



Universidade de Vigo

Trabajo Fin de Máster

---

# Métodos estadísticos para la estimación del área de actividad de un animal: home range.

---

Andrea Vilar Álvarez

Máster en Técnicas Estadísticas

Curso 2018-2019



## Propuesta de Trabajo Fin de Máster

|   |
|---|
| <p><b>Título en galego:</b> Métodos estadísticos para a estimación da área de actividade dun animal: home range.</p>  |
| <p><b>Título en español:</b> Métodos estadísticos para la estimación del área de actividad de un animal: home range.</p>  |
| <p><b>English title:</b> Statistic methods to estimate the area traversed by animals in its normal activities: home range.</p>  |
| <p><b>Modalidad:</b> Modalidad A</p>  |
| <p><b>Autora:</b> Andrea Vilar Álvarez, Universidade de Santiago de Compostela</p>  |
| <p><b>Directora:</b> Beatriz Pateiro López, Universidade de Santiago de Compostela</p>  |
| <p><b>Breve resumen del trabajo:</b></p> <p>El área de actividad de un animal (home range) es el área atravesada por un individuo durante sus actividades normales de recolección de alimento, apareamiento, y cuidado de crías (Burt, 1943). Desde esta primera definición, el concepto de área de actividad ha evolucionado, dando lugar a una cantidad considerable de literatura sobre el tema, véase por ejemplo la revisión de Powell (2000). La estimación del área de actividad de un animal se suele realizar a partir de un conjunto de localizaciones del mismo recogidas durante un período de tiempo. Son varias las metodologías propuestas en la literatura para la estimación del área de actividad.</p> <p>El objetivo de este trabajo es que el alumno haga una revisión de las distintas metodologías existentes para la estimación del área de actividad y realice una comparativa aplicada a datos reales.</p> |
| <p><b>Recomendaciones:</b> Es recomendable que el alumno tenga experiencia en programación con R.</p>   |
| <p><b>Otras observaciones:</b></p>  |



Doña Beatriz Pateiro López, Profesora contratada doctora de la Universidade de Santiago de Compostela, informa que el Trabajo Fin de Máster titulado

**Métodos estadísticos para la estimación del área de actividad de un animal: home range.**

fue realizado bajo su dirección por doña Andrea Vilar Álvarez para el Máster en Técnicas Estadísticas. Estimando que el trabajo está terminado, da su conformidad para su presentación y defensa ante un tribunal.

En Santiago de Compostela, a 5 de septiembre de 2019.

La directora:

La autora:

Doña Beatriz Pateiro López

Doña Andrea Vilar Álvarez



# Agradecimientos

En primer lugar me gustaría agradecer a mi directora de TFM, Beatriz Pateiro López, toda la dedicación y entusiasmo que ha puesto en este trabajo, y todo el apoyo y confianza que me ha brindado durante estos últimos meses.

Del mismo modo, quisiera agradecer al investigador Bruce D. Patterson del Museo de Historia Natural de Chicago, por sus esfuerzos en el estudio de la evolución de los ecosistemas, en particular, por los datos recogidos sobre las localizaciones de los leones en el Parque Natural de Tsavo, en Kenia, que amablemente ha cedido para propósitos académicos posibilitando así su aparición en este trabajo. También al profesor Roger Bivand, de la Universidad de Bergen (Noruega), por sus aclaraciones y sus consejos en la utilización del paquete `rgeos` de R [Bivand and Rundel, 2019] a la hora de tratar con coordenadas geográficas. Ha sido realmente útil.

No quisiera olvidarme del profesor Alberto Rodríguez Casal, al que me gustaría agradecer su pasión por la enseñanza y por la estadística, con la que ha conseguido inculcarme la curiosidad y meticulosidad que tan importantes son en matemáticas. Sin sus consejos probablemente mi vocación estadística y este TFM no existirían.

Quiero agradecer también a mis compañeros de ITMATI (Instituto Tecnológico de Matemática Industrial) todo lo que me enseñan día a día y la paciencia y empatía con la que lo hacen: Fito, Rabi, Jorge, Dani, Javi, Damián, Gabriel Capeáns, Diego, Gabriel Álvarez y Óscar. Me siento enormemente afortunada de formar parte de vuestro equipo.

Y para finalizar, me gustaría dar las gracias a mis amigas de toda la vida, en especial a Inés y Lidia pero también a las nuevas incorporaciones, Cris y María, porque a menudo el apoyo moral es tan importante como el científico.





