necesariamente tienen el mismo clima. Prueba de lo anterior es la diferente vegetación que existe en distintos lugares. Entonces, el tiempo se traduce en algo instantáneo y que cambia constantemente; mientras que el clima, aunque refiere a los mismos fenómenos, los traduce a una dimensión más permanente, duradera y estable. El clima se puede ver como una sucesión periódica de tipos de tiempo.

Capítulo 2

Métodos descriptivos aplicados a nichos ecológicos.

Como se mencionó en la Sección 1.4.2, la primera forma de abordar la descripción de nichos ecológicos se realiza a través de la teoría de Estadística Multivariada. Al describir los nichos ecológicos como conjuntos de puntos en un espacio p-dimensional, se pueden distinguir dos objetivos distintos. En el primero de ellos, se quiere extraer información de un solo nicho: tomar una especie y describir su nicho mediante técnicas de Estadística Descriptiva. El segundo objetivo consiste de realizar comparaciones entre dos nichos: tomar los nichos correspondientes a dos especies y medir su similaridad utilizando alguna distancia.

En la Sección 1.2.3 se describió la discretización del espacio geográfico mediante una rejilla compuesta por m celdas. También se mencionó que, una configuración ambiental es la colección de los valores de las p variables ambientales tomadas de una misma celda. Si n $(\leq m)$ configuraciones corresponden al conjunto que constituye el nicho ecológico de una especie, entonces el conjunto completo de datos puede ser colocado en una matriz $n \times p$, la cual está dada por

$$X_{n\times p} = \{x_{ij}\} = \begin{bmatrix} x_1^t \\ x_2^t \\ \vdots \\ x_n^t \end{bmatrix} \begin{array}{c} \text{primera configuración ambiental} \\ \text{segunda configuración ambiental} \\ \vdots \\ \text{última configuración ambiental,} \end{array} \tag{2.1}$$

donde cada renglón de X es un vector diferente en R^p que representa una configuración ambiental.

2.1 Descripción de un solo nicho ecológico.

2.1.1 Estadísticas Descriptivas.

Aprender más sobre las especies implica la descripción parcial o total de su respectivo nicho ecológico. Se tendrían pocas dificultades si existieran medidas descriptivas apropiadas y representativas, pero como se mencionó en el Capítulo 1 esto no es así. La descripción debe ser tal, que el conocimiento de ciertas medidas permita tener una apreciación clara del nicho. Entonces, en la presente sección se definen cantidades que son medidas numéricas descriptivas de un conjunto de valores en R^p: medidas de centralización, dispersión, asimetría, curtosis y volumen. El conjunto de datos es un conjunto de configuraciones ambientales que constituyen el nicho ecológico.

Media.

Una primera manera de resumir la información contenida en el nicho ecológico es la extensión de la noción univariada de media hacia un promedio multivariado:

$$\overline{x} = \begin{bmatrix} \overline{x}_{*\,1} \\ \vdots \\ \overline{x}_{*\,p} \end{bmatrix},$$

donde $\overline{x}_{*\,i} = \frac{1}{n}\sum_{r=1}^{n} x_{ri}$; $i = 1, 2, ..., p$. Este promedio multivariado, que contiene las medias univariadas, representa el centro de gravedad o geométrico del conjunto de configuraciones que definen el nicho de la especie. La Tabla 2.1 presenta los valores de las medias multivariadas para las ocho especies de orioles.

La media multivariada del nicho solamente localiza el centro de la distribución de las configuraciones ambientales; por sí misma, no ofrece una descripción adecuada del conjunto de configuraciones. Dos conjuntos podrían estar distribuidos con diferente forma pero tener la misma media. La diferencia entre las distribuciones puede estar en la variación o dispersión de las configuraciones con respecto a esta media, por lo que obviamente es necesario considerar medidas complementarias.

Variables	pust	nigr2	leuc	gula	galb	bull	aura	abei
Bio7	194	130	117	181	312	359	177	237
Bio6	130	198	138	145	−17	−66	165	40
Bio5	324	328	254	326	295	293	342	277
Bio4	1785	589	1290	1634	6333	7513	1990	2307
Bio15	85	62	34	71	48	46	61	91
Bio14	6	34	89	17	33	16	27	6
Bio13	235	319	256	251	147	73	211	155
Bio12	1198	1975	1946	1435	999	475	1236	705
Bio1	231	260	193	239	141	103	258	164

Tabla 2.1: Media de nicho ecológico.

Medidas de dispersión.

Una medida de dispersión multivariada que sea sensata debe incluir, además de la dispersión de cada variable ambiental, la relación que pueda existir entre dos variables. Un modo de lograr esto es a través de la covarianza, la cual mide la dependencia lineal entre éstas. Entonces, una generalización de la varianza en un espacio p-dimensional está dada por la matriz de varianzas y covarianzas:

$$S_{pxp} = \{s_{ij}\},$$

donde $s_{ij} = \frac{1}{n}\sum_{r=1}^{n}(x_{ri} - \overline{x}_i)(x_{rj} - \overline{x}_j)$.

La matriz S es una generalización de la noción varianza en una dimensión, como medida de dispersión sobre la media. Sin embargo, es conveniente contar con un solo número que resuma la información contenida en S, como medida de dispersión. es usual proponer para esto dos medidas: la *Varianza generalizada*, $\det(S)$; y la *Variación total*, $\mathrm{tr}(S)$.

Para ambas medidas, valores grandes representan un alto grado de dispersión alrededor de $\overline{x}$ y por el contrario, valores bajos representan concentración alrededor de $\overline{x}$. Sin embargo, cada medida refleja diferentes aspectos de la variabilidad en los datos. En particular, la varianza generalizada juega un papel importante en la Estimación por Máxima Verosimilitud y la Variación total es un concepto usado en Análisis de Componentes Principales.

Variación total. La variación total se denota como la suma de los elementos de la diagonal de la matriz S, es decir, $\mathrm{tr}(S) = \sum_{j=1}^{n} s_{jj}$. Observar que, la variación total no toma en cuenta la orientación o estructura de correlación de los datos. Por lo tanto, la variación total asigna el mismo valor a dos distintos conjuntos con idénticas varianzas, pero covarianzas distintas.

Varianza generalizada. Para entender mejor a la varianza generalizada se utilizará una representación alternativa de la matriz de datos X. La representación puede construirse considerando los datos como p vectores en un espacio de dimensión n, es decir, utilizar las variables ambientales como si fueran vectores:

$$X_{n\times p} = \{x_{ij}\} = \begin{array}{c} \begin{array}{cccc} \text{var1} & \text{var2} & \dots & \text{varp} \end{array} \\ \left[\begin{array}{cccc} y_1 & y_2 & \dots & y_p \end{array}\right] \end{array}.$$

Entonces, las coordenadas del primer punto $y_i^t = [x_{1i}, x_{2i}, \dots, x_{ni}]$, son las n mediciones de la i-ésima variable ambiental.

Con el objetivo de facilitar la interpretación de la varianza generalizada se supondrá que las variables son centradas, es decir, las p variables tienen media cero. De esta manera el producto punto entre dos vectores y_i y y_k es proporcional a la covarianza entre las variables i y k, es decir, $y_i^t y_k = \sum_{j=1}^{n}(x_{ji})(x_{jk}) = ns_{ik}$, porque $\overline{x}_k = \overline{x}_i = 0$. Por otro lado, si θ_{ik} es el ángulo formado por los vectores y_i y y_k, entonces el coseno del ángulo θ_{ik} es el coeficiente de correlación entre las variables i y k. Este último se expresa como: $r_{ik} = s_{ik}/(\sqrt{s_{ii}}\sqrt{s_{kk}}) = \cos(\theta_{ik})$. Por lo tanto, si las variables tienen poca correlación el ángulo entre sus vectores será cercano a 90^o.

Finalmente, se puede demostrar que el determinante de la matriz S es proporcional al volumen del paralelepípedo generado por los vectores y_i, $i = 1, 2, \dots, p$ [10]. En este caso, cuando se tengan dos conjuntos con idénticas varianzas el determinante de la matriz S será más pequeño para el conjunto cuyas variables se encuentren más correlacionadas. En otras palabras, cuando se tienen dos conjuntos con la misma variación total, entonces, la varianza generalizada será más pequeña para el conjunto cuyas variables presenten alguna estructura de correlación.

La Tabla 2.2 presenta los valores de la varianza generalizada ($\det(S)$) y variación total ($\text{tr}(S)$) para las ocho especies de orioles. Se observa que el valor más grande en variación total le corresponde a ictegalb. Por otro lado, aunque esta especie tenga más configuraciones ambientales ocupa un tercer lugar en varianza generalizada seguida de ictebull e ictegula. En este caso, existe cierta estructura de correlación en las configuraciones que definien el nicho de galbula que provocan que el valor de la varianza generalizada sea más pequeño en comparación con las especies ictebull e ictegula. Esto se puede ver en las gráficas de las primeras tres componentes principales en la Figura C.1; la mayor parte de los configuraciones de ictegalb siguen la forma de un plátano. Ahora, se observa que aunque icteaura, icteabei, e icteleuc poseen una pequeña cantidad de configuraciones, esta última ocupa un cuarto lugar en varianza generalizada. En la gráfica de los componentes principales de estos nichos en la Figura C.2 se puede observar que la especie icteleuc tiene mucha dispersión. Lo anterior ilustra que, aunque ambas medidas se encuentran

Especie	n	Determinante	Traza
pust	1478	1.84×10^{13}	2.32×10^{06}
nigr2	1385	1.01×10^{12}	3.73×10^{05}
leuc	63	4.19×10^{13}	1.12×10^{06}
gula	2487	1.33×10^{14}	1.90×10^{06}
galb	5706	1.07×10^{14}	1.49×10^{07}
bull	2531	2.71×10^{14}	3.61×10^{06}
aura	78	2.71×10^{06}	6.08×10^{04}
abei	99	6.51×10^{09}	6.37×10^{05}

Tabla 2.2: Medidas de dispersión de nicho ecológico.

muy relacionadas, la estructura de correlación importa al momento de medir dispersión. Además, se puede concluir que la cantidad de configuraciones que forman los conjuntos no tienen ninguna relación obvia con alguna de las medidas de dispersión.

Por otro lado, una medida de dispersión o volumen utilizada en la literatura biólogica es la suma de las distancias euclidianas al cuadrado a pares entre las configuraciones ambientales [21]. En el apendice B se demuestra que esta cantidad dividida entre n es igual al producto de n por la traza de la matriz de varianzas. Además se muestra que ambos valores son equivalentes a la suma de las distancias euclidianas al cuadrado de las configuraciones con respecto a su media multivariada. A partir de lo anterior, se tiene que es posible utilizar cualquiera de estas tres cantidades para medir dispersión. Debe tenerse presente que, ninguna de estas medidas involucra la relación que pueda existir entre las variables ambientales.

Coefiente de Simetría y coeficiente de Curtosis.

Aunque, las medidas de dispersión presentadas ofrecen una mejor descripción del nicho, es conveniente utilizar otro tipo de medidas para especificar de una manera más completa las características de cada conjunto de configuraciones ambientales. Las primeras medidas más utilizadas, después de la media y varianza, para describir un conjunto de datos son asimetría y curtosis. La generalización de los coeficientes de asimetría y curtosis al caso multivariante, no es inmediata. Una de las propuestas más utilizadas es debida a Mardia [12], quien formuló los siguientes coeficientes:

$$A_s = \frac{1}{n^2} \sum\nolimits_{r,s=1}^{n} g_{rs}^3, \tag{2.2}$$
$$K = \frac{1}{n} \sum\nolimits_{r=1}^{n} g_{rr}^2;$$

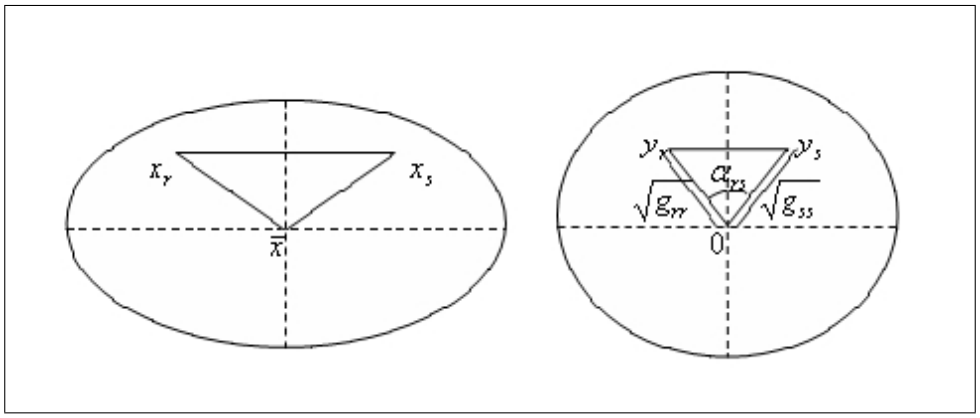


Figura 2.1: Ángulos y distancias de Mahalanobis. Espacio normal (izquierda), y puntos en el espacio transformado (derecha).

donde $g_{rs} = (x_r - \overline{x})' S^{-1}(x_s - \overline{x})$. Estas medidas están relacionadas con las distancias y ángulos de Mahalanobis [13], los cuales proveen de una interpretación geométrica al coeficiente de asimetría.

Se denota como d_{rs}^2 a la distancia de Mahalanobis entre x_r y x_s; además, notar que g_{rr} es la distancia de Mahalanobis entre x_r y $\overline{x}$. Entonces, la relación entre g_{rs} y d_{rs}^2 está dada por $g_{rs} = (1/2)(g_{rr} + g_{ss} - d_{rs}^2)$, donde $d_{rs}^2 = (x_r - x_s)' S^{-1}(x_r - x_s)$. Ahora, se define α_{rs} como el ángulo de Mahalanobis entre los vectores $(x_r - \overline{x})$ y $(x_s - \overline{x})$, como aquel que satisface

$$\cos(\alpha_{rs}) = \frac{g_{rs}}{\sqrt{g_{rr}}\sqrt{g_{ss}}}. \tag{2.3}$$

La distancia de Mahalanobis corresponde a la distancia euclidiana de los vectores en el espacio transformado $y = S^{-\frac{1}{2}}(x - \overline{x})$. A su vez, el ángulo de Mahalanobis también corresponde a el ángulo entre estos vectores transformados. En la Figura 2.1 se representan, el espacio normal por una elipse y el espacio transformado con un círculo. Es posible expresar A_s en términos del ángulo y las distancias de Mahalanobis, despejando g_{rs} en la expresión (2.3) y al sustituirlo en (2.2) se obtiene

$$A_s = \frac{1}{n^2} \sum\nolimits_{r,s=1}^{n} \left(\sqrt{g_{rr}}\sqrt{g_{ss}}\cos(\alpha_{rs})\right)^3.$$

Entonces, si las configuraciones ambientales son uniformemente distribuidas en una esfera de dimensión p, se tiene que $A_s \simeq 0$. En general, si las configuraciones se encuentran uniformemente distribuidas en una elipse en un espacio p-dimensional, se tendrá que $A_s \simeq 0$. Entre más asimétrico sea el conjunto de configuraciones mayor será el valor de A_s. Por otro lado, el estadístico

Especie	n	Asimetría	Curtosis
pust	1478	7.12×10^{07}	1.79×10^{05}
nigr2	1385	6.53×10^{08}	7.66×10^{05}
leuc	63	4.69×10^{06}	3.07×10^{04}
gula	2487	6.03×10^{07}	1.60×10^{05}
galb	5706	5.80×10^{06}	3.52×10^{04}
bull	2531	8.88×10^{05}	1.08×10^{04}
aura	78	6.41×10^{11}	7.45×10^{07}
abei	99	1.42×10^{08}	2.81×10^{05}

Tabla 2.3: Coeficientes de asimetría y curtosis de nicho ecológico.

K es una medida de lo puntiagudo de una distribución de probabilidad. En este caso, K es una medida de lo puntiagudo del histograma multivariado construido a partir de las configuraciones que definen el nicho ecológico.

En la Tabla 2.3 se presentan los valores de los estadísticos de asimetría y curtosis calculados a las ocho especies de orioles. Se tiene que las especies más asimétricas son ictenigr2 e icteaura. En la Figura C.3 de los primeros componentes principales de estas especies se observa que poseen cierta orientación como si las configuraciones ambientales siguieran a una línea en tres dimensiones. Por el contrario, la especie con menor asimetría es ictebull; en la Figura C.4 se obseva una forma más esférica de ictebull en comparación con los nichos de las otras especies. Por otro lado, el valor de curtosis más pequeño es otorgado a icteaura. Esto es porque sus configuraciones se encuentran muy concentradas, lo cual se observa en la Figura C.4, y por lo tanto su distribución es más puntiaguda en comparación con las otras especies. Por el contrario, la especies icteleuc, ictegalb e ictebull se encuentran muy dispersas. Esto provoca que la distribución de estas especies se extienda sobre el espacio induciendo un coeficiente de curtosis más pequeño, y por lo tanto, una distribución menos puntiaguda, esto también se puede observar en las Figuras C.1 y C.4. Esto muestra que es recomendable calcular el índice de curtosis sólo cuando las primeras medidas descriptivas son muy similares; de lo contrario, la nueva información no será más que la que se tenía con las primeras medidas.

Volumen de Elipsoide.

En la presente subsección se propone un índice que representa una medida de volumen del nicho ecológico. Este índice se construye utilizando los conceptos de distancia de Mahalanobis y la varianza generalizada. La medida volumen de un conjunto de configuraciones ambientales está dada por

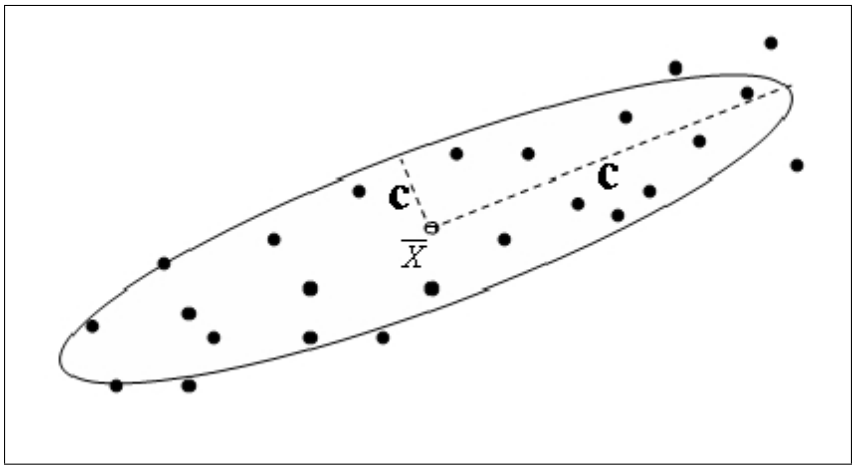


Figura 2.2: Elipsoide a un nivel c con dirección dada por la matriz de correlación S.

$$\begin{aligned} \nabla_c &= \text{vol}\left\{x|\ (x-\overline{x})^t S^{-1}(x-\overline{x}) \leq c\right\} \\ &= c^{\frac{p}{2}}\kappa_p \det(S), \end{aligned} \tag{2.4}$$

donde c es la mediana de las distancias de Mahalanobis de las configuraciones ambientales con respecto a su centroide,

$$c = \text{mediana}\left\{d|\ d = (x-\overline{x})^t S^{-1}(x-\overline{x}) \text{ y } x \in \text{Nicho}\right\}$$

y κ_p es una constante que sólo depende de p.

La idea intuitiva del índice es presentar el volumen de un elipsoide que sigue una orientación según la variabilidad de los datos (estructura de correlación de las variables ambientales). Este elipsoide está centrado en la media multivariada del conjunto, y su tamaño se define en función de las distancias de Mahalanobis de las configuraciones a dicha media multivariada. En la Figura 2.2 se ilustra una representación de un elipsoide, cuya orientación sigue la dirección de la correlación de las variables ambientales. Las configuraciones ambientales que se encuentran dentro de este elipsoide se encuentran a una distancia de Mahalanobis menor que c.

Quizá otros valores adecuados para c, en la expresión (2.4), sean la media de las distancias de Mahalanobis, $\overline{c} = \frac{1}{n}\sum_x (x-\overline{x})^t S^{-1}(x-\overline{x})$ ó el máximo de estas distancias, $c_m = \max\left\{d \mid d = (x-\overline{x})^t S^{-1}(x-\overline{x}) \text{ y } x \in \text{N}\right\}$. Es cierto que el máximo de las distancias de Mahalanobis consideraría el volumen del mínimo elipsoide que contiene a todas las configuraciones ambientales. Sin

Especie	n	Volmen
pust	1478	7.26×10^{16}
nigr2	1385	4.43×10^{15}
leuc	63	2.77×10^{17}
gula	2487	4.39×10^{17}
galb	5706	3.76×10^{17}
bull	2531	1.25×10^{18}
aura	78	1.94×10^{10}
abei	99	4.54×10^{13}

Tabla 2.4: Volumen de nicho ecológico.

embargo, esta medida no sería muy razonable, debido a las distintas formas que puede tomar un nicho. Cuando se tengan conjuntos con forma de plátano, ó conjuntos con configuraciones demasiado alejadas de la media multivariada, se podría exagerar el valor real de su volumen. Por otro lado, utilizar la media de las distancias de Mahalanobis es equivalente a utilizar sólo la varianza generalizada. La razón de lo anterior es porque dicha media siempre es igual p, y por lo tanto $\nabla_{\overline{c}}$ es proporcional a la varianza generalizada. Esto provee de una interpretación adicional de la medida de dispersión.

En la Tabla 2.4 se presentan los valores del volumen elipsoide obtenido con la mediana de las distancias de Mahalanobis. Se observa que icteleuc, ictegula e ictegalb poseen casi el mismo volumen en comparación con las otras especies. Esto parece razonable al observar los nichos en la Figura C.5. Por otro lado, observando la especie icteaura en la Figura C.4, parece obvio que la especie con menor volumen debería ser esta.

A lo largo de esta sección se realizaron propuestas de medidas muméricas descriptivas para el conjunto de configuraciones ambientales que definen el nicho de una especie. Además, se ilustraron dichas medidas con un conjunto de datos de especies de orioles. En la siguiente sección se presentan métodos gráficos para representar al nicho ecológico. Dichos métodos pretenden continuar y mejorar la descripción de estos conjuntos de configuraciones.

2.1.2 Representaciones gráficas de nicho ecológico.

Un problema elemental en la descripción es representar visualmente los elementos (configuraciones ambientales) del nicho ecológico, ya que éste se encuentra en un espacio multidimensional. Para la solución de este tipo de problemas muchos medios gráficos han sido propuestos. Entre los más famosos se encuentran las representaciones con estrellas, árboles o castillos [11], caritas [6] y curvas [3]. Estas técnicas representan cada punto en R^p mediante una estrella, árbol, castillo, carita o curva en dos dimensiones. Esta última tiene

como ventaja sobre las demás que puede representar a más de un elemento del conjunto en la misma figura, es decir, se pueden sobreponer varias curvas en una misma gráfica. Por este motivo sólo se proponen a las curvas de Andrews como un método razonable para la representación de un nicho ecológico.

En esta sección también serán expuestos otros dos métodos de graficación útiles para ilustrar estructuras o relaciones multivariadas entre los elementos de un nicho. El primero es un histograma construido a través del concepto de distancia de Mahalanobis. El segundo es un método de aglomeración que se propone como herramienta de representación gráfica. Además, se estudiarán las posibles transformaciones de las variables que conduzcan a una descripción más simple del conjunto. Ciertamente, los resultados de estos métodos dependerán únicamente de las estructuras subyacentes en el conjunto de configuraciones ambientales; y no de presuposiciones de alguna estructura de clasificación o de algún modelo.

Transformaciones de los datos. Las transformaciones comúnmente utilizadas en Estadística se pueden clasificar en tres tipos. La primera es la *estandarización por variable*: restar la media de cada variable y dividir entre su respectiva desviación. En este caso, la matriz de covarianzas de los nuevos datos es la matriz de correlación de los datos originales. La segunda es la *estandarización multivariada*: aplicar la transformación que elimina la estructura de correlación de los datos originales. Aquí, la matriz de covarianzas de los nuevos datos es la matriz identidad. Por último, la *transformación en componentes principales*: multiplicar la matriz de los datos por la matriz transpuesta que contiene los eigenvectores de la matriz de varianzas de los datos originales. La matriz de covarianzas de los nuevos datos es una matriz diagonal que contiene los eigenvalores ordenados de la matriz de varianzas de los datos originales. Esta última transformación es utilizada en Análisis de Componentes Principales.

Al considerar la matriz de datos X definida en la sección anterior en la expresión 2.1, se tiene que cada transformación consiste de obtener una nueva matriz de datos Z. Esta matriz tiene como renglones a los vectores $z_i^t = [B(x_i - \overline{x})]^t$, $i = 1, 2, ...n$, donde la matriz B depende de la transformación que se busca. En la Tabla 2.5 se muestra cómo es la forma de B para cada transformación.

Histogramas de Mahalanobis.

El primer método de graficación propuesto se realiza a través de la distancia de Mahalanobis. Esta distancia es una medida que involucra la estructura de correlación de las variables. Se denota como D_m a la distancia de Mahalanobis del punto x a su centroide $\overline{x}$:

$$D_m = (x - \overline{x})^t S^{-1} (x - \overline{x}).$$

Transformación			
	por variable	multivariada	en CP
B	$D_s^{-\frac{1}{2}}$	$S_x^{-\frac{1}{2}}$	V^t
S_z	R	I	Q

Tabla 2.5: Forma de la matriz B para cada Transformación. Se denota como D_s a la matriz diagonal con las varianzas de las variables, S_x es la matriz de varianzas de X y $S_x = VQV^t$ denota la descomposición espéctral de S_x; y R para denotar la matriz de correlación de X, I es una matriz identidad y Q es una matriz diagonal con los eigenvalores de S_x.

Si las variables son no correlacionadas, esta distancia se reduce a la distancia euclidiana al origen de los configuraciones transformados mediante la estandarización por variable, descrita al inicio de esta sección.

Para interpretar esta distancia, se define el conjunto

$$\xi_c = \left\{x|\ (x-\overline{x})^t S^{-1}(x-\overline{x}) \leq c\right\},$$

como se aprecia en la Figura 2.2 este conjunto es un elipsoide, cuya forma sigue la dirección de la correlación de las mediciones de las variables ambientales asociadas al nicho ecológico. El conjunto ξ_c es la región contenida en el elipsoide utilizado en la sección anterior para construir una medida de volumen del Nicho. Las distancias de Mahalaobis corresponden a las distancias euclidianas de las configuraciones ambientales en el espacio transformado por la estandarización multivariada. En la Figura 2.1 se puede apreciar que, la distancia de Mahalanobis de un punto al centroide de las configuraciones ambientales, a diferencia de la distancia euclidiana que sigue círculos, sigue la forma de elipses. En otras palabras, la distancia de Mahalanobis al centroide es la misma para dos configuraciones que se encuentran sobre el elipsoide ξ_c, mientras que, la distancia euclidiana es la misma sobre puntos que se encuentran sobre un círculo.

Tal vez el método más simple para visualizar un vector de datos sea gráficar un perfil o histograma. Entonces, con el objetivo de lograr una representación gráfica sencilla, que simplifique la información contenida en el nicho ecológico, se propone gráficar el histograma de las distancias de Mahalanobis de las configuraciones ambientales con respecto a su centroide. En el caso particular, cuando el conjunto de configuraciones es compatible con una muestra de una variable aleatoria normal p-multivariada, un refinamiento en el histograma ajustaría bien a una distribución ji-cuadrada.

En la Figura 2.3 se presentan los histogramas de las distancias de Mahalanobis de las configuraciones ambientales para cada especie. Lo primero que se puede observar en estos histogramas es la dispersión de cada conjunto,

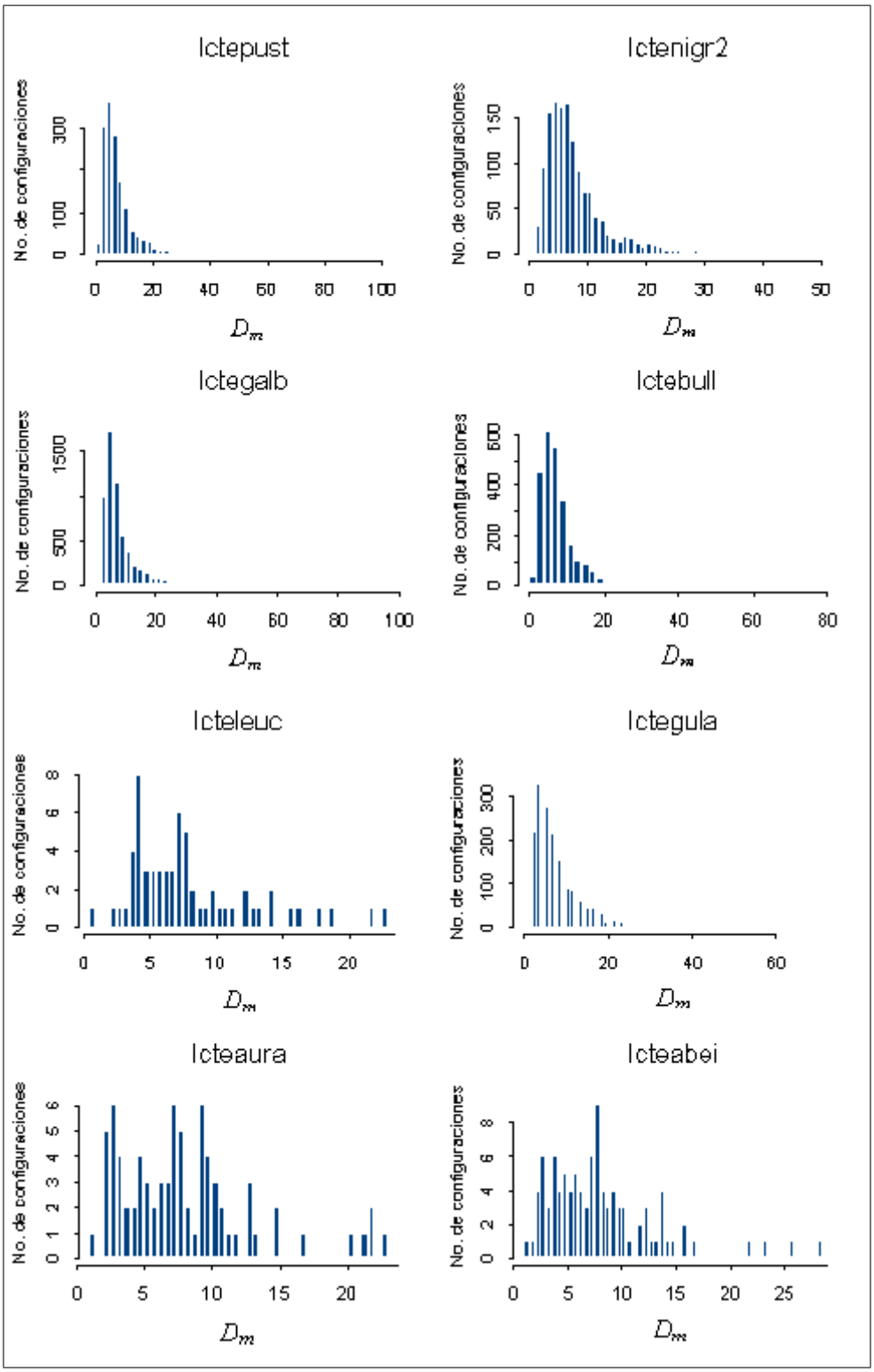


Figura 2.3: Histogramas de Mahalanobis.

lo cual se aprecia claramente en el eje horizontal. Por ejemplo las especies ictepust e ictegalb poseen configuraciones con una distancia de Mahalanobis de hasta 100. Lo anterior coincide con la medida de dispersión presentada en la sección anterior: ambas especies poseen el valor más alto en variación total. Por otro lado, en los histogramas también se pueden observar un poco sobre la estructura de las configuraciones ambientales al identificar modas ó conjuntos demasiado alejados del origen. Las especies icteleuc, icteaura e icteabei tienen en común la presencia de un conjunto de pequeño de configuraciones alejado del centroide lo cual también se puede ver en la Figura C.2. En estos casos el histograma no tiene un descenso gradual en la frecuencia conforme aumenta la distancia, sino que hay un espacio intermedio donde las configuraciones son escasas. Finalmente, una limitación de los histogramas de Mahalanobis es que no es posible identificar simetría en los conjuntos.

Curvas de Andrews.

La constante relación que se tiene con la interpretación de las graficas de funciones es la idea fundamental en la cual se basan las Curvas de Andrews [3], donde cada vector es representado por una función. Para cada configuración ambiental x_i, $i = 1, 2, ...n$, se define la funcion $f_x : [-\pi, \pi] \to \mathbb{R}$ dada por

$$f_{x_i}(t) = x_{i1}/\sqrt{2} + x_{i2}\sin(t) + x_{i3}\cos(t) + x_{i4}\sin(2t) + ..., \ i = 1, 2, ...n.$$

De esta manera un conjunto de puntos en R^p aparecerá como un conjunto de curvas en una grafica en R^2.

La función f_x es una suma ponderada de funciones. La aportación que cada función tiene para definir f_x depende de la variable ambiental asociada a dicha función. Por ejemplo, la influencia que tiene la función $\cos(t)$ en f_x depende del comportamiento de la variable ambiental Bio3. Cuando exista una variable cuyas unidades sean demasiado grandes con respecto a las otras variables ambientales, se tendrá que, la forma de f_x estará completamente determinada por esa variable. Una manera de evitar que las funciones asociadas a ese tipo de variables determine la forma de f_x es reescalar los datos. En este caso, se recomienda aplicar, a cada nicho ecológico, la transformación por variable descrita al inicio de la sección.

Una propiedad de las curvas de Andrews es que la distancia en L_2 entre las funciones correspondientes a dos configuraciones ambientales x_i y x_j es proporcional a la distancia euclidiana entre las configuraciones,

$$\int_{-\pi}^{\pi} \left[f_{x_i}(t) - f_{x_j}(t)\right]^2 dt = \pi D_e^2(x_i, x_j).$$

Entonces cuando se tiene un área entre dos curvas relativamente grande significa que las configuraciones ambientales asociadas a estas curvas también

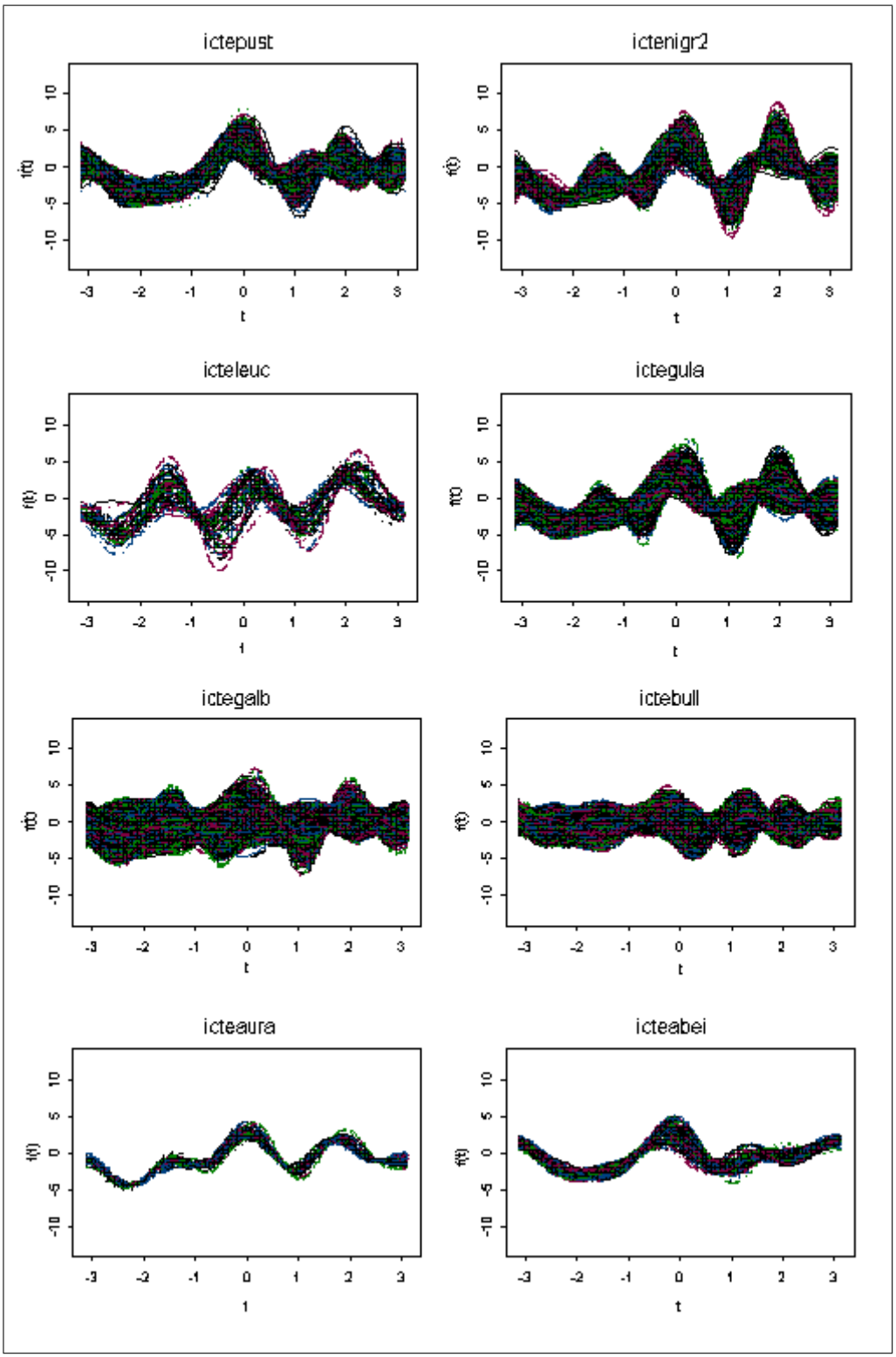


Figura 2.4: Curvas de Andrews de las especies de orioles.

tienen distancia ecuclidiana grande, y viceversa.

En la Figura 2.4, se presentan las curvas de Andrews de las especies. Aunque icteleuc, icteura e icteabei poseen la menor cantidad de configuraciones ambientales se observa que icteleuc es la especie que se encuentra más dispersa. Por el contrario, ictegalb que es la especie con mayor cantidad de configuraciones, en el espacio ecológico se encuentra menos dispersa que ictenigr2, o que ictegula. Estas últimas tienen formas muy parecidas.

Cluster.

Originalmente, el análisis de conglomerados o cluster es utilizado para agrupar elementos; en este caso los elementos a agrupar son las configuraciones ambientales. En el presente trabajo se utilizarán los dendogramas como representaciones gráficas de los nichos ecológicos. Un dendograma es una gráfica que contiene el resultado del proceso de agrupamiento en forma de árbol. Existen diversos métodos de aglomeración [14]. Entre los más utilizados se encuentran el encadenamiento simple, encadenamiento completo, media de grupos, método del centroide y el método de Ward. Los primeros cuatro métodos parten de una matriz de distancias o similaridades, mientras que en el método de Ward se define una medida de heterogeneidad de una agrupación de observaciones en grupos.

La aplicación directa de los métodos de aglomeración que parten de una matriz de distancias, en muchos casos resulta inoperable, sobre todo cuando se trata de grandes conjuntos de datos. Desafortunadamente, algunos de los conjuntos que definen los nichos ecológicos son demasiado grandes. En este caso, el cálculo de la matriz de distancias involucra operaciones computacionalmente costosas debido al tamaño de los nichos. Por esta razón, los dendogramas obtenidos mediante el método de Ward son los que se utilizan para representar a los nichos ecológicos.

Se definirá la siguiente notación: G denota el número de grupos, n_g denota el número de elementos en el grupo g, $x_{i(g)}$ para denotar la i-ésima configuración del grupo g y, $\overline{x}_g$ para denotar el promedio multivariado del grupo. El criterio de agrupación utilizado por el Método de Ward, como se mencionó anteriormente, se basa en la definición de una medida de heterogeneidad de los puntos en grupos. Esta medida es

$$W = \sum_{g=1}^{G} \sum_{i=1}^{n_g} (x_{i(g)} - \overline{x}_g)^t (x_{i(g)} - \overline{x}_g),$$

y es la suma de las distancias euclidianas al cuadrado entre cada configuración y la media de su grupo. El criterio comienza suponiendo que cada configuración ambiental forma un grupo $G = n$ y, por lo tanto, $W = 0$. A continuación se unen las configuraciones que produzcan el mínimo incremento de W. Obviamente, esto implica tomar las más próximas con la distancia

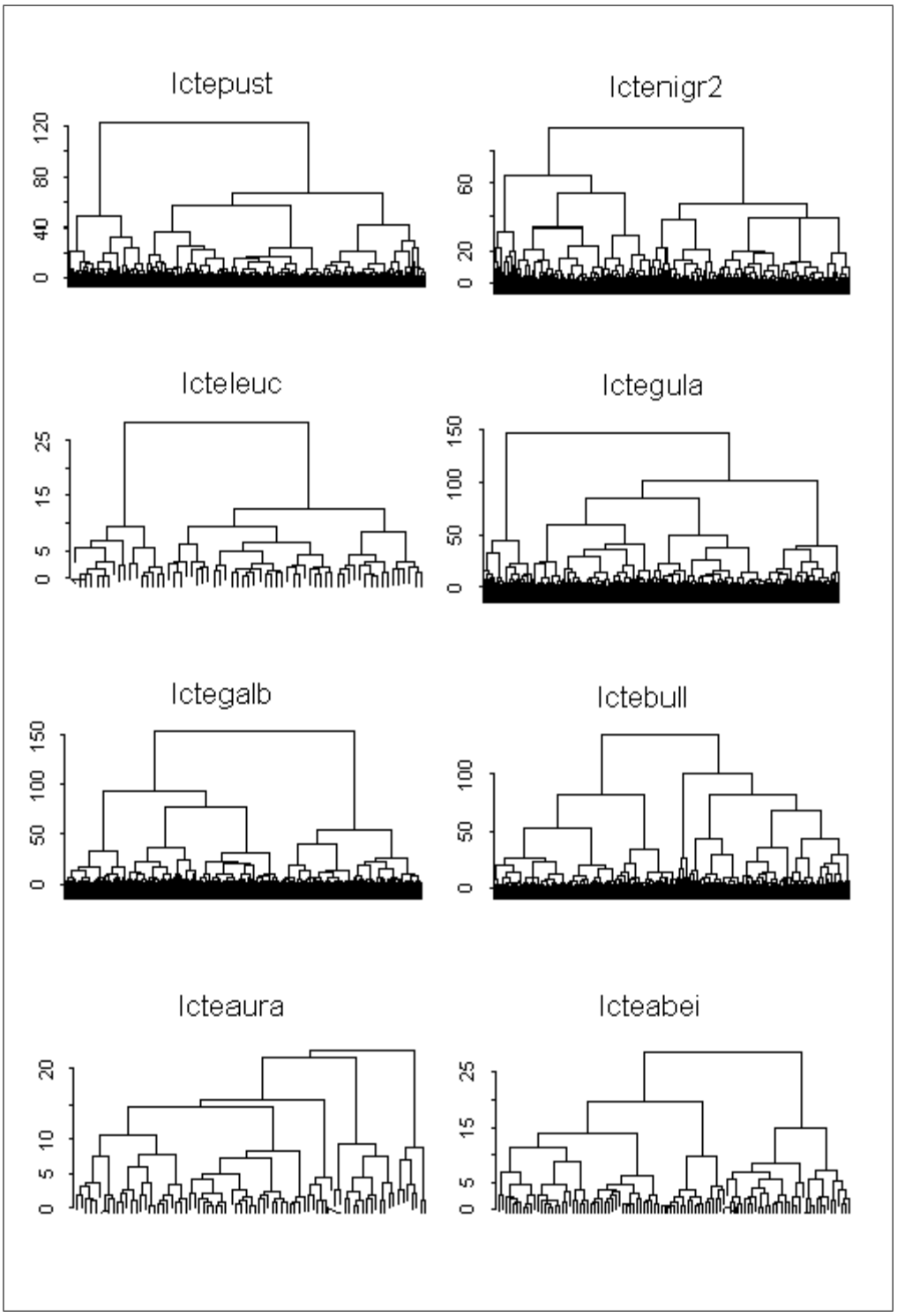


Figura 2.5: Dendogramas de las especies de orioles, utilizando el método de Ward.

euclidiana. En la siguiente etapa se tienen $n-1$ grupos, $n-2$ de una configuración y uno de dos configuraciones. Se decide de nuevo unir dos grupos para que W crezca lo menos posible, con lo que se pasa a $n-2$ grupos y así sucesivamente hasta obtener un único grupo. Los valores de W en cada paso, van indicando el crecimiento del criterio al formar los grupos y puede utilizarse para decidir cuántos grupos naturales contiene el nicho ecológico.

En las Figura 2.5 se presentan los dendogramas de las especies. Estos dendogramas fueron obtenidos con las variables transformadas utilizando la estandarización por variable, descrita al inicio de la sección, para evitar que únicamente las variables con unidades muy grandes determinen la formación de los grupos. Como se mencionó, el dendograma presenta el proceso de agrupamiento de las configuraciones ambientales mediante el método descrito anteriormente. Entonces, cada especie posee una única representación a través de su dendograma. Aunque es complicado distingir las formas de los nichos sólo con los dendogramas, se puede extraer información de éstos. Las configuraciones y sus distancias de unión comienzan a expresar los grupos naturales dentro de cada nicho ecológico. Además, la magnitud de estas distancias (dividida entre el número de configuraciones que se unen) muestran dispersión de los subconjuntos de configuraciones ambientales que constituyen el nicho.

2.2 Comparación de nichos arbitrarios.

Como se mencionó al inicio del Capítulo, el segundo objetivo consiste en realizar comparaciones entre nichos ecológicos. Una manera de comparar nichos es establecer una medida de distancia entre ellos, es decir, determinar la distancia entre nubes de puntos en un espacio p-dimensional. La medida debe ser definida para dos conjuntos de configuraciones ambientales $A, B \subset R^p$, de tamaño n y m respectivamente. Además, una distancia razonable debe estar definida en términos de las distancias de los elementos de los conjuntos. Algunos métodos de aglomeración, mencionados en la sección anterior, utilizan ciertas medidas de distancia entre conjuntos de puntos. Estas distancias son: vecino más cercano, vecino más lejano, distancia promedio, distancia entre centroides. Por otro lado, una medida utilizada en reconocimiento de patrones es la distancia de Hausdorff [19]. Una representación esquemática de estas distancias es presentada en la Tabla 2.6.

El objetivo es proponer una medida adecuada para comparar nichos ecológicos. Las primeras tres medidas (vecino más cercano, vecino más lejano y distancia promedio) podrían ser utilizadas para calcular la distancia entre dos conjuntos de configuraciones ecológicas. Sin embargo, la distancia entre un conjunto y el mismo no siempre es cero. Además, no necesariamente satisfacen la desigualdad del triángulo, por lo tanto, ninguna de estas distancias cumple con las propiedades de una métrica. Por esta razón, no son recomendables

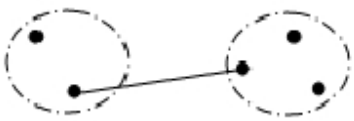
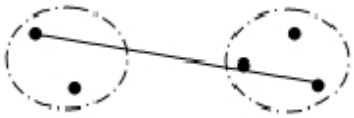
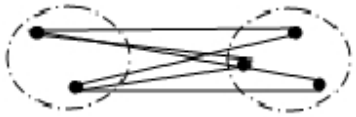
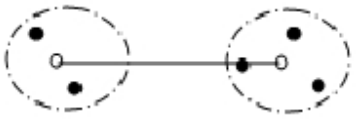
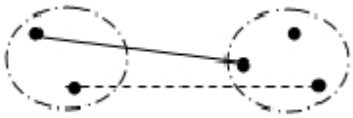
$d(A,B)$	A B
a) $\min_{a\in A, b\in B} D_e^2(a,b)$	
b) $\max_{a\in A, b\in B} D_e^2(a,b)$	
c) $\frac{1}{n_a n_b} \sum_{a\in A, b\in B} D_e^2(a,b)$	
d) $D_e^2(\overline{a},\overline{b})$	
e) $\inf\{r>0 \mid A\subseteq B^r y B\subseteq A^r\}$	

Tabla 2.6: Distancias entre conjuntos: a) Vecinos más cercanos; b) Vecinos más lejanos; c) Distancia promedio, donde n_a es la cardinalidad del conjunto A; d) Distancia entre centroides; e) Distancia de Hausdorff, donde B^r denota la unión de las esferas de radio r centradas en un punto de A. Se denota a $D_e(a,b)$ como la distancia euclidiana entre los puntos a y b.

para la comparación entre nichos ecológicos. Por otro lado, la distancia entre centroides tampoco parece ser un buen candidato para contrastar conjuntos de configuraciones, debido a que no toma en cuenta la dispersión de éstos. Esta medida asigna un valor pequeño cuando se trata de dos conjuntos con promedios similares. Esto ocurre, sin importar que posiblemente uno de los conjuntos es más disperso que el otro, inluso, aunque este último se encuentre contenido totalmente en el conjunto mayor. Finalmente, en este trabajo se sugiere la distancia de Hausdorff como medida para la comparación debido a que no tiene ninguno de los inconvenientes anteriores.

Distancia de Hausdorff. La distancia de Hausdorff es una medida de similaridad definida sobre conjuntos arbitrarios no vacíos A y B como el ínfimo de la distancia de los puntos en A hacia B y los puntos en B hacia A. Esta distancia se formula como sigue:

$$d(A,B) = \inf\{\varepsilon > 0 | A \subseteq B^{\varepsilon} y B \subseteq A^{\varepsilon}\},$$

donde A^{ε} denota la unión de todas las bolas de radio ε centradas en un punto de A.

Cuando se trata de conjuntos finitos de puntos, la distancia de Hausdorff se puede obtener calculando la "distancia" $\overrightarrow{d}$. Ésta consiste en asignar a cada punto de A el valor de la distancia euclidiana al punto más cercano de B; el máximo sobre todos esos valores será la distancia $\overrightarrow{d}$ de A hacia B. De manera similar se obtiene la distancia $\overrightarrow{d}$ de B hacia A. Por último, se toma el máximo de estas dos distancias:

$$d(A,B) = \max\left\{\overrightarrow{d}(A,B), \overrightarrow{d}(B,A)\right\},$$

donde $\overrightarrow{d}(A,B) = \underset{a\in A}{\max}\underset{b\in B}{\min} D_e(a,b)$ y $D_e(a,b)$ denota la distancia euclidiana del punto a al punto b.

La distancia de Hausdorff es una métrica y se puede utilizar como medida de la diferencia entre nichos ecológicos. Esta distancia no tiene los inconvenientes de las medidas mencionadas anteriormente. Es claro que si se calcula la distancia de Hausdorff entre un nicho y él mismo esta medida es cero. Además, cuando se tiene un nicho anidado en otro más grande, pero ambos tienen la misma media multivariada, esta distancia no es cero. En la Figura 2.6 se muestra como actúa la distancia de Hausdorff sobre tres distintos casos: cuando se tienen dos nichos disjuntos, intersección de nichos y un nicho contenido en otro mayor.

Cuando las unidades de las variables no son comparables se debe aplicar la transformación por variable, descrita en la sección anterior, para evitar que las distancias sean determinadas únicamente por las variables con unidades

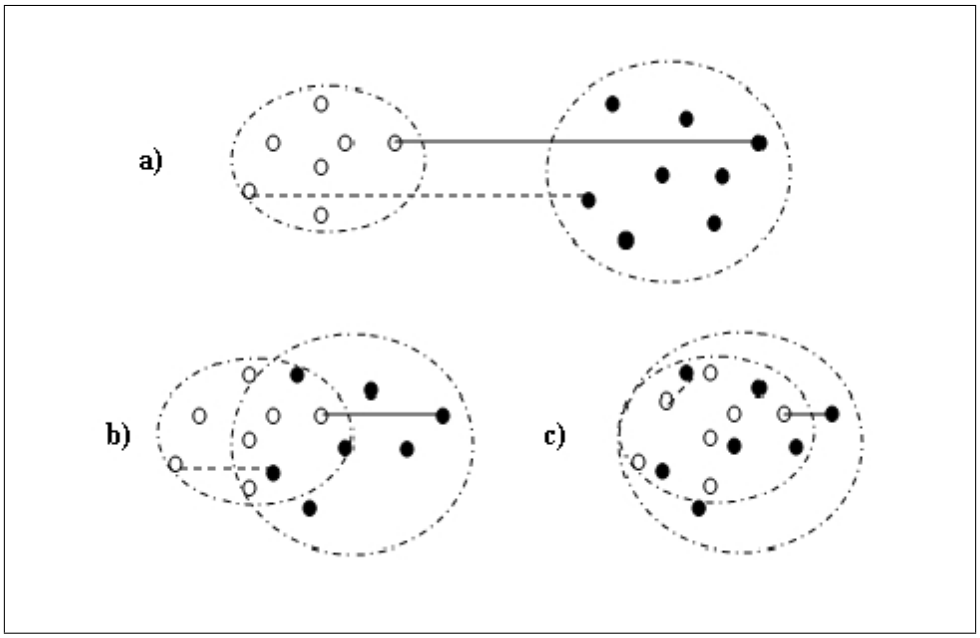


Figura 2.6: Distancia de Hausdorff. a) Conjuntos separados, b) Conjuntos intersectados y c) Conjunto anidado.

más grandes. En este caso, al estandarizar cada nicho con respecto a su media y varianza se pierde la estructura original de los datos. Esto es porque se está enviando cada nicho al mismo lugar, dando como resultado nichos anidados con centroide en el origen. En cambio, estandarizar con respecto a la media y varianza globales (media y varianza de todo el espacio ecológico disponible) produce un cambio menor en el orden original de los conjuntos de configuraciones ambientales. Esta transformación, simplemente, reduce escalas y envia a los centroides de los nichos cerca del origen, sin colocarlos exactamente en él.

En la Tabla 2.7 se presentan las distancias de Hausdorff entre los nichos de las especies de orioles. También en la misma Tabla abajo de las distancias de Hausdorff se muestran las "distancias promedio" [21]. Es claro que cuando la distancia entre dos especies sea pequeña se bebe interpretar que éstas son muy parecidas. Se observa que las especies ictenigr2 e ictegula son las más cercanas de acuerdo a Hausdorff, lo cual coincide con lo que muestran las curvas de Andrews en la Figura 2.4 y los componentes principales en la Figura C.3. Por otro lado, recordar que la especie icteaura tiene muy pocas configuraciones ambientales en comparación con ictenigr2 e ictegula. Además, en la Figura C.8 se observa que icteaura se encuentra contenida en estas especies. Es claro que, cuando el nicho de una especie muy pequeña se encuentre contenido en otro

	nigr2	leuc	gula	galb	bull	aura	abei
pust	2.68	5.82	2.45	4.38	3.60	3.07	3.49
	5.73	13.67	2.60	16.82	11.85	2.46	3.01
nigr2		4.39	**1.93**	4.52	4.23	**4.15**	5.35
		7.10	**3.63**	21.73	19.15	**3.34**	10.68
leuc			4.46	4.80	5.19	6.06	6.85
			9.85	16.27	19.08	7.95	16.69
gula				3.43	4.31	**3.79**	5.56
				16.27	12.43	**1.68**	4.95
galb					3.46	4.58	5.08
					3.70	13.79	12.42
bull						4.17	3.72
						10.04	7.38
aura							2.64
							4.50

Tabla 2.7: Distancias entre las especies de orioles: Distancias de Hausdorff (arriba) y Distancias Promedio (abajo).

mayor no necesariamente significa que ambas especies sean parecidas. En este caso, no parece razonable que icteaura tenga que ser más parecida a ictegula que a ictenigr2, tal como lo sugiere la distancia promedio (tampoco que asigne la misma distancia de ictenigr2 a ictegula que de ictenigr2 a icteaura). En cambio, Hausdorff asigna aproximadamente la misma distancia de icteaura a ictenigr2 que de icteaura a ictegula y una distancia más pequeña de ictegula a ictenigr2, observar la Figura C.8. Esto ilustra, además de las desventajas mencionadas anteriormente, que es mejor utilizar la distancia de Hausdorff en lugar de la "distancia promedio".

Capítulo 3

Idealización probabilística en el espacio ecológico para describir nichos.

Como se mencionó en el Capítulo 1, existen diversas aplicaciones en las que se ha utilizado en la práctica el conocimiento del nicho ecológico para la toma de decisiones estratégicas y para predecir la distribución de especies. Estas aplicaciones son clara evidencia de que los resultados que proporcionan los MENE son en general razonables.

Los subconjuntos producidos como salida por los MENE en ocasiones se interpretan de una forma determinística, en el sentido de que una configuración ambiental pertenece o no pertenece al subconjunto. En esta sección se asume de entrada que el fenómeno no es determinístico sino aleatorio, porque dos localidades distintas pueden poseer idénticas variables ambientales sin que la especie necesariamente se presente en ambos sitios a la vez. Esta presunción obliga a buscar una interpretación probabilística más formal al subconjunto que produce un MENE.

En esta sección se definirá una noción probabilística que proporcione una interpretación apropiada a dichos subcojuntos. Tomando como ejemplos varias salidas auténticas de MENE—que en la práctica han sido tomados por biólogos como nichos fidedignos—se investigará si la noción desarrollada es adecuada. Si dicha noción, en efecto, es adecuada, entonces indirectamente se habilitará un nuevo mecanismo de descripción de un nicho ecológico basado en la descripción del objeto probabilístico asociado.

3.1 Los MENE vistos como regiones de alta conveniencia para la especie

Para lograr una mejor interpretación de las salidas de los MENE, será conveniente comprender un poco más sobre el funcionamiento interior de estos procedimientos, y a su vez recurrir a una analogía con un método de clasificación.

Aunque existan diversos procedimientos, unos más sofisticados que otros, para obtener un nicho ecológico con base en observaciones empíricas, se puede visualizar una generalización de lo que ocurre dentro de un MENE típico como la determinación de una función $f : R^p \to R$. La clasificación de un sitio con ambiente z como dentro o fuera del nicho, está basada en el valor $f(z)$. Por ejemplo, $f(z)$ podría ser una escala continua que mide la conveniencia del ambiente z para sostener a la especie. Aun en el caso particular de que las salidas de los MENE son de naturaleza binaria, en realidad en el interior se trata de una función f cuya imagen es el conjunto $\{0, 1\}$. En este caso, los MENE utilizan información sobre presencias de la especie en las configuraciones ambientales para determinar la función clasificadora de éstas en dos clases: nicho y no nicho.

Una dificultad técnica-operativa desde el punto de vista biológico, es la determinación de sitios donde la especie es ausente. Es interesante notar, por lo tanto, que los MENE no son más que la clasificación de configuraciones ambientales en dos clases, donde la determinación del clasificador es llevada al cabo sin contar con información sobre configuraciones de la clase denominada como no nicho. Es decir, los MENE no son métodos convencionales de clasificación.

Dada una resolución de la rejilla para el espacio geográfico, se define $Đ'$ como el conjunto de configuraciones disponibles en el planeta y $M = \{x_1, x_2, \ldots, x_m\}$ como el conjunto de configuraciones ambientales donde se detectó la presencia de la especie. Notar que $M \subset Đ'$. El problema que abordan los MENE es encontrar la etiqueta 0 ó 1 para cualquier configuración $x \in Đ'$, lo cual se logra a través de la determinación del clasificador f. Ejemplos de las formas explícitas del clasificador para algunos MENE son las siguientes:

- Bioclim.

$$f(z) \quad = \quad \begin{cases} 1 & \text{si} \quad m_k - C_o v_k \leq z_k \leq m_k + C_o v_k \text{ para toda } z_k;\ k = 1, ..., p \\ 0 & \text{en otro caso} \end{cases},$$

donde, $C_0 \in R$ es un valor arbitrario, inducido por un umbral en la suposición de una distribución normal de la variable k-ésima; m_k y v_k son la media y

la desviación estandar de la variable k-ésima, respectivamente. La idea es construir hiper rectángulos que cubren las configuraciones observadas.

- Domain.

$$f(z) \quad = \quad \begin{cases} 1 & \text{si} \quad D_a(z, x_i) \leq C_1; \text{ para algún } x_i \in M \\ 0 & \text{en otro caso} \end{cases},$$

donde $C_1 \in R$ es un valor umbral arbitrario y $D_a : R^p \times R^p \to R^+$ es una medida de disimilaridad entre ambientes. La idea es medir la disimilaridad que tiene una configuración arbitraria, z, respecto a las configuraciones descritas por las presencias observadas.

- Floramap.

$$f(z) \quad = \quad \begin{cases} 1 & \text{si} \quad \widehat{p}(z) > C_2 \\ 0 & \text{en otro caso} \end{cases},$$

donde $C_2 \in R$ es un valor umbral arbitrario y $\widehat{p} : R^p \to [0, \infty)$, es la probabilidad de que z pertenezca a cierta distribución de probabilidad descrita por las configuraciones de presencia de la especie. La idea aquí es decidir si una configuración arbitraria, z, proviene de una distribución de probabilidad aproximada con las configuraciones descritas por las presencias observadas.

- GARP.

  En este caso, no se puede dar una forma explícita del clasificador debido a la complejidad del método. Este procedimiento es una combinación de otros métodos que incluyen un clasificador (Bioclim y Regresión logística, entre otros). Por otro lado, existe la noción de umbral en el momento de la selección de los mejores modelos que caen dentro de la categoria de "Best Sub-sets". Sin embargo, sí es cierto que al final del proceso se trata de una función $f(z)$ que sirve para clasificar una configuración z en nicho o no-nicho.

Ulteriormente, en cada caso se define el conjunto $N = \{z \in \text{Đ}' |\ f(z) = 1\}$ como el nicho estimado. Notar que los nichos que se producen por los MENE presentados anteriormente, son obtenidos truncando alguna función a partir de un umbral. Dicha función ordena las configuraciones ambientales, y un punto es considerado dentro del nicho para valores altos de f. Esto permite interpretar los valores de f como un "grado de conveniencia" para la especie. Lo que estos métodos están entonces produciendo, son subconjuntos que tienen alto "grado de conveniencia" en algún sentido, y ésta será la clave para encontrar una interpretación probabilística más precisa.

3.2 Idealización probabilística.

En el capítulo anterior, implícitamente se le otorgó al nicho ecológico una interpretación binaria, en el siguiente sentido: Si una configuración ambiental se encuentra dentro de la región, se implica la supervivencia de una especie, mientras que fuera de la región la interpretación es que la especie no se da. En la realidad esto no sucede así. Por un lado, existen lugares donde la especie crece y se desarrolla en mejores condiciones mientras que existen otros sitios donde la vida de la especie no es tan propicia aunque ésta puede subsistir. Por otro, hay una componente aleatoria, porque aún en el caso de que dos sitios diferentes tengan el mismo ambiente, la ocurrencia de la especie en un lugar no necesariamente implica la ocurrencia en el otro, no obstante que ambos sitios sean igualmente accesibles a la especie.

Esto sugiere que el fenómeno no es tan simple como sugiere la interpretación binaria, sino que es necesario involucrar algún concepto que permita diferentes grados de conveniencia del ambiente, así como esta noción aleatoria. Se verá que una forma natural de involucrar ambas nociones es mediante una densidad de probabilidad que especifique los niveles de preferencia de la especie sobre las configuraciones ambientales. A esta densidad se le denominará función de preferencia. Los parámetros de esta función, y por consiguiente la función misma, caracterizarían a los nichos ecológicos. En lugar de hablar de un nicho (subconjunto de configuraciones ambientales), se hablaría entonces de una densidad sobre el espacio ecológico.

3.2.1 Función de preferencia.

Con la premisa de que es posible que una especie se establezca en lugares que pueden tener distintos grados de conveniencia, parece razonable sugerir que sobre el espacio ecológico existen ambientes que prefiere la especie sobre otros. Su preferencia sería determinada por las necesidades particulares de la especie, es decir, por las condiciones que le son convenientes. Dentro de estos ambientes preferenciales pueden encontrarse uno o varios óptimos.

Una manera de representar el fenómeno anterior es concebir una función sobre el espacio ecológico que indique el nivel de preferencia de la especie para cada configuración ambiental. Dicha función es una densidad de probabilidad $\Phi : R^p \to R$ que tiene el propósito de describir cómo varía la preferencia de la especie sobre las configuraciones ambientales. Se puede considerar a esta función de preferencia como un "tiro al blanco" idealizado, donde cada especie está apuntando hacia su óptimo ambiental, y donde su posición relativa al óptimo es aleatoria debido a condiciones fortuitas en su dispersión, etc. Es como si cada espécimen determinara al azar su ambiente alrededor de un óptimo formando una distribución definida sobre el espacio ambiental. Esta función de preferencia tiene la interpretación de ser una propiedad inherente

de la especie, es decir, que el grado de conveniencia es irrespectivo de que físicamente se encuentre disponible el ambiente óptimo o no.

En efecto, una cosa es el espacio ambiental completo (sobre el que se define la función de preferencia idealizada para la especie), y otra muy distinta es la colección de ambientes disponibles sobre el planeta Tierra, Đ. Es posible que un espécimen resulte sorteado con una configuración según la probabilidad dictada por Φ, pero que dicha configuración no exista disponible sobre el terreno. Así, el concepto asociado es una densidad condicional de Φ dado Đ, para indicar que una especie debió elegir lo que prefiere biológicamente entre lo que se encuentra disponible físicamente. Dicha densidad condicional, que se representará por g, es la que representa la preferencia de la especie sobre la Tierra. Con más precisión,

$$g(z) = \Phi(z|Đ) = \begin{cases} \frac{\Phi(z)}{\int_Đ \Phi(u)\ du} & \text{si} \quad z \in Đ \\ 0 & \text{en otro caso} \end{cases} . \tag{3.1}$$

Mientras que Φ es un propiedad de la especie por sí misma, g es una propiedad de la especie sujeta al mundo físico.

En el entendido de que nicho está relacionado con conveniencia alta, la densidad g recoge el concepto de conjunto nicho de una manera similar a la inducida por los MENE. La noción de "valores altos" puede describirse mediante un umbral, c_α, especificando que el nicho significa el conjunto $\{z \in Đ|\ g(z) \geq c_\alpha\}$. Es decir, se puede construir un método de clasificación a partir de la idealización anterior de una función de preferencia:

$$f(z) \quad = \quad \begin{cases} 1 & \text{si} \quad g(z) \geq c_\alpha \\ 0 & \text{en otro caso} \end{cases} .$$

Sin embargo, debido a que la función de preferencia, g en este caso es una densidad de probabilidad, existe una cantidad relacionada con este conjunto, dada por

$$\alpha = \int_{\{z \in Đ|\ g(z) \geq c_\alpha\}} g(u)du, \tag{3.2}$$

que tiene una interpretación clara. Se trata de la probabilidad (condicional) de que un espécimen resulte sorteado con una configuración ambiental con conveniencia al menos c_α. Recíprocamente, si se fijara el valor de α, al determinarse c_α que cumpla (3.2) se obtendría un subconjunto de Đ, $\{z \in Đ|\ g(z) \geq c_\alpha\}$, el cual se denominará Conjunto de Probabilidad α, utilizando la notación CP(α). El CP(α) tiene la interpretación de ser un conjunto de predicción de tamaño mínimo, de certidumbre α, para la configuración ambiental que prefiere la especie. Es claro que para fines de estudiar nichos, que los valores de interés para α serían valores altos (cercanos a 1). Esto se debe a que sería poco informativa una aseveración respecto a un subconjunto de Đ que tenga poca probabilidad.

Resumiendo, si se concibe que la función de preferencia—al igual que los MENE—ordena configuraciones ambientales según su conveniencia para la especie, entonces los nichos a los que daría lugar son conjuntos CP(α). Una pregunta natural a partir de lo anterior es: ¿La idealización por vía de la función de preferencia es viable empíricamente? Más adelante se investigará ésto abordando la pregunta indirecta ¿Los nichos comúnmente aceptados, obtenidos por los MENE pueden ser representados como CP(α) para alguna α y para alguna Φ?

Caso rejillas discretas. Como se mencionó en el Capítulo 1, en lo relativo a posiciones geográficas, en la práctica generalmente se opera sobre una rejilla discreta. Por otra parte, para algunas de las variables ambientales se utilizan escalas discretas (por ejemplo, grados centígrados enteros). El conjunto de ambientes disponibles en el planeta, Đ, puede abarcar un continuo de valores. Pero en la práctica, al considerar una restricción del conjunto de configuraciones ambientales para una cierta resolución de la rejilla geográfica, el conjunto relevante, Đ′, es discreto. Por facilidad de especificación, en lo que sigue el espacio E se considerará continuo, con correspondiente función de densidad $\Phi(z)$ continua. Para adaptar la expresión 3.1 al caso de Đ′ se recurrirá al siguiente artificio:

$$g(z) = \Phi(z|Đ') = \begin{cases} \frac{\Phi(z)}{\sum_{u \in Đ'} \Phi(u)} & \text{si} \qquad z \in Đ' \\ 0 & \text{en otro caso} \end{cases} .$$

Esto equivale a una discretización de la función de preferencia condicional 3.1 sobre Đ′. A final de cuentas, se trata de una manera de especificar el comportamiento discreto a partir de un comportamiento continuo e idealizado de la especie (la función de preferencia). A su vez, se trata de una aproximación, porque una posición sobre una retícula en el espacio geográfico (coordenada longitud-latitud), representa una región a su alrededor, ya que no es posible hacer aseveraciones con más precisión que la que otorga la resolución de la rejilla.

3.3 Aproximaciones de MENE usando CP(α).

Para investigar la pregunta sobre la posible interpretación de las salidas de los MENE como CP(α), en esta sección se desarrolará un algoritmo para determinar parámetros de la función de preferencia de una especie, de tal manera que un CP(α) aproxime como conjunto un nicho arbitrario obtenido por un MENE. Posteriormente se presentarán resultados de las aproximaciones obtenidas para las especies de orioles. La conclusión será que se pueden obtener ajustes razonables de nichos a través de CP(α). Entonces, esta inves-

tigación muestra que se puede utilizar un instrumento probabilístico (función de preferencia) para describir nichos ecológicos (salidas de los MENE).

3.3.1 Algunos ejemplos de funciones de preferencia

Como primera exploración para describir el comportamiento de la especie idealizado en E a través de su función de preferencia, se considerará una densidad normal p-variada:

$$\phi(z;\mu,\Sigma) = \frac{1}{[(2\pi)^p \det(\Sigma)]^{\frac{1}{2}}} e^{-\frac{1}{2}(z-\mu)^t \Sigma^{-1}(z-\mu)}.$$

Esta función posee las siguientes interpretaciones. La moda—que es coincidente con la media μ—representa el concepto de configuración ambiental óptima de la especie, es decir el ambiente más conveniente para su desarrollo. Correspondientemente, conforme la intensidad de alguna de las p variables ambientales se distancia de este óptimo, se recoge la idea de una conveniencia gradualmente decreciente. Esta densidad se considera por su simplicidad en cuanto a describir este concepto de óptimo, y a que posee la interpretación física de que conveniencia es consecuencia de un efecto acumulado de un gran número de pequeños efectos. En particular, la simetría (elipsoidal) parecería ser compatible con la idealización de "tiro al blanco" mencionada en la sección anterior.

Por otra parte, en el contexto específico de biología, puede suceder que para una especie existan enmascaradas subespecies, cada una de las cuales definan un óptimo ambiental diferente. En los ejemplos de orioles se concluirá que tal es el caso para dos de las ocho especies consideradas. La manera de abordar esta situación, será a través de una función de preferencia que es la densidad de una mezcla de dos densidades normales multivariadas:

$$\Phi(z;\theta) = \beta_1\phi_1(z;\mu_1,\Sigma_1) + \beta_2\phi_2(z;\mu_2,\Sigma_2),$$

donde $\theta = (\beta_1, \mu_1, \mu_2, \Sigma_1, \Sigma_2)$ es el vector que parametriza a la función, y $\beta_1 + \beta_2 = 1$. De manera similar a la función anterior, μ_1 y μ_2 representan los óptimos ambientales de las posibles subespecies. La mezcla se interpreta como una propiedad inherente de la especie, producida por la existencia de dos subespecies con preferencias diferentes entre sí. Es claro que cuando $\beta_1 = 0$, la función de preferencia que se obtiene no es más que una densidad normal p-variada. Cabe notar que esta propuesta conlleva implícitamente una noción de continuidad en θ. El que la función de preferencia se modula en forma continua como función de θ será importante más adelante cuando se discuta el algoritmo para examinar la pertinencia de esta concepción.

Haciendo las consideraciones anteriores, cualquiera de estas funciones de supervivencia basadas en la distribución normal multivariada, deberán some-

terse al concepto de condicionamiento sobre el conjunto de configuraciones ambientales disponibles, Đ. En virtud de que se están considerando funciones paramétricas, la expresión (3.1) debe reescribirse como

$$g(z;\theta) = \Phi(z;\theta|Đ) = \begin{cases} \frac{\Phi(z;\theta)}{\int_Đ \Phi(u;\theta)\ du} & \text{si} \qquad z \in Đ \\ 0 & \text{en otro caso} \end{cases}.$$

Por otra parte, se tiene la restricción debida a la rejilla definida en el espacio geografico, la cual provocó que los datos en la práctica sean un conjunto de vectores en R^p, denotado por Đ′. Uniendo estas dos consideraciones, se obtiene como propuesta final una discretización de una densidad mezcla de densidades normales condicionada, dada por

$$g(z;\theta) = \Phi(z;\theta|Đ') = \begin{cases} \frac{\Phi(z;\theta)}{\sum_{u\in Đ'} \Phi(u;\theta)} & \text{si} \qquad z \in Đ' \\ 0 & \text{en otro caso} \end{cases}. \tag{3.3}$$

Entonces, la función $\Phi(\cdot;\theta|Đ)$ describe un comportamiento continuo de la especie idealizado en E, restringido a Đ. En contraparte, la función $\Phi(\cdot;\theta|Đ')$ se puede interpretar como una restricción natural de $\Phi(\cdot;\theta|Đ)$ debida a la resolución discreta de la rejilla en el espacio geográfico. Sobre esta última función se implementará el algoritmo para la determinación de un valor de θ que logre una descripción aproximadamente equivalente al conjunto producido por un MENE, con base en la interpretación CP(α).

Cabe notar que, la propuesta de la función de preferencia, dada en la expresión (3.3), sugiere que las especies tienen comportamientos relativamente básicos sobre el espacio ecológico, y que la complejidad de su distribución geográfica proviene más bien del condicionamiento sobre los ambientes disponibles. Además, esta interpretación separa muy bien dos conceptos que tienen ambitos distintos. Lo biológico que sucede en el espacio ecológico, mientras que las configuraciones ambientales disponibles son producto de los accidentes geológicos en la Tierra.

3.3.2 Algortimo para determinar parámetros de una función de preferencia.

La determinación de θ, vector de parámetros de la función de preferencia, requiere la obtención de $1 + 3p + p^2$ valores: parámetro de ponderación β_1, dos vectores μ_1 y μ_2 (especificando p valores por cada vector) y dos matrices de covarianzas Σ_1 y Σ_2 (especificando $p(p+1)/2$ valores por cada matriz). Esta dimensionalidad tan extrema por una parte dificulta la interpretación de la función de preferencia, y por otra puede introducir problemas numéricos por sobreparametrización. Por este motivo se introducirá una suposición simplificadora para reducir la dimensión de θ. Dicha suposición será forzar a que

las variables de cada mezcla tengan covarianza cero, es decir, considerar a Σ_1 y Σ_2 como matrices diagonales. Éstas matrices se pueden ver entonces como vectores en R^p que contienen las varianzas de las variables de cada función mezcla. El supuesto anterior reduce la dimensión de θ a $1 + 4p$. Esta suposición, biológicamente bien podría ser criticable, ya que implica que en el "tiro al blanco" idealizado, no hay correlación entre humedad relativa y temperatura, por ejemplo. Sin embargo, se verá que aun bajo esta simplificación, los resultados son muy prometedores.

Cabe notar que, el espacio de busqueda de los parámetros,

$\theta = (\beta_1, \mu_1, \mu_2, \Sigma_1, \Sigma_2)$, de la función de preferencia realmente es un subconjunto de R^{1+4p}. Esto se debe a que $\beta_1 \in [0, 1]$ y cada una de las entradas de los vectores de varianzas es mayor que cero. Se utilizará la notación Θ para representar el espacio parámetrico

$\left\{\theta \in R^{1+4p} \mid \beta_1 \in [0,1] \ \Sigma_{ij} > 0 \text{ para } i = 1, 2 \text{ y } j = 1, 2, \ldots, p\right\}$.

Sea $N \subset \text{Đ}'$ un nicho estimado por un MENE. No obstante que un MENE no necesariamente posee una interpretación probabilística transparente ni intencional, en lo que sigue se asumirá que N es un CP(α) para algún valor grande de α. Esto significa que N, irrespectivamente de su génesis implícita en el MENE, tiene la propiedad de abarcar una configuración conveniente para la especie con probabilidad alta. Se asumirá también que la función de preferencia $g(\cdot;\theta)$ asociada con el CP(α), es aproximada por la familia $\{g(z;\theta) \mid \theta \in \Theta\}$ descrita en la sección anterior. Esto es, que para algún valor grande de α, el conjunto N satisface:

$$\alpha = P(z \in N) \simeq \sum\nolimits_{\{z \in \text{Đ}' \ \mid \ g(z;\theta) > c_\alpha\}} g(z;\theta) \tag{3.4}$$

para alguna constante c_α. De esta manera, cada valor de $\theta \in \Theta$ determina una función de preferencia, la cual tiene el potencial de tener a N como un CP(α).

Para cuantificar la efectividad de θ para aproximar al conjunto N, se introducirá una función de discordancia que tendrá por objeto evaluar la diferencia entre los conjuntos N y $\left\{z \in \text{Đ}' | g(z;\theta) > c_\alpha\right\}$. Así, el objetivo del algoritmo enunciado enseguida, será determinar el valor de θ que produzca la menor discordancia, es decir, la mejor aproximación posible para un valor preespecificado de α.

Algoritmo basado en minimización de una función de discordancia.

Se utilizará la notación $N_c(\theta)$ para representar al conjunto $\left\{z \in \text{Đ}' | g(z;\theta) > c\right\}$. La motivación de la función de discordancia propuesta se puede dividir en dos pasos. En el primero, para un valor de α preespecificado, se determina el valor del umbral c_α que satisface la expresión (3.4). De esta manera se obtiene una posible aproximación de N a través de $N_{c_\alpha}(\theta)$ (CP(α) que depende de θ). En el segundo paso, se desea forzar a que $N_{c_\alpha}(\theta)$ sea lo más

similar posible a N. Es así, como surge una distancia entre estos conjuntos, $d^*(N, N_{c_\alpha}(\theta))$. Esta distancia será el valor de discordancia entre los conjuntos $N_{c_\alpha}(\theta)$ y N. Entonces, asociando cada valor de θ con este valor de discordancia se construye una función $h : \Theta \to R$. Esta función se denominará como función de discordancia, donde $h(\theta)$=$d^*(N, N_{c_\alpha}(\theta))$. Finalmente, el principio de aproximación descrito en este trabajo consiste en adoptar el modelo con la función de discordancia más pequeña, por lo tanto, el "estimador" de mínima función de discordancia para θ será el valor $\widehat{\theta}$ que minimiza la función h en Θ.

Determinación del valor umbral c_α. Se adoptará la notación ψ_θ para representar la función $\psi_\theta : [0, \max(g(x;\theta))] \to [0,1]$ dada por $\psi_\theta(c) = \sum_{N_c(\theta)} g(z;\theta)$. Esta función no es más que la probabilidad del conjunto de configuraciones mayores que un umbral c. Se desea encontrar un valor de c que produzca una probabilidad α, es decir, se debe encontrar el valor de c que satisfaga $\psi_\theta(c_\alpha) \simeq \alpha$ (Esto se hace númericamente implementando el método de bisección). Una vez que se obtiene este valor se consigue el CP(α) que depende de θ:

$$N_{c_\alpha}(\theta) = \left\{ z \in \text{Đ}' \mid g(z;\theta) > c_\alpha \right\}$$

Distancia entre N y $N_c(\theta)$. Se pretende que, el CP(α) que depende de θ obtenido en el punto anterior sea muy similar al nicho estimado, es decir, $N \simeq N_{c_\alpha}(\theta)$. Ahora, se propone una distancia entre estos conjuntos basada en diferencias simétricas:

$$\begin{aligned} d^*(N, N_{c_\alpha}) &= \#((N \cup N_{c_\alpha}) \setminus (N \cap N_{c_\alpha})) \\ &= \#(N \setminus N_{c_\alpha}) + \#(N_{c_\alpha} \setminus N) \end{aligned} \tag{3.5}$$

Esta distancia es la cardinalidad del conjunto de los puntos que no coinciden: los *errores de comisión* (incluir una configuración como parte del CP(α) que no es parte del nicho) más los *errores de omisión* (no incluir en el CP(α) una configuración que es parte del nicho).

Como se mencionó en la primera sección, algunos de los MENE utilizan únicamente información sobre presencias de la especie para extrapolar nubes de puntos, mientras que las ausencias son difíciles o imposibles de determinar experimentalmente. Por otra parte, el contexto de la aplicación dicta que no necesariamente es simétrico el concepto de error de comisión y error de omisión. Por ejemplo, en una aplicación de diseño de reservas ecológicas para conservación de la especie, es más grave dejar de proteger el nicho que proteger un sitio que no pertenece al nicho. Todo esto sugiere que sería indicado adoptar distancias entre conjuntos que no sean simétricas. Un ejemplo de una modificación de (3.5), que penalizaría más un error de omisión está dado por

$$d^*(N, N_{c_\alpha}) = 2 \times \#(N \setminus N_{c_\alpha}) + \#(N_{c_\alpha} \setminus N).$$

Esta asimetría otorga mayor flexibilidad en el proceso de la determinación del conjunto aproximador.

El algoritmo en pseudocódigo tendría el siguiente aspecto:

Algortimo basado en la minimización de una función de discordancia.

//Entradas: un conjunto de configuraciones ambientales N, un valor α y un punto inicial θ_0//
//Salidas: un valor de $\widehat{\theta}$ que minimiza la función de discordancia y el valor de discordancia asociado a $\widehat{\theta}$//

$h(\theta)$ //función de discordancia//
{
$g(x, \theta) \leftarrow$función de probabilidad condicional discreta //corresponde a la expresión (3.3)//
$N(c, \theta) \leftarrow \{z \in \text{Đ}' | g(z; \theta) > c\}$
$\psi(c, \theta) \leftarrow \sum_{N(c,\theta)} g(z; \theta)$
$f(c, \theta) \leftarrow \psi(c, \theta) - \alpha$
//utilizando un método númerico encontrar el valor de c tal que $f(c) = 0$ (método de bisección*)//
$c_\alpha \leftarrow$ el valor de c tal que $f(c) = 0$
$N_{c_\alpha}(\theta) \leftarrow N(c_\alpha, \theta)$
resultado1$\leftarrow 2 \times \#(N \setminus N_{c_\alpha}(\theta)) + \#(N_{c_\alpha}(\theta) \setminus N)$
devolver resultado1
}

//utilizando algún procedimiento encontrar el valor de θ que minimice h (algoritmo simplex de Nelder-Mead*)//
$\widehat{\theta} \leftarrow$ el valor de θ que minimiza h
disc$\leftarrow h(\widehat{\theta})$
devolver $\widehat{\theta}$ y disc

*Métodos númericos [18] utilizados para la aproximación de la función de preferencia de las especies de orioles.

Unicidad de los parámetros.

Identificabilidad de la función. Una familia paramétrica de funciones $\{g(z;\theta) \mid \theta \in \Theta\}$ es identificable si, para dos vectores de parámetros distintos θ_1 y θ_2, se tiene que $g(\theta_1) \neq g(\theta_2)$. Notar que para $\widehat{\theta}$= ($\widehat{\beta}_1$, $\widehat{\mu}_1$, $\widehat{\mu}_2$, $\widehat{\Sigma}_1$, $\widehat{\Sigma}_2$) un valor obtenido mediante el método descrito anteriormente, se tendría que $\widehat{\theta}^* = (1-\widehat{\beta}_1, \widehat{\mu}_2, \widehat{\mu}_1, \widehat{\Sigma}_2, \widehat{\Sigma}_1)$ también es un valor de mínima discordancia. Es decir, g no es una función identificable, y por lo tanto su representación no es única. Por este motivo se debe incorporar una restricción al procedimiento presentado anteriormente para garantizar la identificabilidad de la función. Esta restricción es la imposición de un orden en los parámetros de ponderación [1], $\beta_1 < \beta_2$, lo que se traduce en $\beta_1 < 0.5$ porque $\beta_1 + \beta_2 = 1$. La idea es restringir el espacio de búsqueda del vector de parámetros a un espacio más pequeño, $\Theta_o \subset \Theta$, donde a cada función $g(\theta)$ especificada en (3.3) sea representado por uno y sólo un vector θ. Entonces, la familia parámetrica de interés será $\{g(z;\theta) \mid \theta \in \Theta_o\}$, donde $\Theta_o = \Theta \cap \left\{\theta \in R^{1+4p} \mid \beta_1 < 0.5\right\}$.

Función condicionada. Se utilizará la Figura 3.1 para explicar que a pesar de imponer la restricción anterior, sigue habiendo una particularidad causada por el condicionamiento sobre una rejilla discreta $\text{Đ}'$. En efecto, es posible que para más de un vector de parámetros se produzca el mismo valor objetivo prescrito en el procedimiento para la determinación de los parámetros. En la figura, las dos elipses pequeñas representan las curvas de nivel de dos funciones de preferencia ligeramente distintas, correspondientes a dos valores distintos de θ. Notar que estas funciones a resolución la rejilla dada, generan el mismo CP(α), debido a que ambas abarcan el mismo subconjunto de configuraciones ambientales. Este subconjunto es representado por once puntos contenidos por ambas elipses. Por lo tanto un mismo subconjunto de configuraciones, en términos de CP(α), puede tener asociados más de dos vectores de parámetros diferentes.

Lo anterior indica que el algoritmo descrito dará como resultado uno de una clase de vectores que son igualmente óptimos. Se denota esta clase de parámetros asociado al conjunto de configuraciones N como:

$$\Theta_N = \{\theta \in \Theta_o \mid \theta \text{ genera el CP}(\alpha) \text{ de mínima discordancia con } N\}.$$

En la figura, si se quiere provocar que las configuraciones contenidas por ambas elipses cambien respecto a los once originales, las elipses tendrían que abarcar por lo menos un punto diferente, y para ello el valor de θ tendría que sufrir un cambio drástico. Aquí es donde la noción de continuidad en la función de preferencia cobra importancia, pues es por continuidad que se concluye que si dos elipses son cercanas, sus valores correspondientes de θ serán cercanos. Esto es, la clase Θ_N es restringida debido a que las configuraciones ambienta-

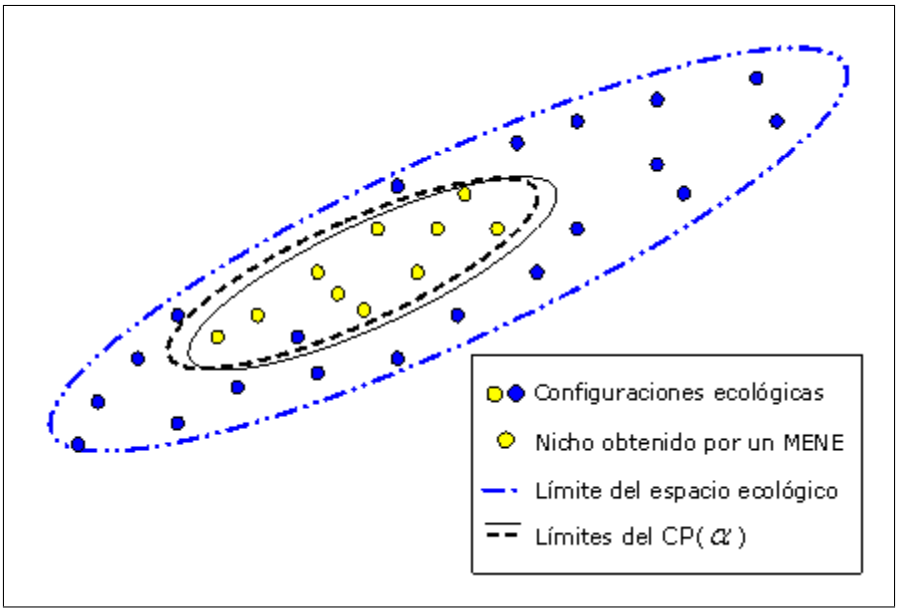


Figura 3.1: Límites del CP(α) que aproxima a un nicho representados por dos elipses pequeñas dentro de una elipse grande.

les acotan la variación en θ. El resultado del algoritmo anteriormente descrito deberá entonces interpretarse como un miembro de una clase Θ_N, y que para fines de descripción, cualquier elemento de esta clase es igualmente conveniente. En las gráficas que se presentan en la siguiente sección se utilizará un representante de Θ_N, el cual es el vector centroide de la región.

3.3.3 Ajuste realizado a especies de orioles.

En la sección anterior se explicó el procedimiento para la determinación de los paramétros de la función de probabilidad que pretende caracterizar a los nichos ecológicos estimados. Mediante este procedimiento se determinaron las funciones de densidad para cada uno de los ocho nichos ecológicos (especies de orioles) obtenidos con el método GARP. Los parámetros son presentados en las Tablas 3.2 y 3.3, el parámetro de ponderación, los dos vectores de medias y los dos vectores de varianzas.

Implementación y bondad de ajuste.

La implementación del procedimiento de estimación de los parámetros fue realizada a través de un programa en el lenguaje matématico *Matlab.* En este software se utilizó la función fminsearch (método simplex), para buscar

Especie	No. iter.	$\#(N)$	$\#(CP(\alpha))$	$\#(N \cap CP(\alpha))$
pust	938	1478	1515	1314
nigr2	287	1385	1484	1237
leuc	674	63	66	32
gula	957	2487	2800	2309
galb	115	5706	6900	5225
bull	1052	2531	2621	2263
aura	74	78	90	67
abei	41	99	82	71

Tabla 3.1: Resultados numéricos obtenidos para las especies de orioles (En la busqueda el vector que minimiza la función de discordancia). Número de iteraciónes necesarias para encontrar el vector óptimo θ, $\#(N)$ denota la cardinalidad de N, $\#(CP(\alpha))$ para denotar la cardinalidad del $CP(\alpha)$ y $\#(N \cap CP(\alpha))$ para la cardinalidad de la intersección; donde, el $CP(\alpha)$ depende del óptimo θ.

	pust	nigr2	leuc	gula	galb	bull	aura	abei
β_1	0.20	0.45	0.30	0.17	0.48	0.49	0.45	0.38
	287	137	149	276	312	285	204	216
	61.4	205.1	12.5	94.0	−18.2	0.0	169.1	34.3
	337	343	161	340	280	273	363	263
	4625	581	3011	4348	6339	6311	2151	1565
μ_1	107	54	17	86	48	59	66	85
	5.1	49.5	140.6	14.2	32.8	14.2	24.4	7.5
	185	531	234	196	171	86	199	165
	709	2617	2206	868	1104	516	516	815
	222	263	81	214	141	139	39	148
	187	128	68	187	329	399	170	244
	150	204	193	179	−16	−107	176	44
	333	312	186	311	294	318	338	294
	1344	655	788	1550	6363	7013	1859	2776
μ_2	119	76	42	79	48	60	59	96
	7.9	22.8	92.4	18.7	32.9	18.6	29.4	5.5
	262	275	278	284	148	69	222	149
	1341	1583	1767	1555	1053	425	1322	639
	206	261	247	226	141	104	259	175

Tabla 3.2: Parámetros de la función de preferencia estimada para las ocho especies de orioles (parámetro de ponderación y parámetros de centralización).

	pust	nigr2	leuc	gula	galb	bull	aura	abei
	4244	882	431	3295	29578	5391	440	390
	7009	772	2667	3652	43426	2840	98	1271
	2995	379	3109	1426	4654	8622	226	1338
	2058100	77797	309750	2002500	36861000	1611200	22847	203820
σ_1^2	503	284	135	1034	1521	1653	142	186
	149	1557	1424	413	1832	273	48	15
	20888	11311	4763	9119	15762	13994	1256	3434
	765810	337570	295890	193940	558830	439760	20961	60695
	3300	253	2881	2464	16999	5520	25	1009
	4879	947	314	2691	29303	7641	315	610
	6796	1087	2090	5816	42846	12836	161	826
	3774	722	1919	2032	4497	8224	145	1014
	2926700	125470	171530	1359400	38364000	4860000	27435	496780
σ_2^2	621	721	340	931	1504	1090	84	212
	56	1097	1318	1120	1844	492	174	10
	14603	12569	6216	23010	15656	2158	2460	2815
	325560	448230	186080	1261000	581540	116510	67101	31272
	4062	500	1829	2085	16831	9876	39	611

Tabla 3.3: Parámetros de la función de preferencia estimada para las ocho especies de orioles (parámetros de dispersión).

Especie	n	discordancia	comisión	omisión	omisión/n
pust	1478	529	201	164	0.11
nigr2	1385	543	247	148	0.11
leuc	63	96	34	31	0.49
gula	2487	847	491	178	0.07
galb	5706	2637	1675	481	0.08
bull	2531	894	358	268	0.11
aura	78	45	23	11	0.14
abei	99	67	11	28	0.28
$m = 14576$ (configuraciones del espacio ecológico)					

Tabla 3.4: Errores de omisión, comisión y discordancia.

un mínimo de una función multivariada. El programa se alimentaría con un conjunto de datos (nicho obtenido por un MENE), un nivel de probabilidad α y un punto inicial. Con el objetivo de acelerar la convergencia (busqueda del mínimo por fminseach) se aplicó un método de aglomeración (Ward) a los conjuntos de configuraciones ambientales: se obtuvieron dos subconjuntos por cada nicho, y para cada uno de éstos se calculó su media multivariada y su vector de varianzas para construir el punto inicial con el que se alimentaría el programa. Por último, aunque no existe una regla universal para especificar el valor de α, es posible proponer un valor cercano a uno. En este caso, en el procedimiento se implementó con un valor de 0.95. Los resultados numéricos obtenidos con la base de datos de orioles se presentan en la Tabla 3.1.

Una primera manera para verificar qué tan adecuado es el ajuste de las funciones de preferencia es observar los errores de omisión y comisión. En la Tabla 3.4 se presentan la cantidad de configuraciones que conforman a cada nicho, los errores de comisión, los errores de omisión, las discordancias, y el porcentaje de errores de omisión (omisión/n). Es obvio que si los errores son pequeños la discordancia tambíen lo será. En el caso de estos resultados, los errores no son tan grandes en comparación relativa con el tamaño de cada nicho. El porcentaje de errores de omisión muestra cuantas configuraciones fueron omitidas del CP(α) (propuesto por el algoritmo) que sí eran parte del nicho estimado por GARP. Excepto por Icteleuc, en las especies se observa un ajuste razonable, lo cual significa que el algoritmo en general reproduce nichos con base en la noción de función de preferencia.

Representación gráfica de las funciones de preferencia estimadas.

En primer lugar, se construyen los 36 diagramas de dispersión de las variables a pares del espacio ecólogico completo (el Continente Americano en este caso) presentadas en las gráficas de la Figura D.1. Posteriormente, se presentan las gráficas que ilustran la función de probabilidad estimada (función de preferencia) para cada una de las especies. Dicha función es representada sobre los diagramas de dispersión a través de "puntos de nivel". Los puntos de nivel tienen un significado similar a las curvas de nivel cuando se quiere representar una función continua. En este caso, el color del punto indica la altura que tiene la función de preferencia, evaluada en ese punto, con respecto a la altura de la función en los puntos con distinto color. Los puntos rojos son los más altos al evaluar la función de probabilidad estimada, el siguiente nivel son los puntos magenta, seguido de éstos los amarillos, después los puntos verdes fuerte, en seguida los puntos verdes claro (son casi cero al evaluar la función), y finalmente, los puntos azules los cuales representan a los puntos más bajos, el valor de la función de preferencia en un punto azul es prácticamente cero.

Para comparar visualmente la función de preferencia (condicional sobre las configuraciones ambientales del Continente Americano), a través de los puntos

de nivel, con los puntos que definen el nicho se muestran los 36 diagramas de dispersión, pero con el nicho estimado por el MENE sobrepuesto. Entonces, para cada especie se repiten dos veces los diagramas de dispersión, la primera vez se ilustra (color rojo y en verde la parte de no nicho) el nicho estimado obtenido vía GARP, y la segunda se presentan los diagramas de dispersión con los puntos de nivel de la función de preferencia. Éstos diagramas de dispersión de los nichos ecológicos son presentados en el Apéndice D.

Cabe señalar que estos puntos de nivel indican la altura de la función de probabilidad mezcla conjunta y no la marginal bivariada, es decir, en los diagramas de dispersión se presenta una proyección de la función de probabilidad conjunta estimada en dos dimensiones. La razón es porque lo que se quiere evaluar visualmente es la estimación de la densidad conjunta, aunque representada con puntos de nivel en dos dimensiones, más no la densidad marginal a dos dimensiones.

Por otro lado, no se debe olvidar que al "estimar" los parámetros lo que se obteniene es una función de probabilidad discreta, que es la condicional de una densidad continua mezcla de normales multivariadas sobre configuraciones ambientales. En la Figura 3.2 se muestran las curvas de nivel de la densidad marginal bivariada (Bio7 y Bio5) de esta función de densidad continua para la especie ictegula. En ésta, los puntos de nivel ilustran cómo una densidad continua, condicionada sobre el espacio ecológico disponible, es utilizada para construir niveles de preferencia de la especie sobre las configuraciones que existen en América.

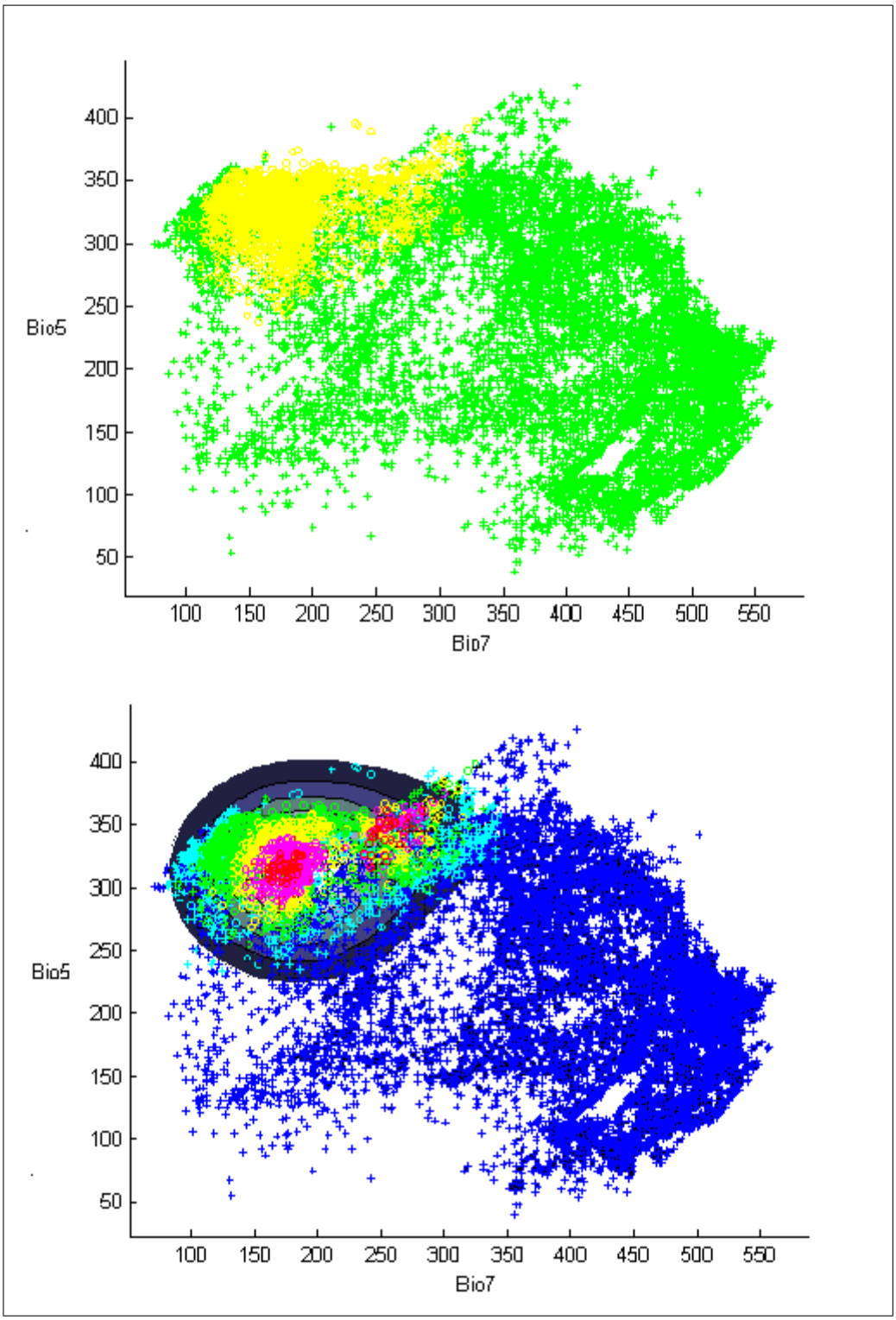


Figura 3.2: Nicho estimado por GARP en Amarillo (arriba). Densidad marginal **continua** para las varibles Bio7 y Bio5 de ictegula (abajo).

Capítulo 4

Conclusiones y comentarios

La motivación principal del presente trabajo fue una inquietud [21] por explorar las propiedades de las especies en el espacio ecológico, a diferencia de prestar énfasis en su distribución geográfica. Este espacio incluye variables ambientales donde se pueden identificar los niveles en que influyen en las especies, definiendo así los nichos ecológicos. La aportación principal de esta tesis fue proporcionar herramientas para la descripción y caracterización de los nichos de las especies mediante dos objetivos distintos.

El primer objetivo conlleva técnicas de Estadística Multivariada, bajo la premisa de que el objetivo es meramente describir un subconjunto del espacio ecológico p-dimensional dado de antemano, aquel obtenido por un MENE. Esta actitud impone una limitación en la interpretación binaria del nicho ecológico. En efecto, bajo la carencia de una interpretación probabilística formal, implícitamente se dicta que todos las configuraciones ambientales dentro de la región que define el nicho implican la supervivencia de la especie. Esto es, se asume que el nivel de supervivencia es el mismo, mientras que en las configuraciones que se encuentran fuera del nicho la vida de la especie no es posible. En el Capítulo 2 se presentan estadísticas que analizan información contenida en cada elemento del nicho ecológico, sin cuestionar que posiblemente alguno de éstos tenga más importancia que otro. Las representaciones gráficas desarrolladas son de gran ayuda para ampliar y comparar resultados obtenidos mediante las estadísticas descriptivas. También, se mencionan posibles medidas para comparar nichos ecológicos, y se propone como medida adecuada para ello a la distancia de Hausdorff por sus propiedades matemáticas y de interpretación.

El segundo objetivo se origina en la búsqueda de una interpretación probabilística para el nicho en el espacio ecológico. Dado que es natural que exista una región óptima en la cual la supervivencia es grande, y regiones subóptimas donde el nivel de supervivencia es menor, se propone una idealización basada en una densidad de probabilidad llamada la función de preferencia. El

concepto resultante es el de un CP(α), desarrollado en el Capítulo 3. Lo anterior se aparta de una interpretación binaria de nicho ecológico, y determina indirectamente una especificación a través de una idealización del comportamiento de la especie en E. La especificación fue llevada al cabo a través de una función de probabilidad condicionada sobre el espacio ecológico disponible sobre la región de estudio.

En la aplicación del algoritmo para determinar los parámetros de la función de preferencia de las especies de orioles, a excepción de icteleuc, se obtuvieron ajustes razonables al observar pocos errores de omisión y comisión. Ésto muestra que se puede utilizar una densidad de probabilidad—función de preferencia—para caracterizar a un nicho ecológico obtenido por un MENE. Un punto importante a resaltar es que no obstante que la función de preferencia es muy simplificada (mezclas de normales multivariadas con covarianzas diagonales), que el grado de similitudes encontradas en la gran mayoría de los casos por este mecanismo de condicionamiento y los MENE es extremadamente notorio. Lo anterior es quizás el resultado más relevante de la tesis: que un MENE puede ser explicado como el resultado de un mecanismo simple que tiene interpretación biológica, y no como el resultado de un complejo algoritmo de clasificación tales como muchos MENE toman en cuenta.

Para decirlo de otra manera: las especies tienen comportamientos relativamente básicos sobre el espacio ecológico, y la complejidad de su distribución geográfica sobre el terreno proviene más bien de condicionamiento sobre los ambientes disponibles. Esta interpretación separa muy bien dos conceptos que tienen distintos ámbitos. Lo biológico sucede en el espacio ecológico, mientras que las configuraciones ambientales disponibles son accidentes de los procesos geológicos y atmosféricos del Planeta Tierra. Esto contrasta con los MENE, los cuales en efecto están ajustando directamente dicha densidad condicional, mientras que el mecanismo probabilístico propuesto aquí lo hace indirectamente (primero una función de preferencia, luego un condicionamiento).

También puede notarse lo siguiente. Dado que el resultado de un MENE (como un subconjunto del espacio ambiental) tiene una interpretación de ser CP(α) para alguna función de preferencia, entonces la colección de métodos descriptivos basados en Estadística Multivariada desarrollados en el Capítulo 2 tienen automáticamente la interpretación de ser métodos para comparar CP(α) entre especies.

La especificación a través de una mezcla de normales proporciona información sobre las posibles subespecies. Cada función mezclada representa a cada posible subespecie, como se observa en la Figura D.9 de la función estimada para la especie Ictegula. Ésto es natural, debido a que dicha especie se da en dos costas de México (dos ambientes distintos), lo que significa que existen dos poblaciones posiblemente diferentes. Ciertamente, aumentar el número de mezclas en el modelo aumentará la información contenida en el

nicho, pero esto involucra una perdida de simplicidad en el modelo mismo. En otras palabras, aumentar la cantidad de funciones a mezclar logrará un mejor ajuste en los datos. Sin embargo, se perdería interpretabilidad al aumentar el número de parámetros, al mismo tiempo que, se consumirían más recursos computacionales.

El concepto de función de preferencia también tiene posibles implicaciones en otros ámbitos. En colaboración con Enrique Martínez-Meyer, del Instituto de Biología de la UNAM, se han comenzado a explorar posibles explicaciones a la relación empíricamente observada que existe entre la abundancia de especies y la distancia a un centroide ecológico. La explicación basada en la función de preferencia es simplemente que la abundancia será mayor cuanto más preferente sea el ambiente. El centroide ecológico no sería más que el concepto de "ambiente óptimo" manejado en esta tesis.

Notar que el nicho de la especie Icteleuc, para la cual no fue posible encontrar una descripción apropiada, es conformado por únicamente 63 configuraciones ambientales, no obstante ésta se encuentra entre las especies con mayor dispersión. En las Figuras C.1 y D.6 se observa que para algunas variables las configuraciones de esta especie se encuentran esparcidas sobre el espacio ecológico disponible. En este caso, la forma del nicho no es compatible con la idea de un comportamiento gradual de la conveniencia de la especie. En consecuencia, el algoritmo no puede encontrar una función de preferencia razonable, debido a que no se pueden reducir los errores de omisión y comisión a la vez. Por este motivo se recomienda investigar si en el campo este tipo de nichos es factible. Quizás las configuraciones que se encuentran dentro de las que definen este nicho pueden también ser parte de él, o bien el algoritmo GARP pudiera tener una ideosincrasia particular que explicara por qué arrojó como resultado ese nicho atípico respecto a los demás.

El presente trabajo plantea algunos temas a los que se les puede dar seguimiento. Entre los más importantes se encuentran los siguientes:

- Implementación de la metodología presentada en el Capítulo 3 para salidas de distintos MENE. Los resultados de estas implementaciones pueden proponer una modificación en la función de preferencia, así como una propuesta en las penalizaciones de la función de discordancia.

- Determinar, con información de presencias, los parámetros de la función de preferencia de la especie. Posteriormente, construir el clasificador de configuraciones ambientales, descrito en el Capítulo 3, basado en el truncamiento de la función de preferencia. El objetivo será investigar si en la práctica éste clasificador puede ser implementado en sí mismo como un MENE.

- En el caso de un resultado positivo en el punto anterior, comparar los resultados del MENE propuesto con los métodos convencionales.

Bibliografía

[1] Aitkin, M., and Rubin, D. B. (1985), "Estimation and hypotesis testing in finite mixture models,". *Journal of the Royal Statistical Society*, B 47, 67–75.

[2] Anderson, P. R., Lew A., Peterson T. (2003), "Evaluating predictive models of species' distributions: criteria for selecting optimal models," *Ecological modeling*, 162, 211–232.

[3] Andrews, D. F. (1972), "Plots of high-dimensional data," *Biometrics, 28*, 125–136.

[4] Argáez-Sosa, J., Christen, J. A., Nakamura, M., and Soberón, J. (2005), "Prediction of high potential areas of habitat for monitored species," *Environmental and Ecological Statistics,* 12, 27–44.

[5] Carpenter, G., Gillison, A. N., and Winter, J. (1993), "Domain:A flexible Modelling Procedure for Mapping Potential Distribution of Plants nad Animals," *Biodiversity and conservation*, 2, 667–680.

[6] Chernoff, H. "Using faces to represent points in k-dimensional space graphically," *Journal of de American Statistical Association*, 68, 361–368.

[7] Duran, B., and Olsen, P. (1992), "Customizing BIOCLIM to investigate Spatial and Temporal Variations in highly Mobile Species," GeoComputation CD-ROM, Pullar, D.V.(ed).

[8] Gladkov, A., and Jones, P. G. (1999), "Floramap: a Computer Tool for Predecing the Distribution of plants and Other Organism in the Wild," versión 1, Jones, A.L.(ed.), CIAT CD-ROM, Cali, Colombia: Centro Internacional de Agricultura Tropical.

[9] Hutchinson, G. E. (1957), "Concluding remarks," *Cold Spring harbor Symposia on Quantitative Biology*, 22, 415–427.

[10] Johnson, R. A., and Wichern, D. W. (1998), Applied Multivariate Statistical Analysis, Prentice Hall.

[11] Kleiner, B., and Hartigan, J. A. (1981), "Representing points in many dimensions by trees and castles," *Journal of de American Statistical Association*, 76, 260–269.

[12] Mardia, K. V. (1970a), "Measures of Multivariate skewness and kurtosis with applications,". *Biometrika*, 57, 519–530.

[13] Mardia, K. V. (1974), "Assessment of Multinormality and the Robustness of Hotelling´s T^2 Test,". *Journal of Applied Statistical Science*, 24, 164–166.

[14] Mardia, K.V, Kent, J. T., and Bibby, J. M.(1979), Multivariate Analysis, Academic Press.

[15] Nakamura, M., and Soberon, J. (2007), "Some thoughts on the notions of areas of distribution and environmental niches as related to available data," Manuscrito en preparación.

[16] Peters, D. and Stockwell, D. (1999), "The GARP Modelling System: Problems and Solutions to Automated spatial Prediction," *International Journal of Geographical Information Science*, 13, 143–158.

[17] Phillips, S. J., Anderson, R. P., and Schapire, R. E. (2006), "A Maximum Entropy Modeling of Species Geographic Distributions," *Ecological Modelling,* 190, 231–259.

[18] Quarteroni, A., Sacco, R., and Saleri, F. (2000), Numerical Mathematics, Springer, 245–251, 297–300.

[19] Rucklidge, W. (1996), Efficient visual recognition using the Hausdorff distance, Lecture Notes in Computer Science, Springer.

[20] Soberón, J., Golubov, J., and Sarukhán, J. (2001), "The importance of Opuntia in México and routes of invasion and impact of Cactoblastis cactorum," (Lepidoptera: Pyralidae), Florida Entomologist, 84, 4, 486–492.

[21] Soberón, J., Eaton, M., Menon, S., and Peterson, T. (2007), "Methods for Measuring Enviromental Niches to Estimate Geographical Distributions of Species," Manuscrito en preparación.

Apéndice A

Notación.

E : Espacio ecológico, si se consideran p variables ambientales, entonces $E = R^p$.

G : Espacio geográfico, el conjunto de las coordenadas geográficas, donde $G \subset R^2$.

Đ : Espacio ecológico disponible, compuesto por las condiciones ambientales que existen en el planeta, donde Đ$\subset E$.

Đ$'$: Espacio ecológico disponible, compuesto por las condiciones ambientales que existen en el planeta sujeto a una resolución de la rejilla en el espacio geográfico. Este espacio es discreto, es decir, es un conjunto de m configuraciones ambientales que son consecuencia de la medición del variables ambientales en m celdas geográficas.

φ : Función $\varphi : G \to E$, que especifica la relación entre el espacio geográfico y el espacio ecológico.

D_p : Distribución potencial, formada por las regiones geográficas donde se dan las condiciones para la supervivencia de una especie.

D :Distribución realizada, región de la distribución potencial donde la especie se encuentra presente.

N_f : Nicho fundamental, parte del espacio ecológico donde se dan las condiciones necesarias para la supervivencia de la especie.

N_p : Nicho potencial, subconjunto del nicho fundamental cuyos factores ecológicos ocurren en la tierra.

N_r : Nicho realizado, subconjunto del nicho potencial en cuyos factores ecológicos se encuentra presente la especie.

N : Salida de un MENE, conjunto de configuraciones ambientales que definen el nicho ecológico donde $N \subset$Đ$'$.

m : Número de configuraciones ambientales que definen el espacio ecológico disponible discreto.

n : Número de configuraciones ambientales que definen el nicho.

p : Número de variables ambientales que definen el nicho.

X_{nxp} : Matriz de datos ecólogicos, cuyos renglones contienen las configuraciónes ambientales dada una rejilla en el espacio geográfico.

x_i : Configuración ambiental, vector con las p mediciones de las variables de la i-ésima configuración ambiental.

x_{ij} : Representa la j-ésima medición de la i-ésima configuración ambiental.

$\overline{x}$: Promedio multivariado, centro de gravedad del conjunto de configuraciones ambientales que constituyen el nicho ecológico de una especie.

s_{ij} : Covarianza entre las variables i y k.

r_{ij} : Correlación entre las variables i y k.

y_i : vector con las n mediciones de la i-ésima variable ambiental.

S_{pxp} : Matriz de varianzas, contiene en su diagonal las varianzas de las p variables, y fuera de la diagonal las covarianzas de las variables.

$det(S)$: Determinante de la matriz S.

$tr(S)$: Traza de la matriz S, suma de los valores de la diagonal de S.

A_s : Sesgo multivariado, índice que representa una medida de asímetria del conjunto de puntos que definen el nicho ecológico.

K : kurtosis, estadístico que representa una medida de lo picudo de una distribución de probabilidad.

∇_c : Voumen Elipsoide, volumen de un elipsoide que contiene un conjunto de configuraciones ambientales con una distancia de Mahalanobis (con respecto a su centroide) menor que c.

D_s : Matriz diagonal de $p \times p$ que contiene en la diagonal las varianzas de las p variables ambientales.

V : Matriz que contiene los p eigenvectores de S, obtenida a partir de la descomposisción espectral.

Λ : Matriz diagonal de $p \times p$, contiene en la diagonal los eigenvectores de S, obtenida a partir de la desconposición espectral.

S_y : Matriz de varianzas de $p \times p$, después de aplicar una transfornación a X_{nxp}.

R : Matriz de correlación, contiene en su diagonal unos, y fuera de la diagonal las correlaciones de las variables.

D_m : Distancia de Mahalanobis de un punto a su centroide.

De : Distancia Euclidiana.

ξ_c : Elipsoide formado por las configuraciones con una distancia de Mahalanobis (con respecto a su centroide) menor que c, donde $\xi_c \subset .E$

f_{x_i}: Curva de Andrews de x_i, función $f : R^p \rightarrow R$, que representa a la configuración x_i

W : Medida de heterogeneidad de las configuraciones ambientales en grupos; suma de las distancias euclidianas al cuadrado entre cada elemento y la media de su grupo.

d : Distancia de Hausdorff entre dos conjuntos A y B, donde $A, B \subset R^p$.

f : Función $f : R^p \rightarrow R$ que caracteriza a los Métodos de Etimación de Nichos Ecológicos; el caso particular cuando $f : R^p \rightarrow \{0, 1\}$, f especifica a un clasificador.

g : Función de preferencia de una especie.

α : Nivel de probabilidad, es un valor cercano a 1.

N : Salida de un MENE, conjunto de configuraciones ambientales que definen el nicho ecológico donde $N \subset Đ'$.

c_α : Altura de una función de preferencia asociada a un nivel de probabilidad α.

$CP(\alpha)$: Conjunto de probabilidad alfa, conjunto de configuraciones ambientales, cuya función de preferencia asociada es mayor que c_α y la suma del valor de la función de preferencia sobre este conjunto es aproximadamente α.

$\phi(x; \mu, \Sigma)$: Función de densidad de una v.a. normal p-variada, parametrizada por μ, vector de medias; y Σ, matrix de varianzas.

$\Phi(x; \theta)$: Función de densidad mezcla de dos funciones de densidad de v. a. normales multivariadas, parametrizada por θ

$g(x; \theta)$: Función de preferencia de una especie (condicional), $g : Đ' \rightarrow R$, parametrizada por θ.

θ : vector de parámetros de una densidad mezcla normal multivariada, θ: $= (\beta_1, \mu_1, \mu_2, \Sigma_1, \Sigma_2)$, β_1 es parámetro de ponderación de las densidades mezcladas, μ_1 y Σ_1 parámetros de la primera función mezcla y μ_2 y Σ_2 parámetros de la otra densidad mezcla.

$h(\theta)$: Función de Discordancia, función $h : R^{37} \rightarrow R$ que compara el conjunto de probabilidad alfa propuesto con el nicho estimado.

Θ : Espacio parámetrico, subconjunto de R^p que se origina a partir de una restricción del parámetro de ponderación y los parámetros de dispersión de la función de probabilidad.

Θ_o : Espacio parámetrico, subconjunto de Θ que se origina a partir de una restricción de orden en los parámetros de ponderación.

Θ_N : Clase que contiene vectores de parámetros asociados a un conjunto de configuraciones ambientales en términos del CP(α).

Apéndice B

Traza de S y distancias a pares.

Si se tiene un conjunto de n puntos $\{x_1, x_2, \ldots x_n\}$, donde $x_i \in R^p$ para $i = 1, 2, \ldots, n$; $\overline{x}$ y S son la media multivariada y la matriz de varianzas, respectivamente. Entonces, se demostrará que el número de puntos mutltiplicado por la traza de S es igual a la suma de las distancias euclidianas al cuadrado a pares de los puntos dividida entre n.

$$\begin{aligned} ntr(S) &= ntr\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})(x_i-\overline{x})^t\right) \\ &= tr\left(\sum_{i=1}^{n}(x_i-\overline{x})(x_i-\overline{x})^t\right) \\ &= \sum_{i=1}^{n} tr\left((x_i-\overline{x})(x_i-\overline{x})^t\right) \\ &= \sum_{i=1}^{n} tr\left((x_i-\overline{x})^t(x_i-\overline{x})\right) \\ &= \sum_{i=1}^{n}(x_i-\overline{x})^t(x_i-\overline{x}) \end{aligned}$$

$$\sum_{i=1}^{n}(x_i-\overline{x})^t(x_i-\overline{x})=\sum_{i=1}^{n}x_i^tx_i-2\overline{x}^t\sum_{i=1}^{n}x_i+n\overline{x}^t\overline{x}$$

$$=\sum_{i=1}^{n}x_i^tx_i-n\overline{x}^t\overline{x}$$

$$=\sum_{i=1}^{n}x_i^tx_i-\frac{1}{n}\left(\sum_{i=1}^{n}x_i\right)^t\left(\sum_{i=1}^{n}x_i\right)$$

$$=\sum_{i=1}^{n}x_i^tx_i-\frac{1}{n}\left[\left(\sum_{i=1}^{n}x_i^tx_i\right)+2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}x_i^tx_j\right]$$

$$=\sum_{i=1}^{n}x_i^tx_i-\frac{1}{n}\left[\left(x_1^tx_1+\sum_{i=2}^{n}ix_i^tx_i-\sum_{i=2}^{n}ix_i^tx_i+\sum_{i=2}^{n}x_i^tx_i\right)+2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}x_i^tx_j\right]$$

$$=\sum_{i=1}^{n}x_i^tx_i-\frac{1}{n}\left[\left(\sum_{i=1}^{n}ix_i^tx_i-\left[\sum_{i=2}^{n}(i-1)x_i^tx_i\right]\right)+2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}x_i^tx_j\right]$$

$$=\sum_{i=1}^{n}x_i^tx_i-\frac{1}{n}\left[\left(\sum_{i=1}^{n}ix_i^tx_i-\left[x_2^tx_2+2x_3^tx_3+...+(n-1)x_n^tx_n\right]\right)+2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}x_i^tx_j\right]$$

$$=\sum_{i=1}^{n}x_i^tx_i-\frac{1}{n}\left[\left(\sum_{i=1}^{n}ix_i^tx_i-\left[\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}x_j^tx_j\right]\right)+2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}x_i^tx_j\right]$$

$$=\sum_{i=1}^{n}x_i^tx_i-\frac{1}{n}\left[nx_n^tx_n+\sum_{i=1}^{n-1}ix_i^tx_i-\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}x_j^tx_j+2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}x_i^tx_j\right]$$

$$=\frac{1}{n}n\sum_{i=1}^{n-1}x_i^tx_i-\frac{1}{n}\sum_{i=1}^{n-1}ix_i^tx_i+\frac{1}{n}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}x_j^tx_j-\frac{2}{n}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}x_i^tx_j$$

$$=\frac{1}{n}\left[\sum_{i=1}^{n-1}(n-i)x_i^tx_i+\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}x_j^tx_j-2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}x_i^tx_j\right]$$

$$=\frac{1}{n}\left[\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}x_i^tx_i+\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}x_j^tx_j-2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}x_i^tx_j\right]$$

$$=\frac{1}{n}\left[\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left(x_i^tx_i+x_j^tx_j-2x_i^tx_j\right)\right]$$

$$=\frac{1}{n}\left[\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(x_i-x_j)^t(x_i-x_j)\right]$$

Apéndice C

Gráficas de los Componentes Principales.

.

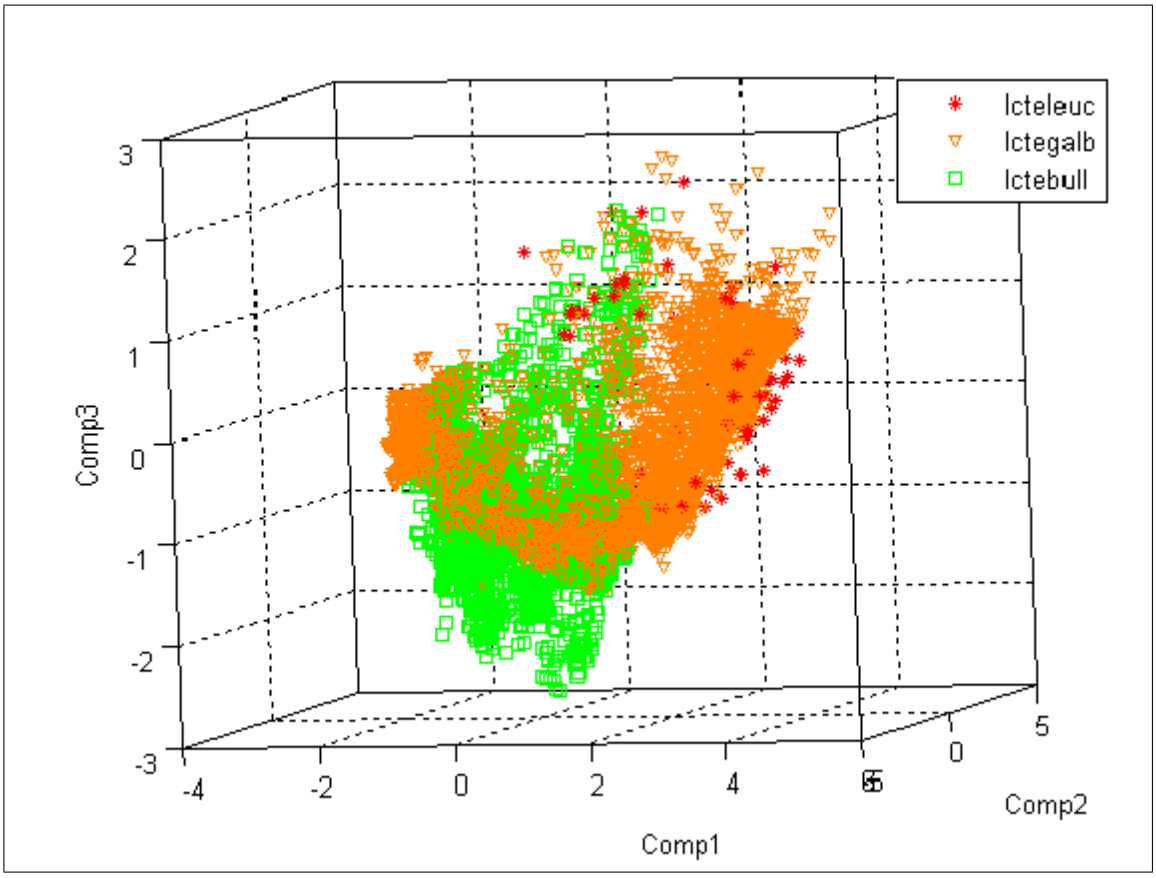


Figura C.1: Primeras tres componentes principales de las especies Icteleuc, Ictegalb e Ictebull.

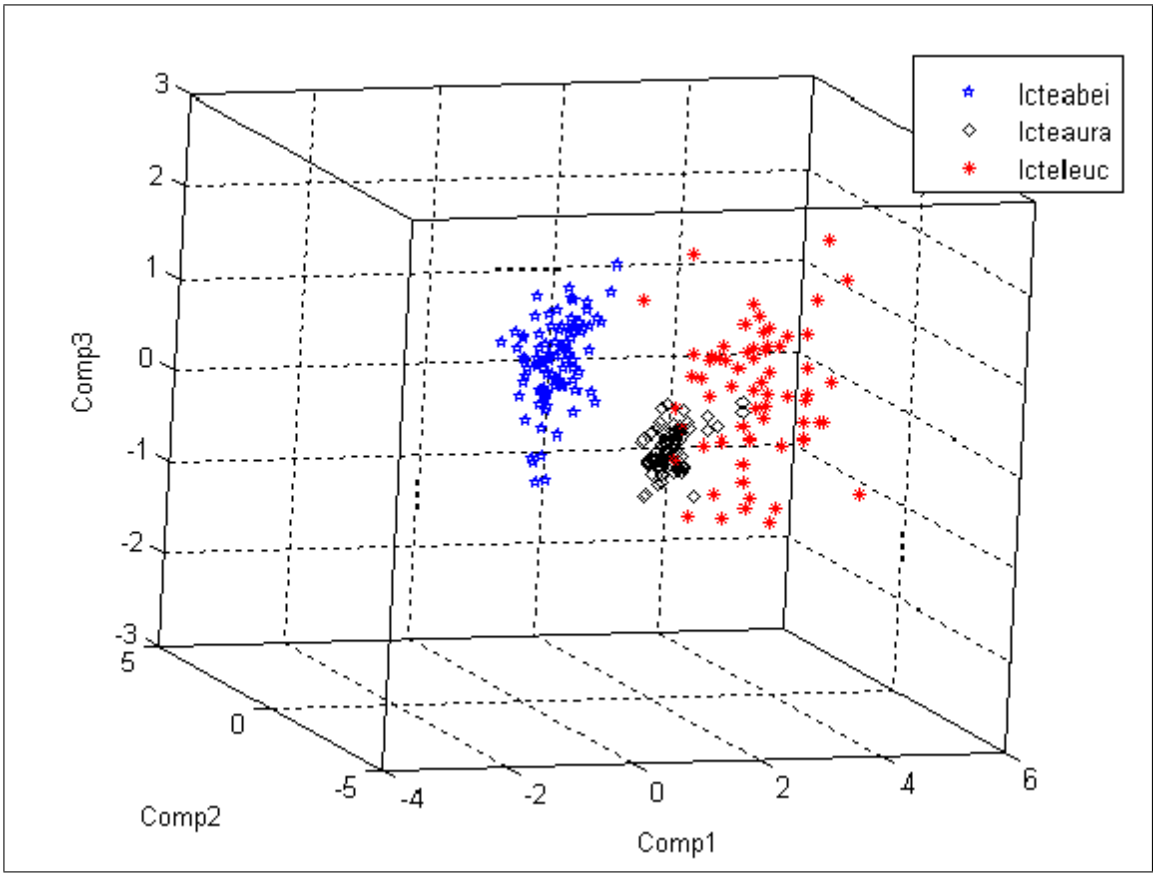


Figura C.2: Primeras tres componentes principales de las especies Icteleuc, Icteaura e Icteabei.

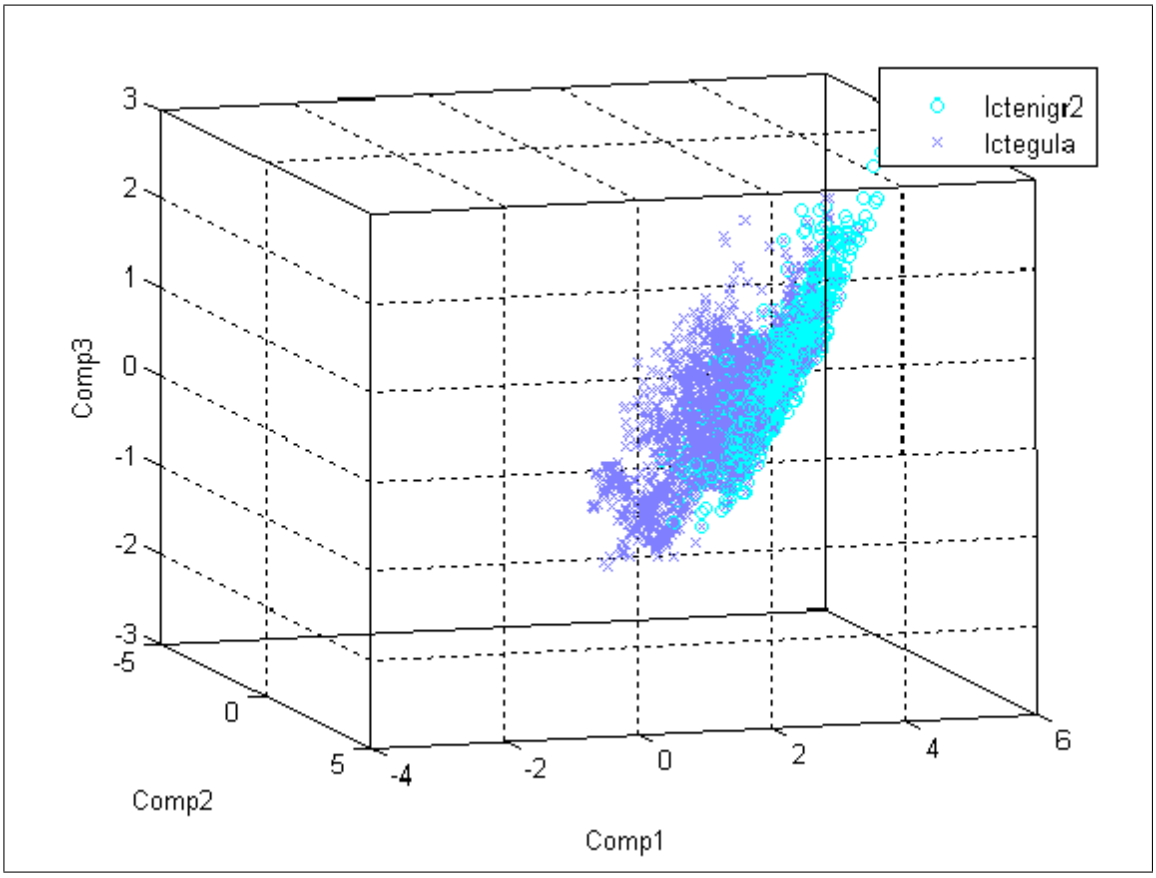


Figura C.3: Primeras tres componentes principales de las especies Ictenigr2 e Ictegula.

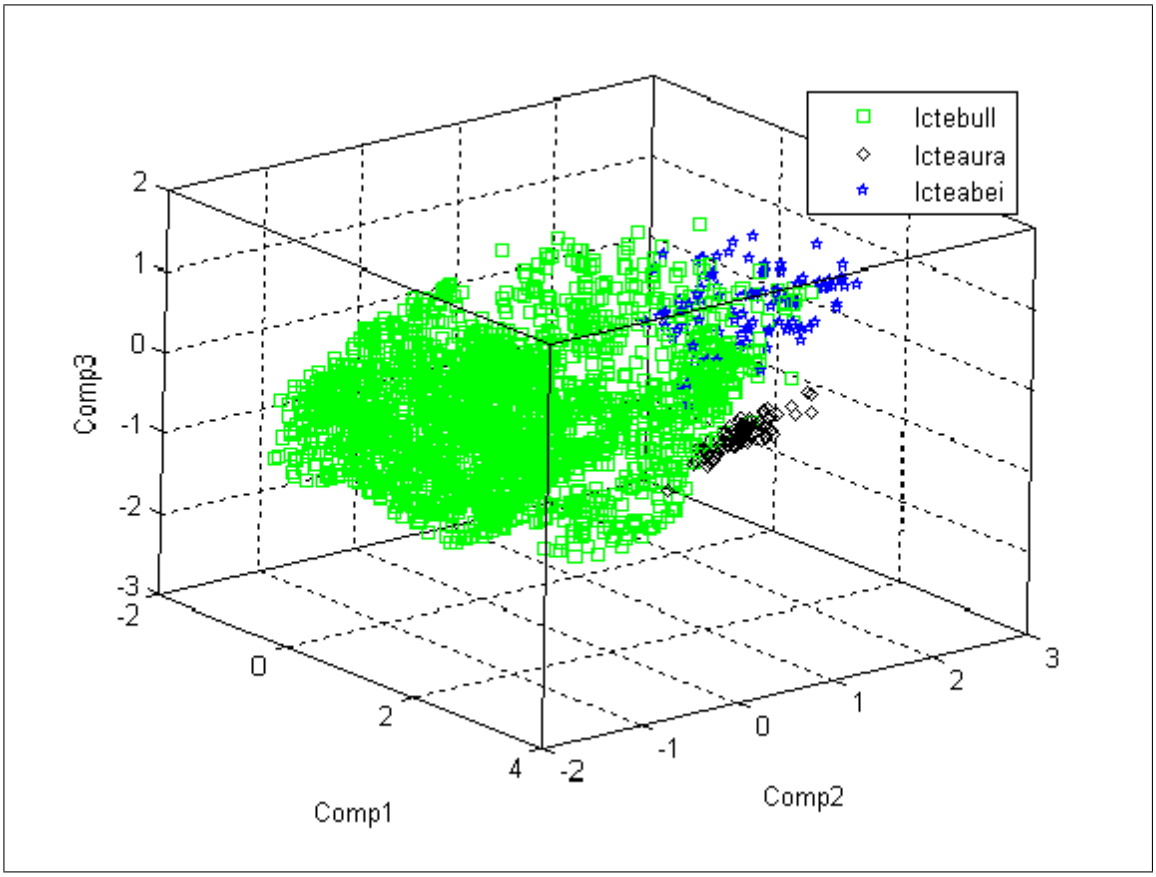


Figura C.4: Primeras tres componentes principales de las especies Ictebull, Icteaura e Icteabei.

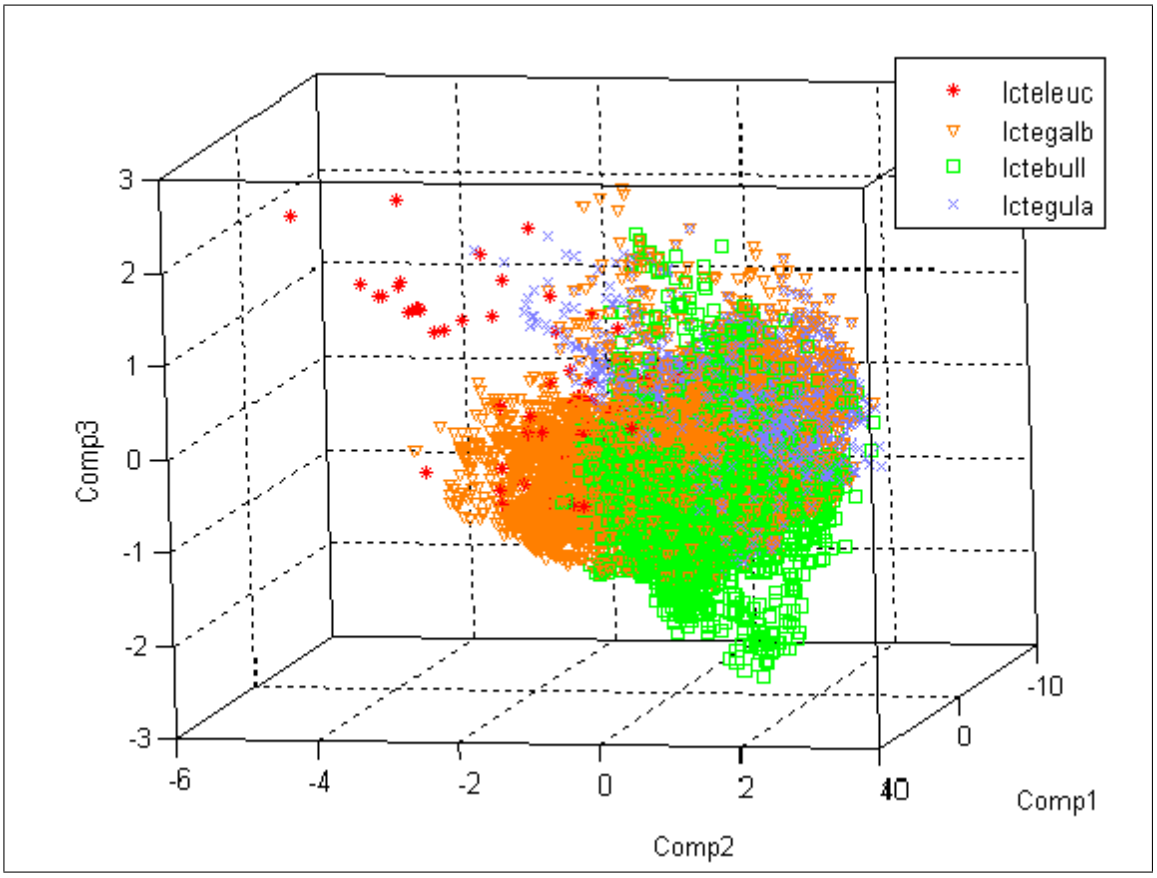


Figura C.5: Primeras tres componentes principales de las especies Icteleuc, Ictegalb, Ictebull, e Ictegula.

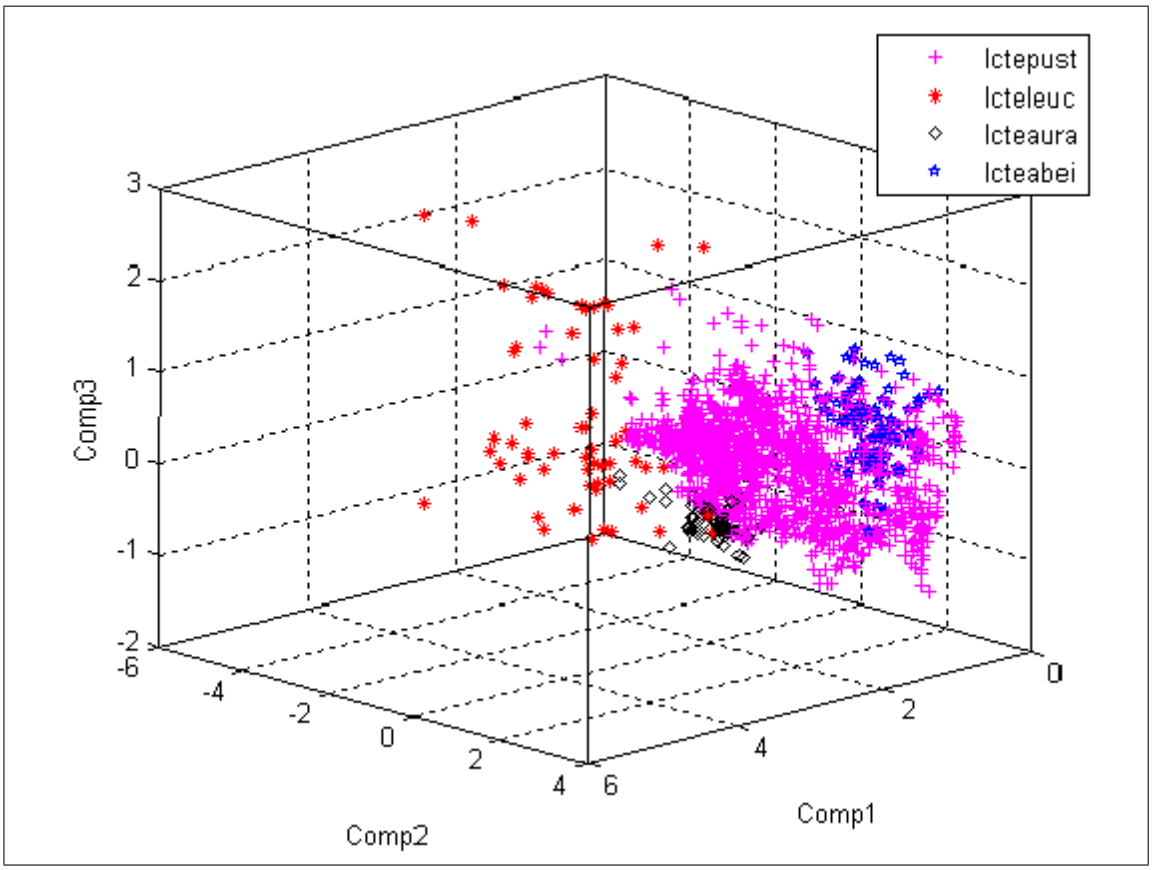


Figura C.6: Primeras tres componentes principales de las especies Ictepust, Icteleuc, Icteaura e Icteabei.

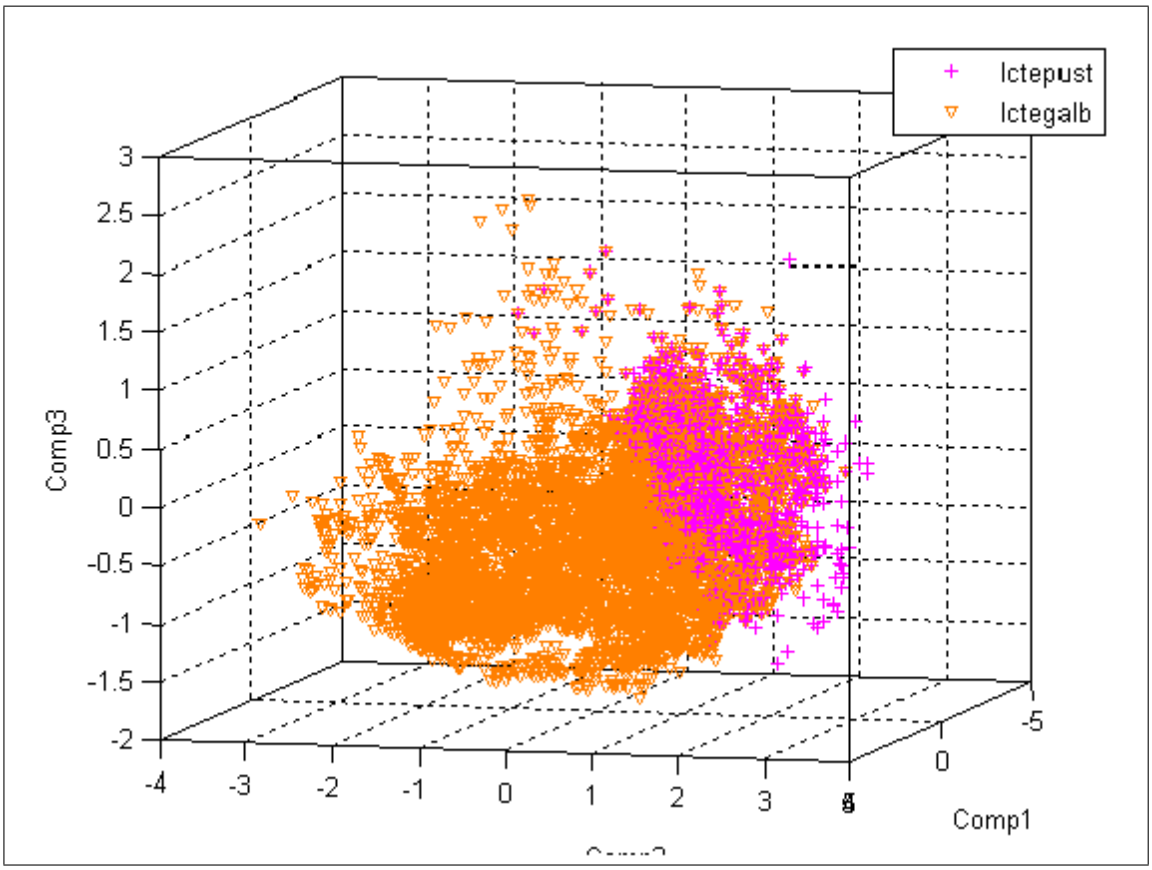


Figura C.7: Primeras tres componentes principales de las especies Ictepust e Ictegalb.

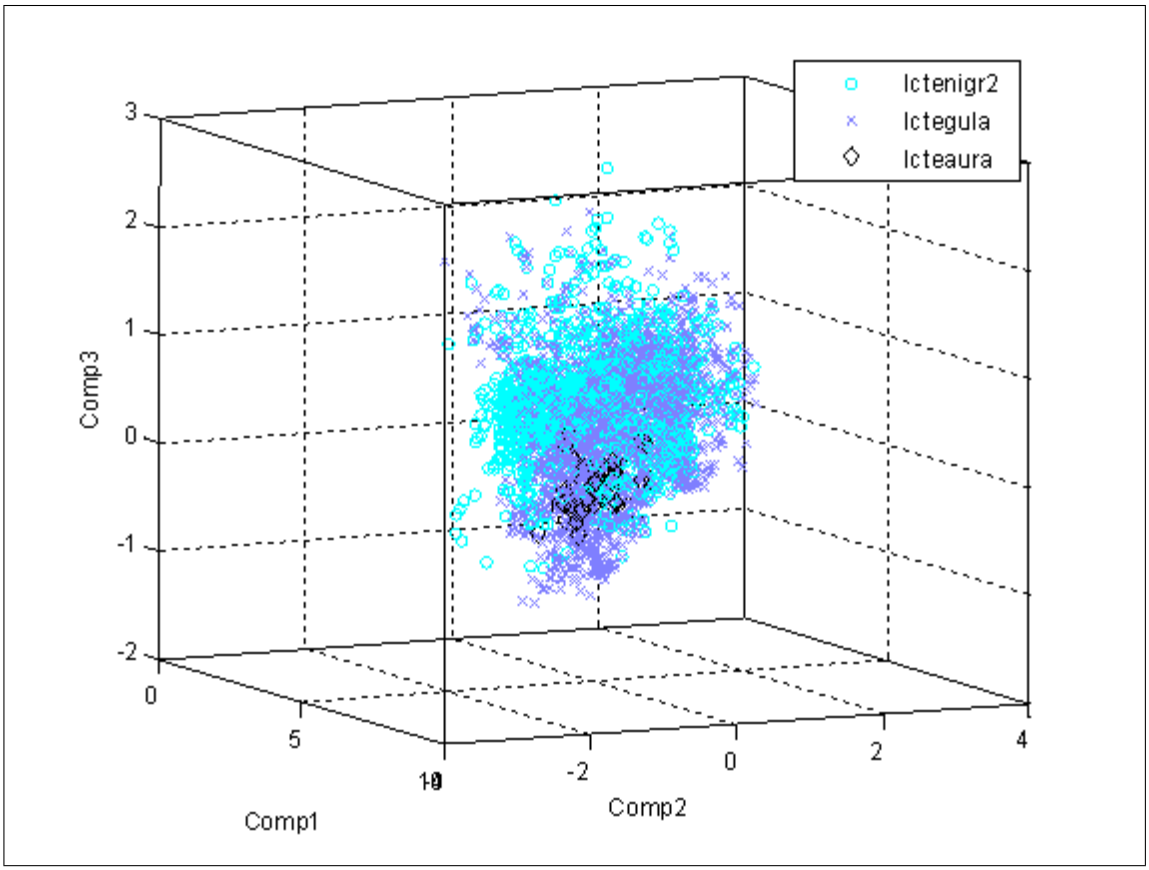


Figura C.8: Primeras tres componentes principales de las especies Ictenigr2, Ictegula e Icteaura.

Apéndice D

Funciones de preferencia estimadas.

.

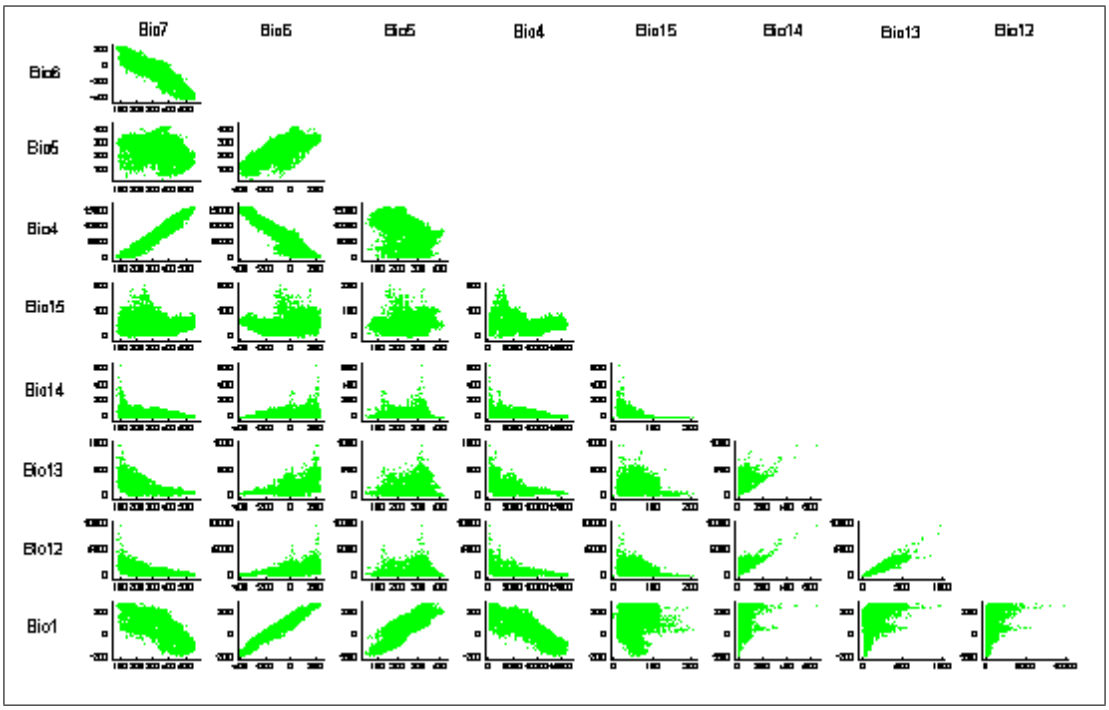


Figura D.1: Diagrama de dispersión de las variables a pares del espacio ecológico.

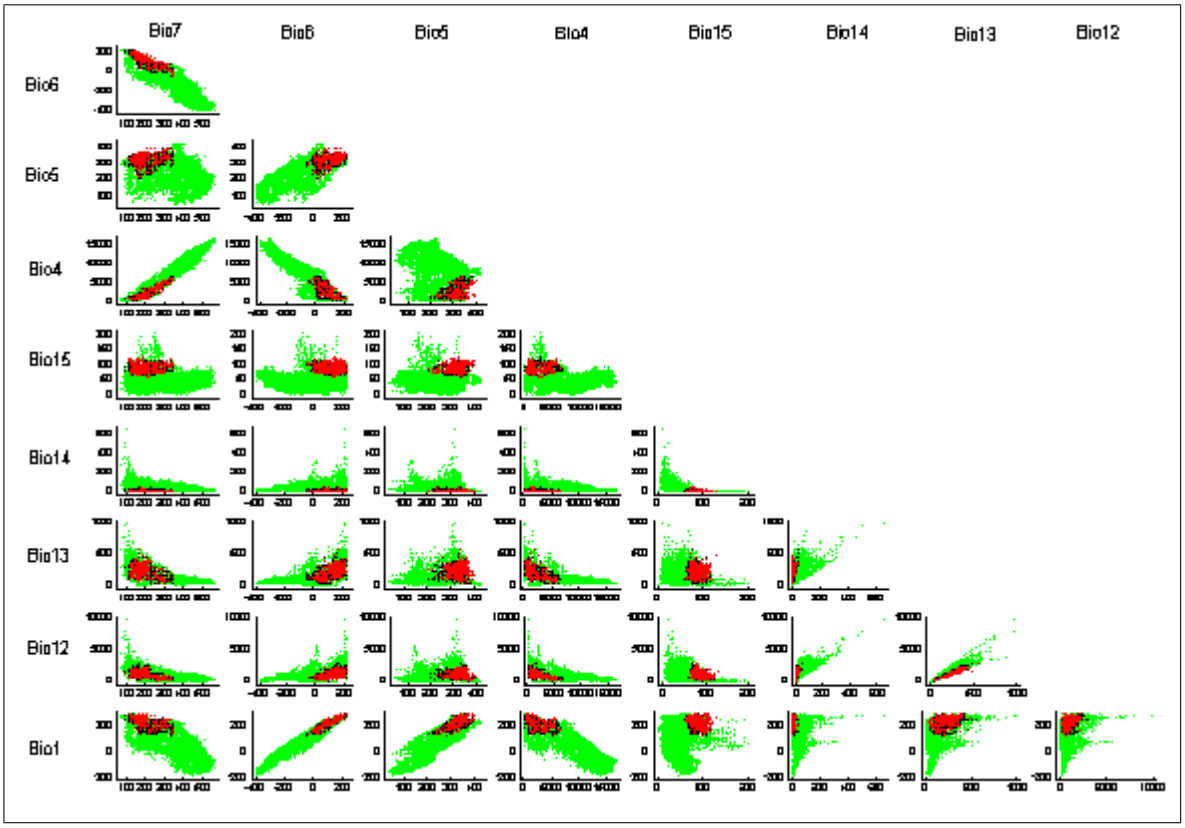


Figura D.2: Diagramas de dispersión del nicho estimado por GARP de la especie Ictepust (color rojo y en verde la parte de no nicho).

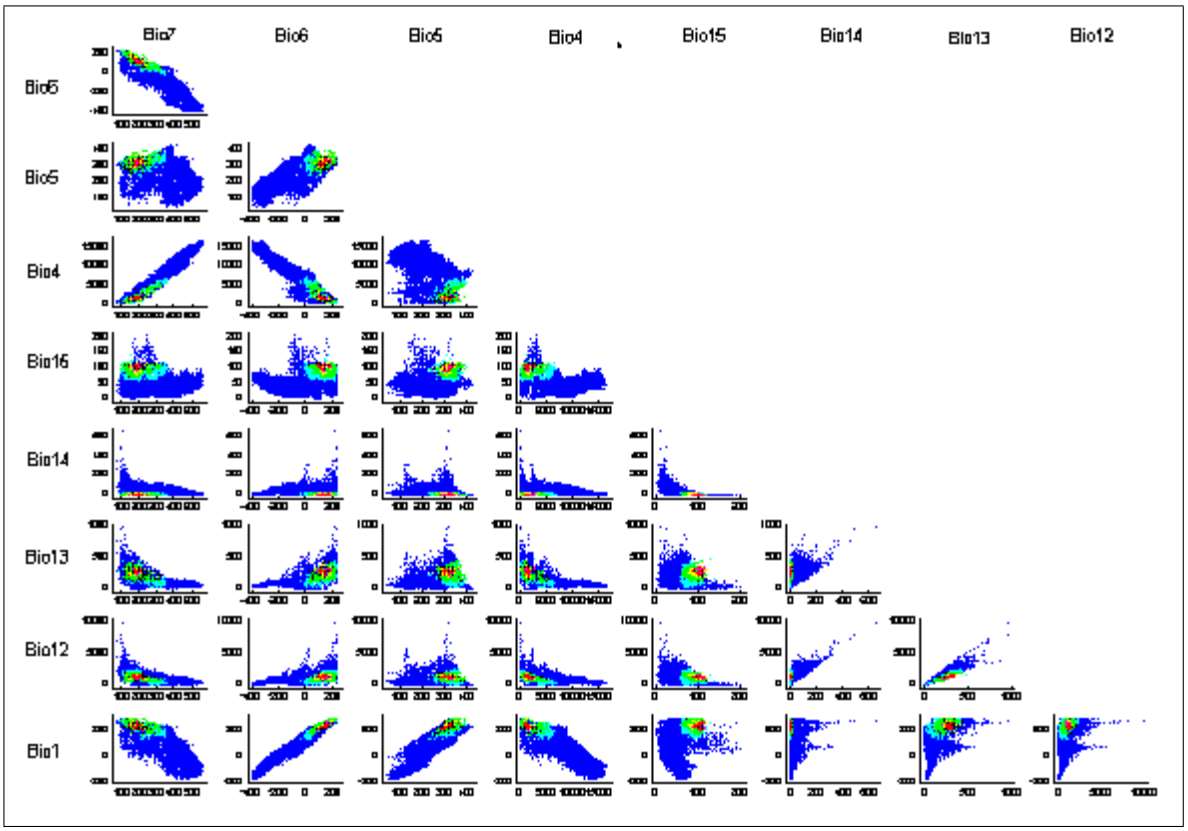


Figura D.3: Diagramas de dispersión de los puntos (configuraciones ambientales) de nivel de la función de preferencia estimada para la especie Ictepust. Los puntos rojos son los más altos al evaluar la función de probabilidad estimada, el siguiente nivel son los puntos magenta, seguido de éstos los amarillos, después los puntos verdes fuerte, en seguida los puntos verdes claro (son casi cero al evaluar la función), y finalmente, los puntos azules los cuales representan a los puntos más bajos, el valor de la función de preferencia en un punto azul es prácticamente cero.

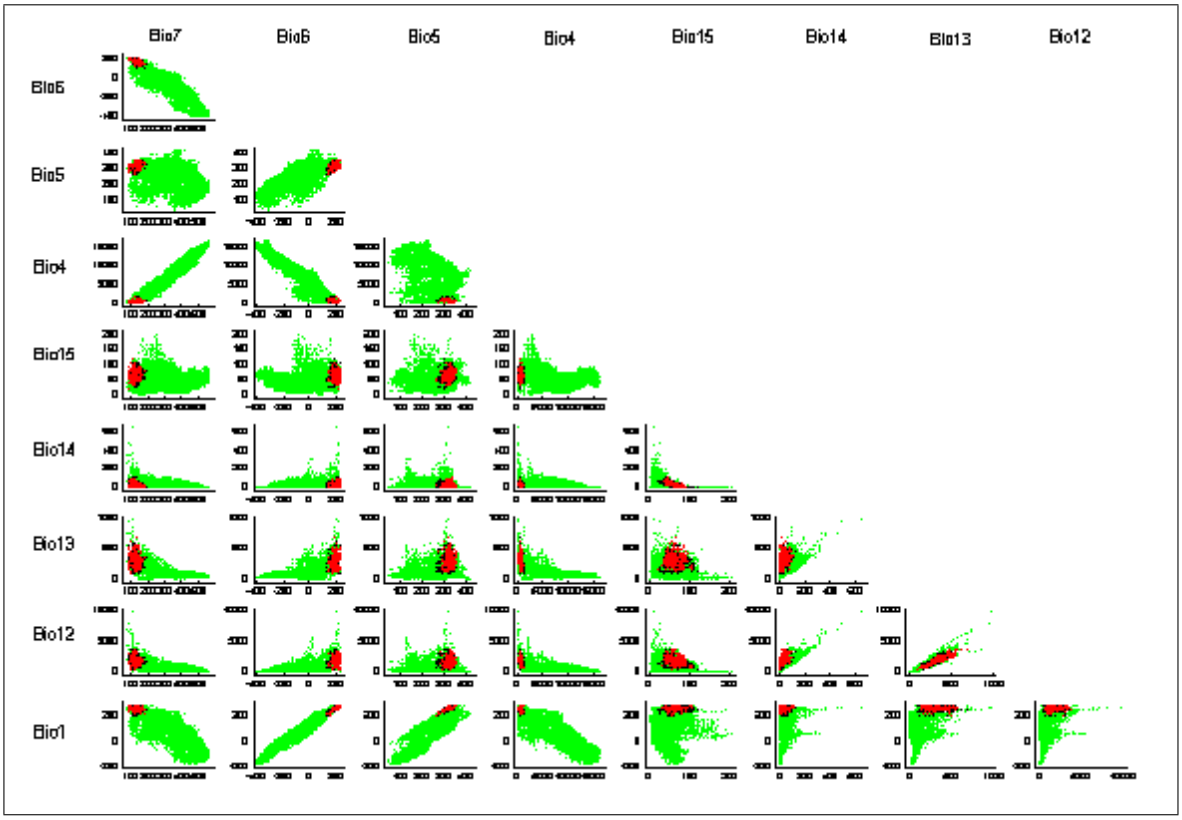


Figura D.4: Diagramas de dispersión del nicho estimado por GARP de la especie Ictenigr2 (color rojo y en verde la parte de no nicho).

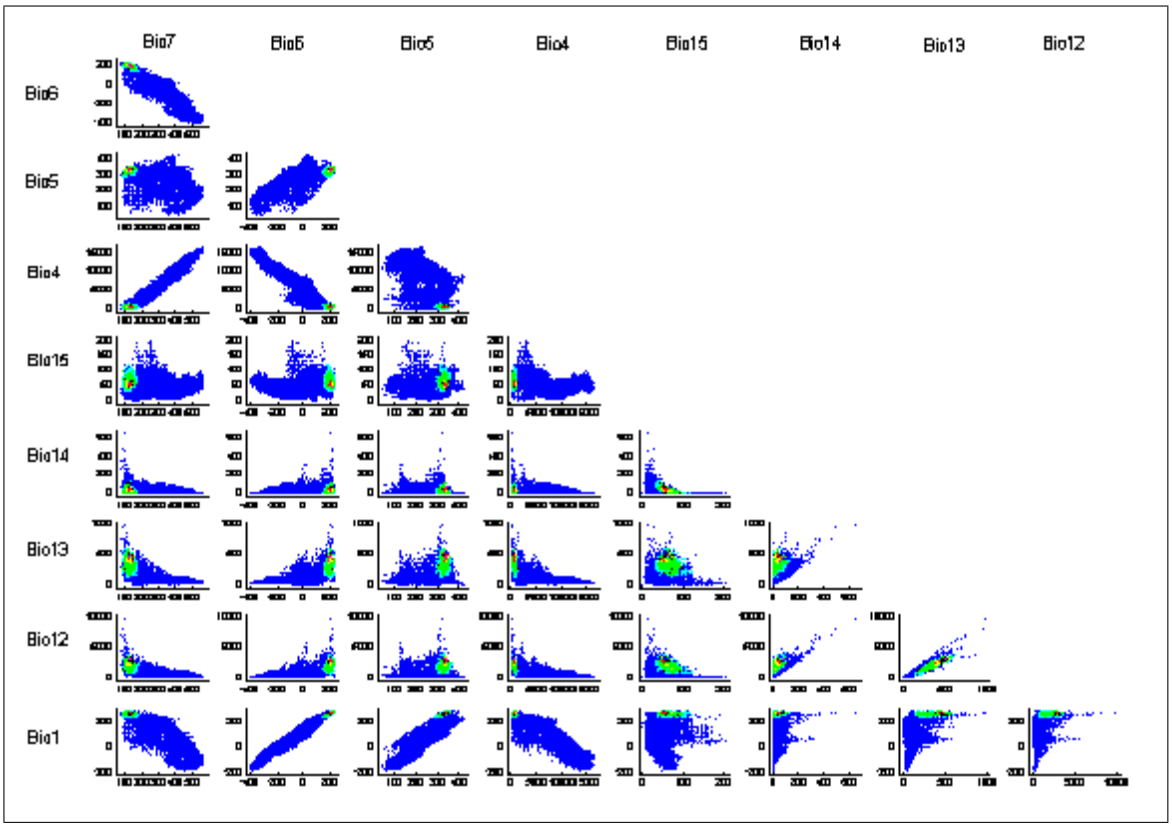


Figura D.5: Diagramas de dispersión de los puntos (configuraciones ambientales) de nivel de la función de preferencia estimada para la especie Ictenigr2. Los puntos rojos son los más altos al evaluar la función de probabilidad estimada, el siguiente nivel son los puntos magenta, seguido de éstos los amarillos, después los puntos verdes fuerte, en seguida los puntos verdes claro (son casi cero al evaluar la función), y finalmente, los puntos azules los cuales representan a los puntos más bajos, el valor de la función de preferencia en un punto azul es prácticamente cero.

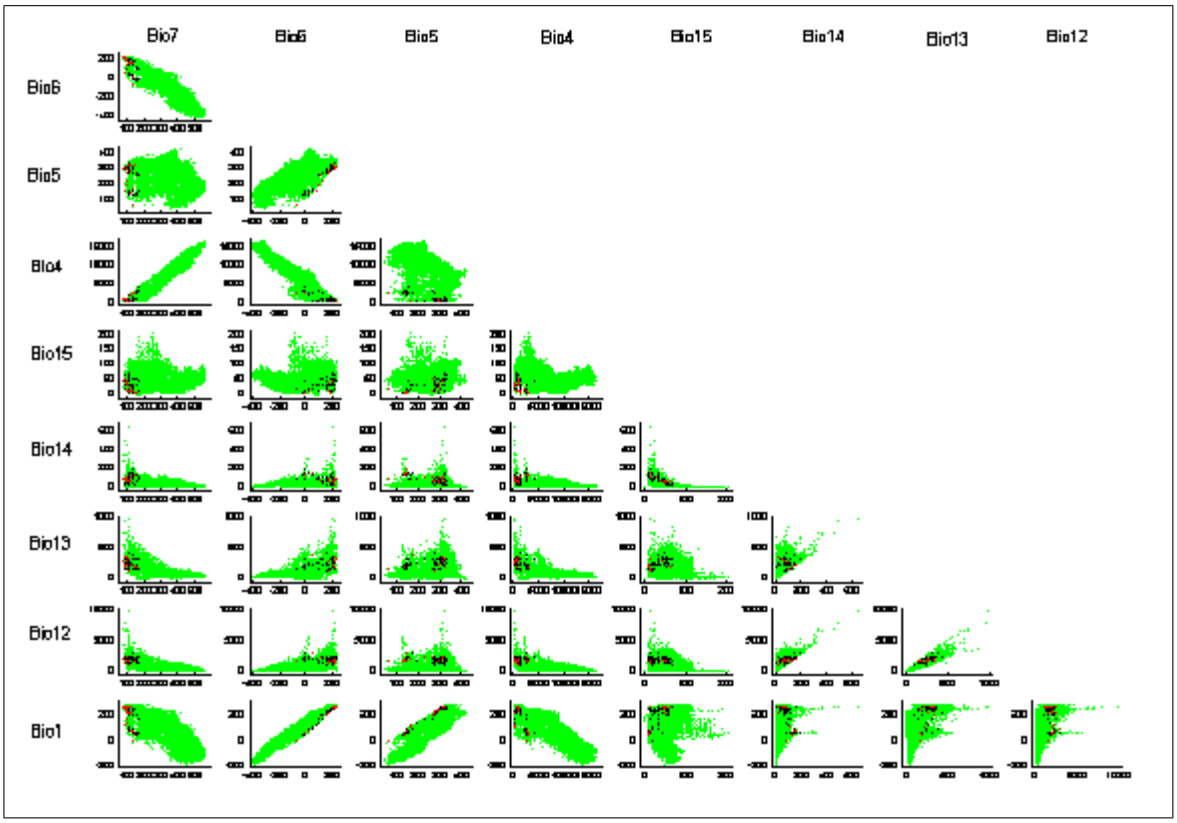


Figura D.6: Diagramas de dispersión del nicho estimado por GARP de la especie Icteleuc (color rojo y en verde la parte de no nicho).

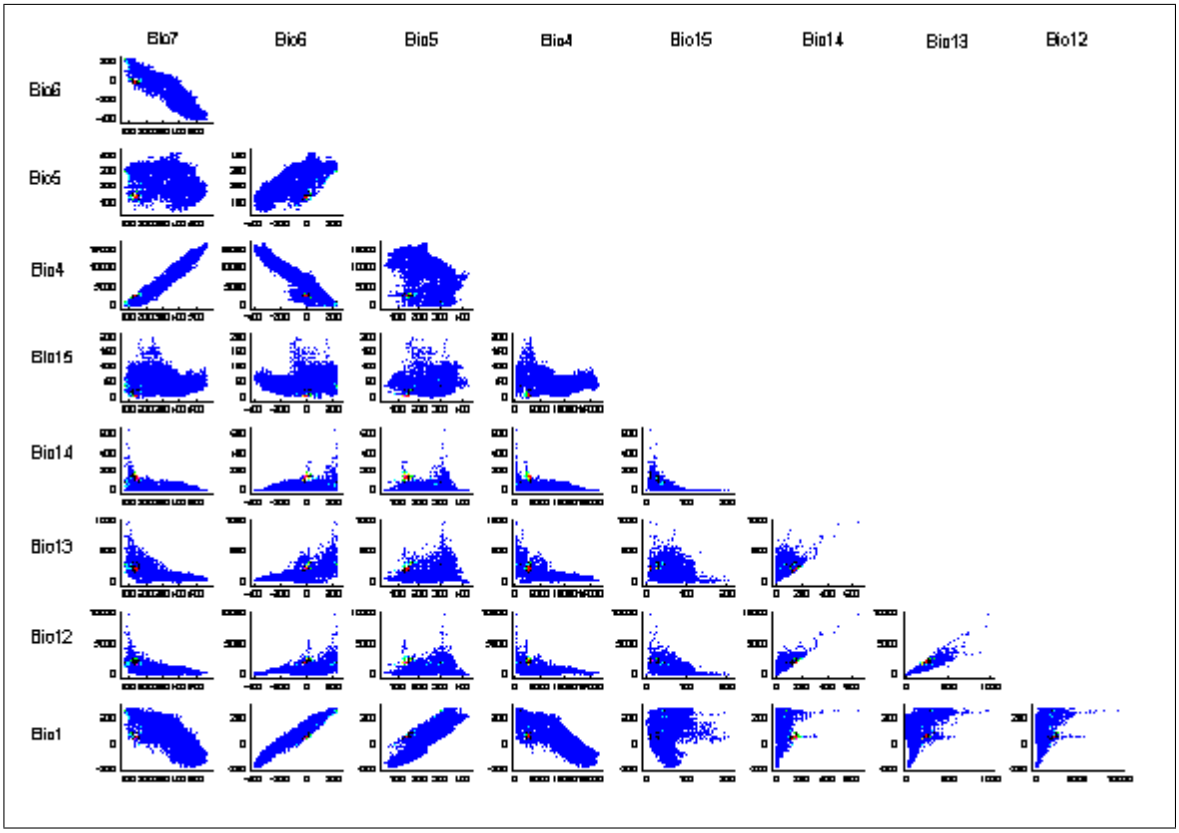


Figura D.7: Diagramas de dispersión de los puntos (configuraciones ambientales) de nivel de la función de preferencia estimada para la especie Icteleuc. Los puntos rojos son los más altos al evaluar la función de probabilidad estimada, el siguiente nivel son los puntos magenta, seguido de éstos los amarillos, después los puntos verdes fuerte, en seguida los puntos verdes claro (son casi cero al evaluar la función), y finalmente, los puntos azules los cuales representan a los puntos más bajos, el valor de la función de preferencia en un punto azul es prácticamente cero.

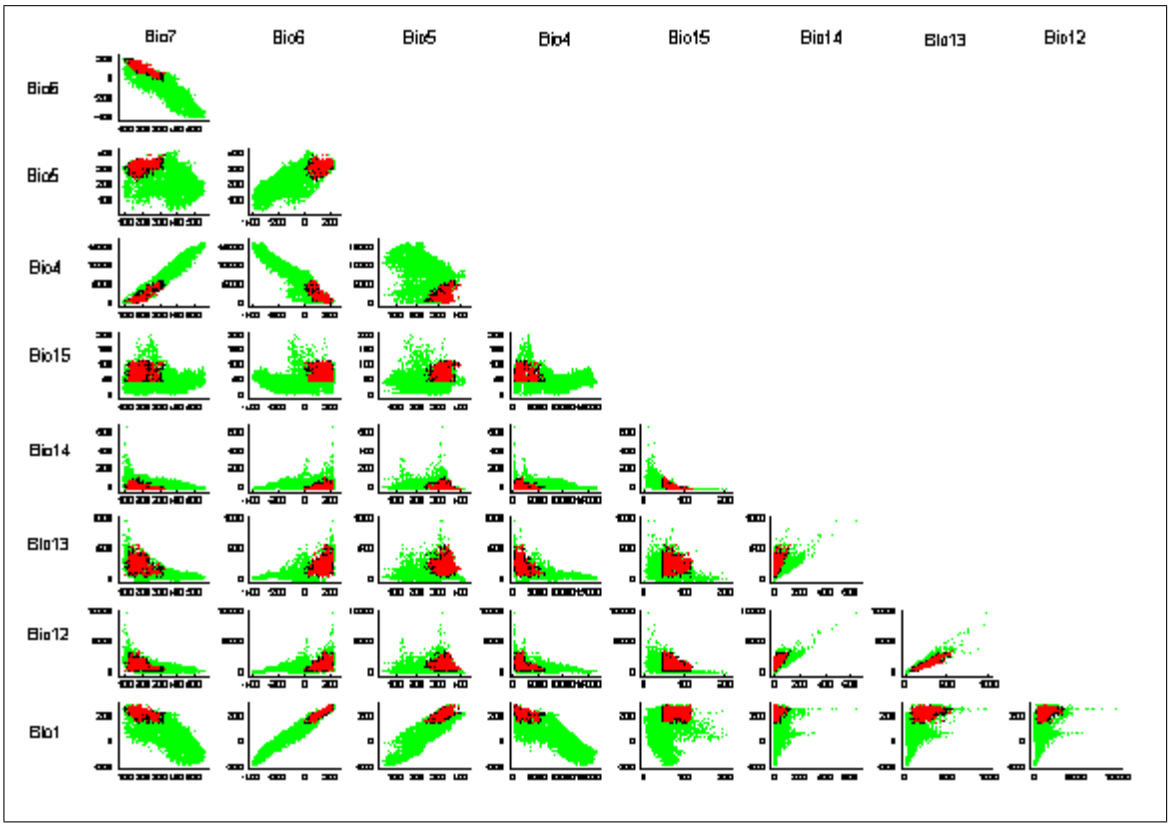


Figura D.8: Diagramas de dispersión del nicho estimado por GARP de la especie Ictegula (color rojo y en verde la parte de no nicho).

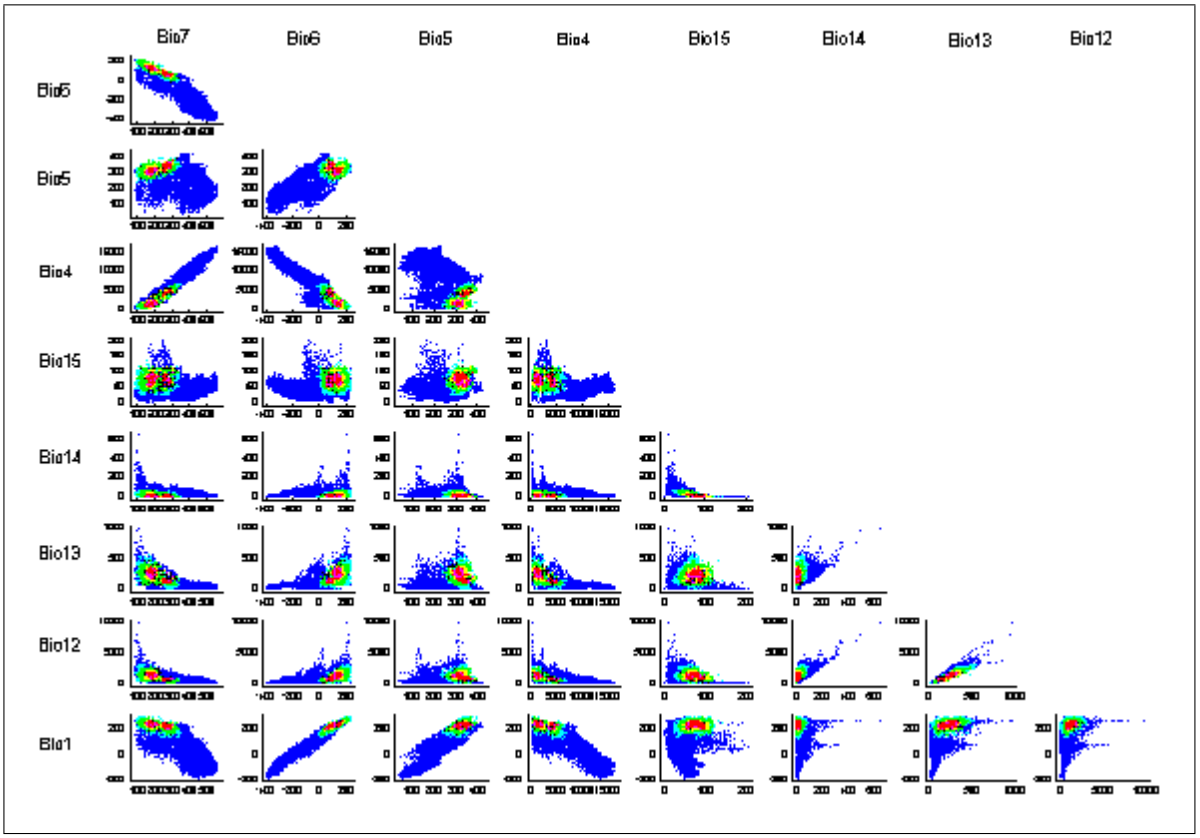


Figura D.9: Diagramas de dispersión de los puntos (configuraciones ambientales) de nivel de la función de preferencia estimada para la especie Ictegula. Los puntos rojos son los más altos al evaluar la función de probabilidad estimada, el siguiente nivel son los puntos magenta, seguido de éstos los amarillos, después los puntos verdes fuerte, en seguida los puntos verdes claro (son casi cero al evaluar la función), y finalmente, los puntos azules los cuales representan a los puntos más bajos, el valor de la función de preferencia en un punto azul es prácticamente cero.

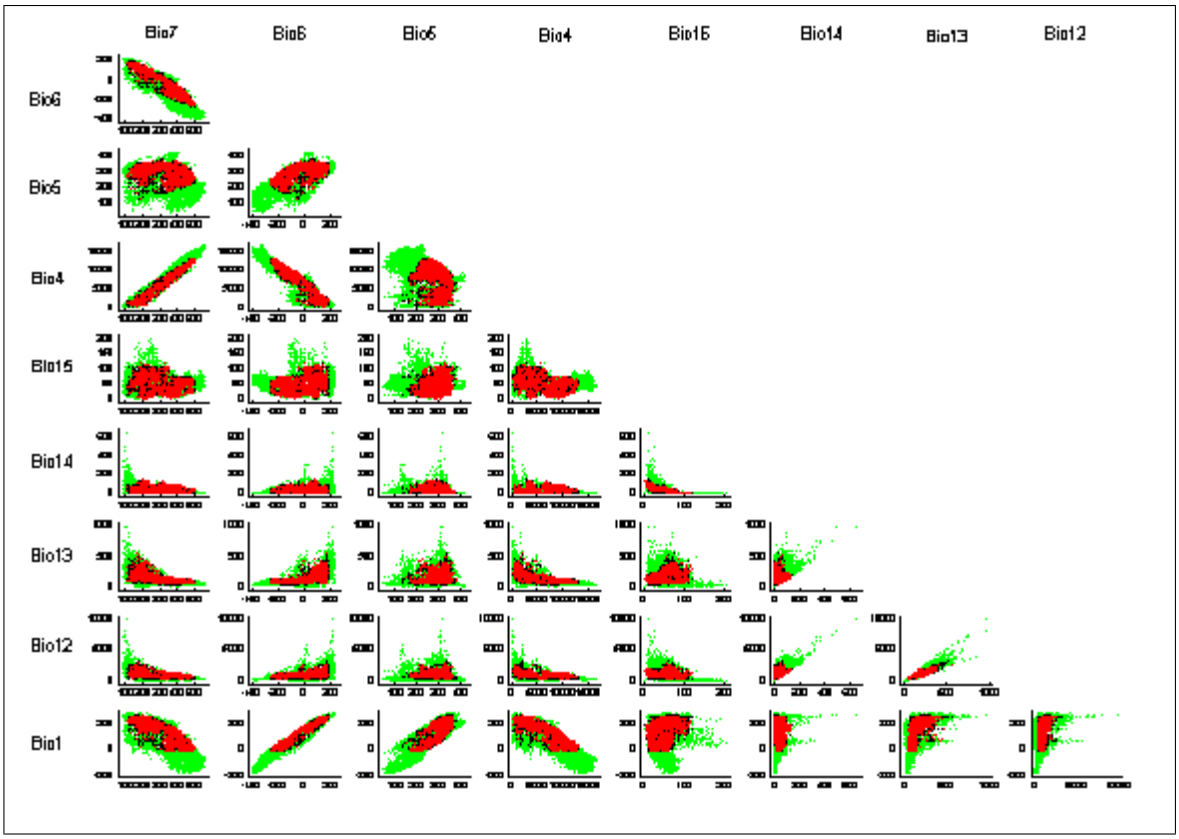


Figura D.10: Diagramas de dispersión del nicho estimado por GARP de la especie Ictegalb (color rojo y en verde la parte de no nicho).

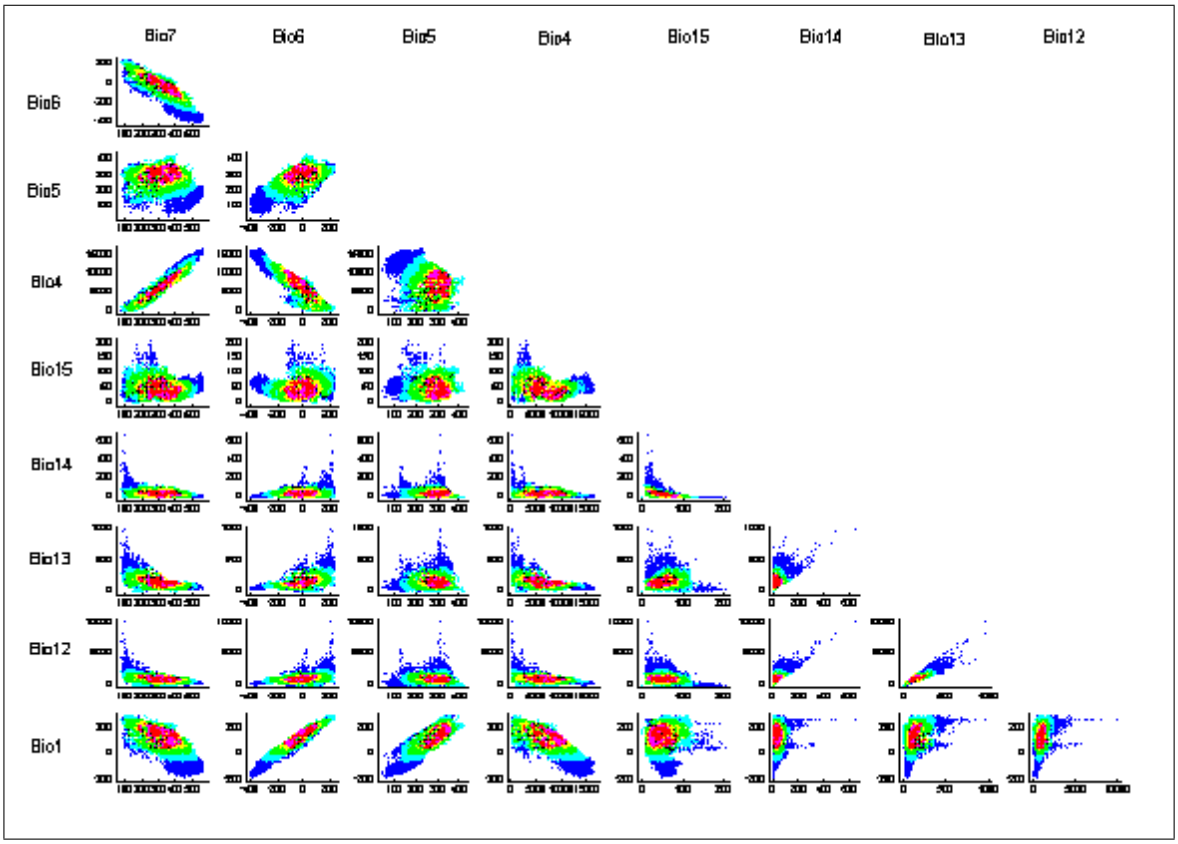


Figura D.11: Diagramas de dispersión de los puntos (configuraciones ambientales) de nivel de la función de preferencia estimada para la especie Ictegalb. Los puntos rojos son los más altos al evaluar la función de probabilidad estimada, el siguiente nivel son los puntos magenta, seguido de éstos los amarillos, después los puntos verdes fuerte, en seguida los puntos verdes claro (son casi cero al evaluar la función), y finalmente, los puntos azules los cuales representan a los puntos más bajos, el valor de la función de preferencia en un punto azul es prácticamente cero.

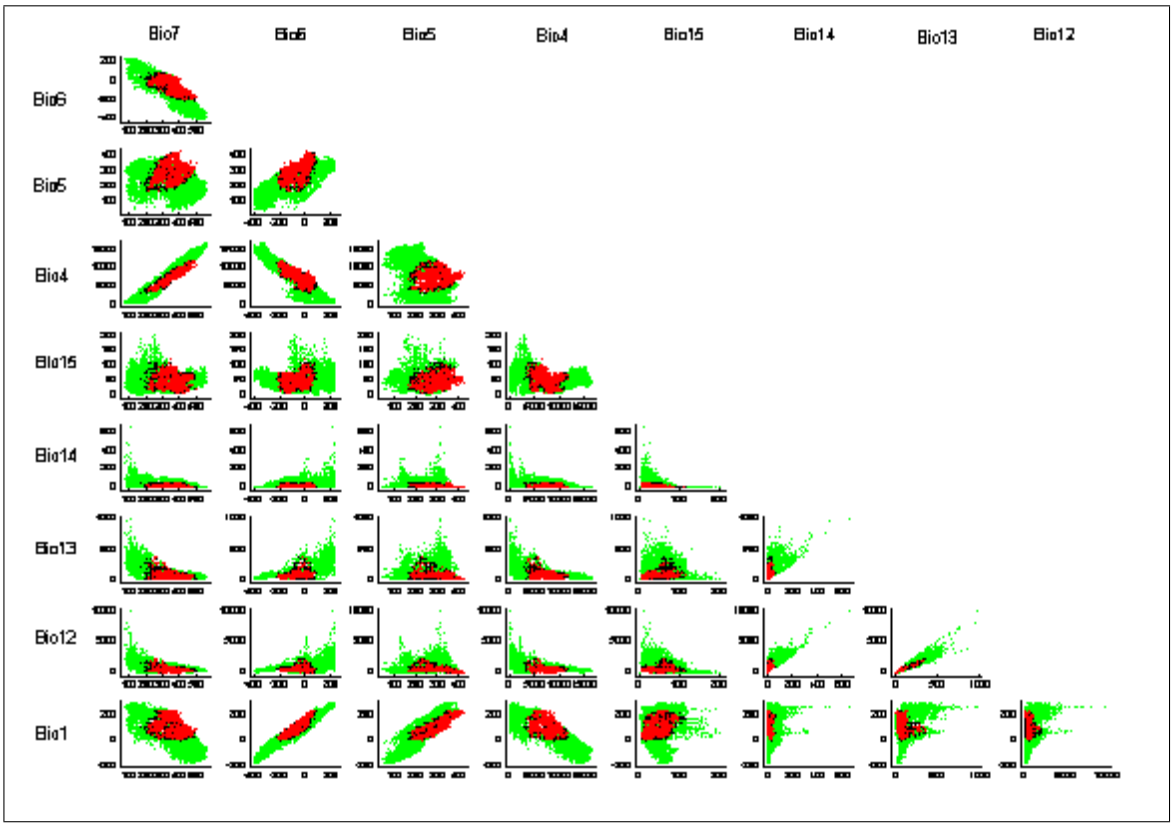


Figura D.12: Diagramas de dispersión del nicho estimado por GARP de la especie Ictebull (color rojo y en verde la parte de no nicho).

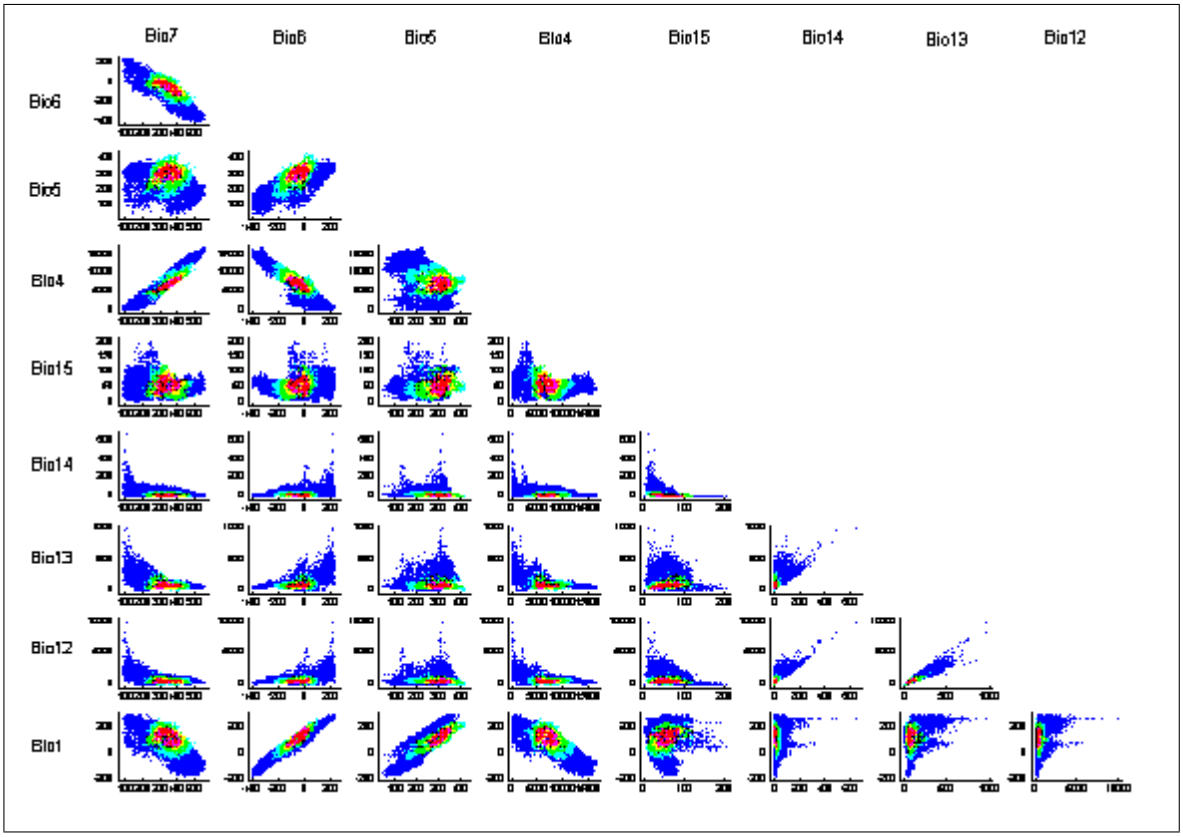


Figura D.13: Diagramas de dispersión de los puntos (configuraciones ambientales) de nivel de la función de preferencia estimada para la especie Ictebull. Los puntos rojos son los más altos al evaluar la función de probabilidad estimada, el siguiente nivel son los puntos magenta, seguido de éstos los amarillos, después los puntos verdes fuerte, en seguida los puntos verdes claro (son casi cero al evaluar la función), y finalmente, los puntos azules los cuales representan a los puntos más bajos, el valor de la función de preferencia en un punto azul es prácticamente cero.

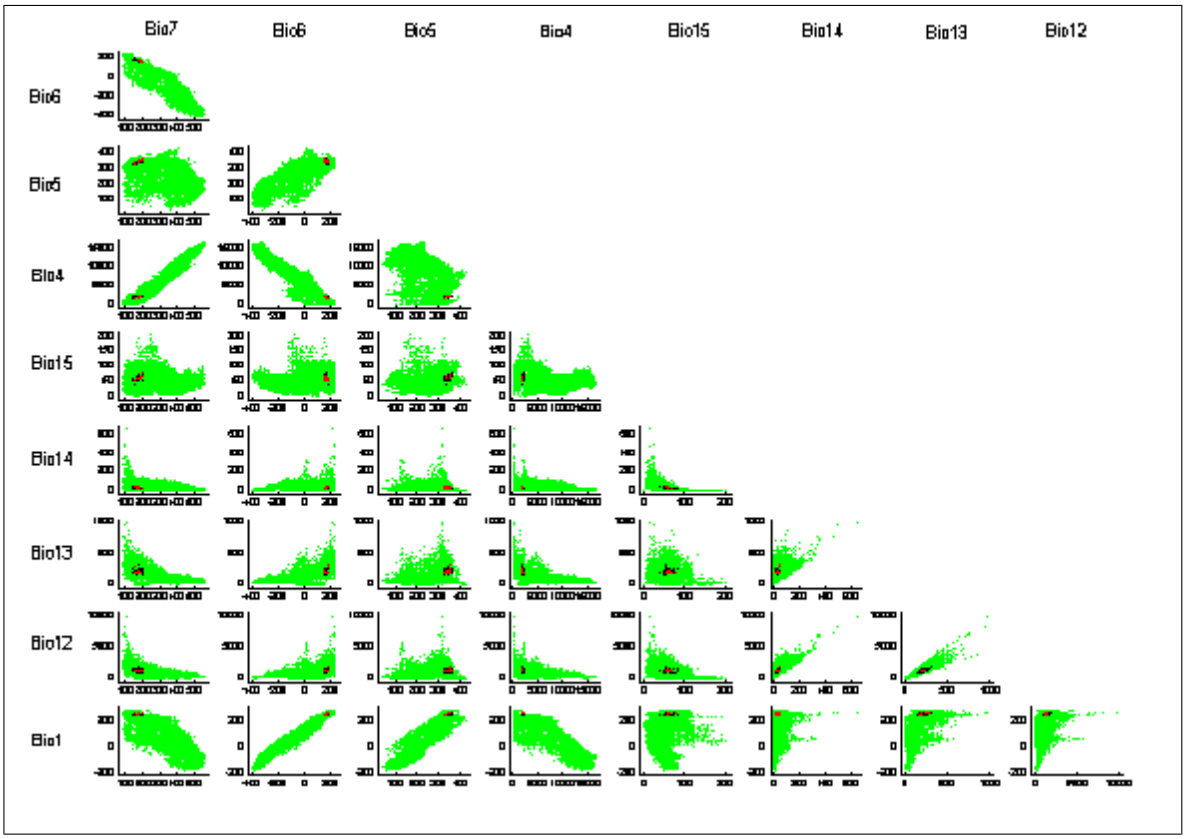


Figura D.14: Diagramas de dispersión del nicho estimado por GARP de la especie Icteaura (color rojo y en verde la parte de no nicho).

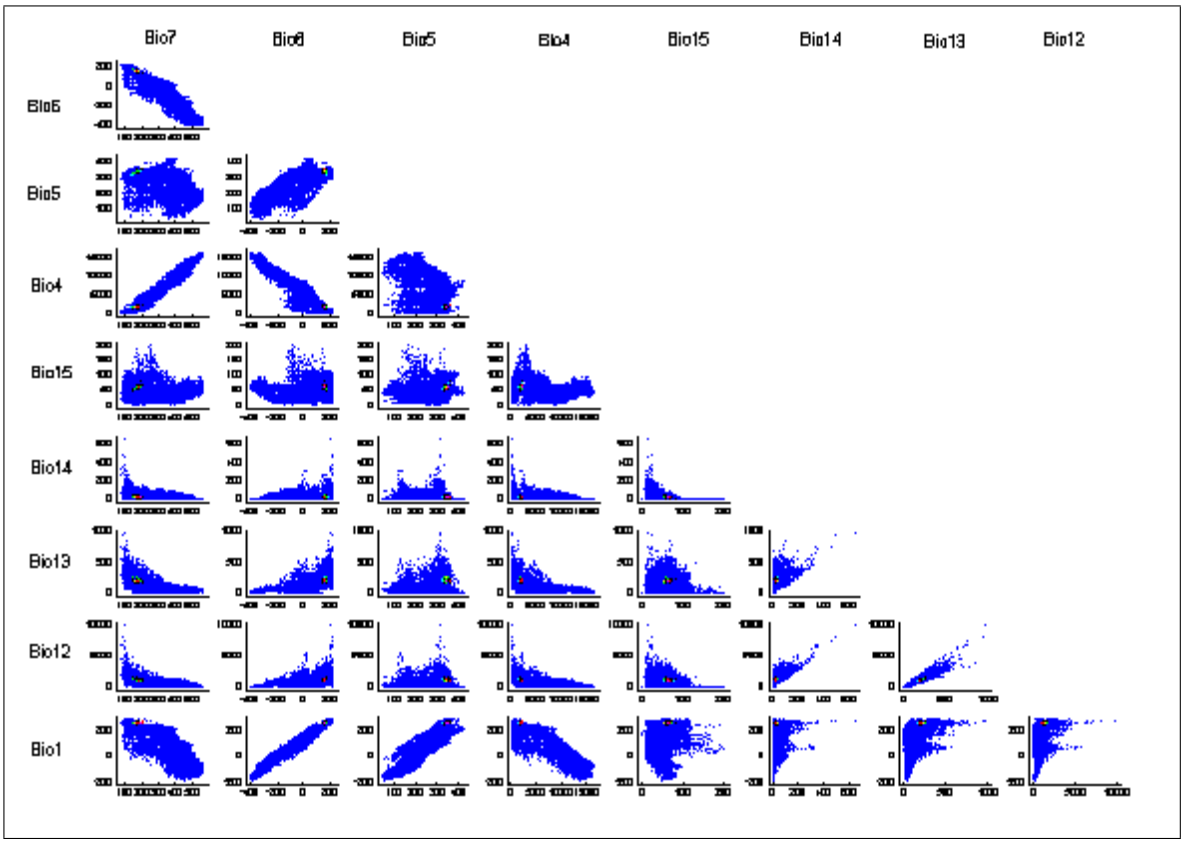


Figura D.15: Diagramas de dispersión de los puntos (configuraciones ambientales) de nivel de la función de preferencia estimada para la especie Icteaura. Los puntos rojos son los más altos al evaluar la función de probabilidad estimada, el siguiente nivel son los puntos magenta, seguido de éstos los amarillos, después los puntos verdes fuerte, en seguida los puntos verdes claro (son casi cero al evaluar la función), y finalmente, los puntos azules los cuales representan a los puntos más bajos, el valor de la función de preferencia en un punto azul es prácticamente cero.

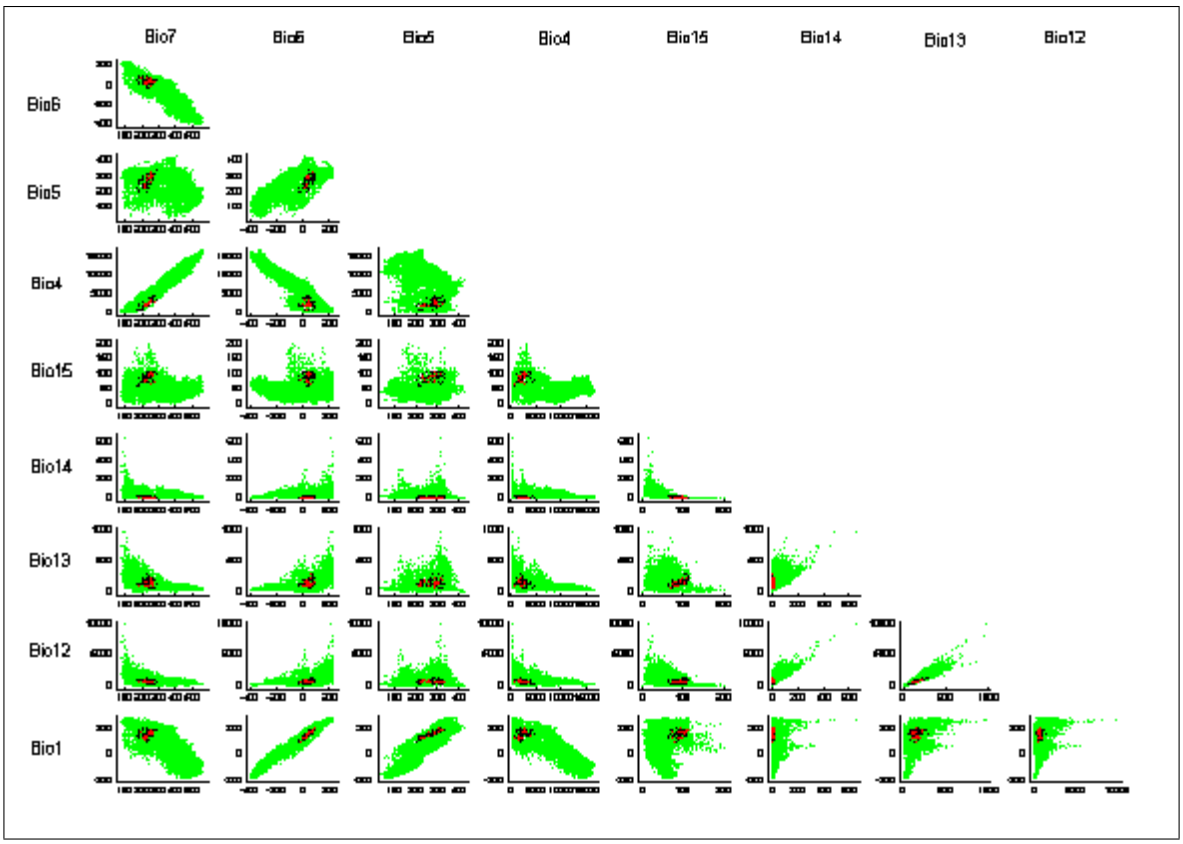


Figura D.16: Diagramas de dispersión del nicho estimado por GARP de la especie Icteabei (color rojo y en verde la parte de no nicho).

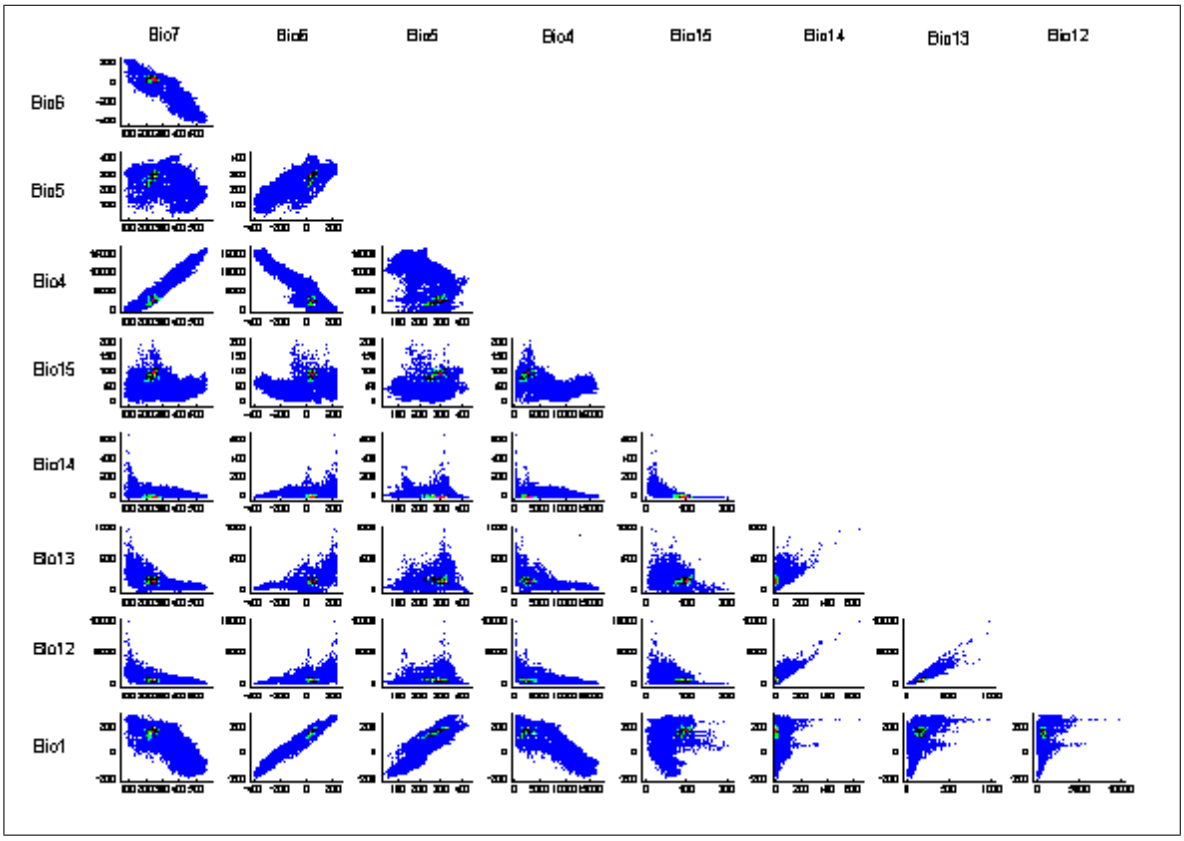


Figura D.17: Diagramas de dispersión de los puntos (configuraciones ambientales) de nivel de la función de preferencia estimada para la especie Icteabei. Los puntos rojos son los más altos al evaluar la función de probabilidad estimada, el siguiente nivel son los puntos magenta, seguido de éstos los amarillos, después los puntos verdes fuerte, en seguida los puntos verdes claro (son casi cero al evaluar la función), y finalmente, los puntos azules los cuales representan a los puntos más bajos, el valor de la función de preferencia en un punto azul es prácticamente cero.