# Índice general

|  |           |
|--|-----------|
| <b>Resumen</b>   | <b>XI</b> |
| <b>1. Definición del problema de estimación del home range</b>                         | <b>1</b>  |
| <b>2. Estimación del home range con datos independientes</b>                           | <b>3</b>  |
| 2.1. Métodos globales para datos independientes . . . . .                              | 3         |
| 2.1.1. Cierre convexo . . . . .  | 3         |
| 2.1.2. Cierre $\alpha$ -convexo . . . . .  | 6         |
| 2.1.3. Estimación a través de la función de utilización . . . . .                      | 6         |
| 2.2. Métodos localizados para datos independientes . . . . .                           | 12        |
| 2.2.1. Cierre convexo local (LoCoH) o k-vecino más próximo del cierre convexo (k-NNCH) | 12        |
| 2.2.2. Cierre poligonal característico (CHP) . . . . .                                 | 13        |
| 2.2.3. Single-linkage cluster . . . . .  | 14        |
| <b>3. Estimación del home range con datos dependientes</b>                             | <b>15</b> |
| 3.1. Métodos globales para datos dependientes . . . . .                                | 15        |
| 3.1.1. Estimación de la densidad geográfica-temporal (TGDE) . . . . .                  | 15        |
| 3.1.2. Modelo de movimiento del puente browniano (BBMM) . . . . .                      | 18        |
| 3.1.3. Modelo de movimiento browniano reflejado (MRBM) . . . . .                       | 20        |
| 3.2. Métodos localizados para datos dependientes . . . . .                             | 21        |
| 3.2.1. T-LoCoH . . . . .   | 21        |
| <b>4. Estimación del home range con datos simulados</b>                                | <b>23</b> |
| 4.1. Simulación de observaciones uniformes en distintos conjuntos . . . . .            | 23        |
| 4.2. Simulación de observaciones aleatorias en un disco . . . . .                      | 35        |
| <b>5. Aplicación sobre un conjunto de datos real</b>                                   | <b>45</b> |
| 5.1. Estudio de la especie y su hábitat . . . . .                                      | 46        |
| 5.2. Estimación del home range . . . . .   | 48        |
| <b>6. Conclusiones</b>   | <b>57</b> |



# Resumen

## Resumen en español

El término home range se utiliza para designar el área de actividad de un animal, es decir, la porción de terreno en la que un animal realiza sus actividades del día a día: alimentación, apareamiento, cuidado de crías...

Desde el punto de vista biológico, conocer el home range de una especie supone conocerla más profundamente, lo que facilita la gestión de recursos y en general, la anticipación a situaciones adversas, por lo que es un concepto con un enorme interés biológico.

El propósito de este trabajo es revisar las técnicas existentes para la estimación del home range. Para ello, se seguirá el esquema utilizado en [Baíllo and Chacón, 2018]. Seguidamente se analizará el funcionamiento de cada técnica sobre un conjunto de datos simulados y finalmente, se hará una estimación del home range con observaciones reales.

Así, en el primer capítulo se hace una breve introducción al problema de estimación del home range, cuya resolución veremos que puede afrontarse desde distintas perspectivas.

Los Capítulos 2 y 3 están dedicados a la exposición pormenorizada de las técnicas disponibles en la actualidad, para datos independientes y datos dependientes, respectivamente.

En el Capítulo 4 se realizará un estudio de simulación para evaluar la calidad de cada estimador y el Capítulo 5 está dedicado a la estimación del home range de un león de la especie *Panthera leo* de la Reserva Natural del Rio Tsavo en Kenia.

Finalmente las conclusiones y futuras líneas de trabajo aparecen en el Capítulo 6.

## English abstract

The word home range is used to describe the area traversed by animals in their normal activities such as food gathering, mating, caring for young...

Since the biological point of view, knowing the home range of a specie means knowing the specie in detail and this fact helps to manage resources and, in general, helps to anticipate adverse situations, so this is a term with an enormous biological interest.

The aim of this document is revise the techniques used to estimate the home range. For this purpose, we follow the same structure as [Baíllo and Chacón, 2018]. Then, we analyse how each technique works using simulated data and finally, we estimate a real home range using real observations.

In the first chapter, we make a short introduction to the problem of estimate the home range, whose resolution can be view from different perspectives.

Chapters 2 and 3 deals with the detailed exposition of all the techniques used nowadays for independent and dependent data respectively.

At Chapter 4 we make a simulation study to evaluate the performance of each estimator and finally, at Chapter 5 we make the estimation of a real home range of a lion of the specie *Panthera leo* from Tsavo National Park in Kenya.

Finally, we make a brief conclusion including future research lines at Chapter 6.

# Capítulo 1

## Definición del problema de estimación del home range

Los restos humanos más antiguos que se conocen, fueron descubiertos en Etiopía y datan de hace 2.8 millones de años. Con los datos disponibles a día de hoy, todo apunta a que la especie humana surgió en África y allí habitó un largo periodo de tiempo hasta que, por factores principalmente medioambientales, fue desplazándose a otros continentes.

Más recientemente, el naturalista inglés Charles Darwin apuntaba en su libro “El origen de las especies”, [Darwin, 1861], que los movimientos de los animales no eran aleatorios, sino que se restringían a una porción de terreno y que, como ocurrió con los humanos, los desplazamientos fuera de esta área delimitada, eran ocasionados por transformaciones en el medio que de alguna forma, empujaban a la especie a marcharse a otros lugares.

A raíz de esta observación, surgió un creciente interés por estudiar esta porción de área, sus características y su utilidad a la hora de proporcionar información sobre las especies que la habitan y su entorno.

El zoólogo estadounidense Henry Burt fue más allá, y en 1943 dio la primera definición de la porción de área a la que Darwin había hecho referencia, es lo que se denominó como **home range** o **área de actividad de un animal** [Burt, 1943].

Burt la definió como **el área atravesada por un individuo en sus actividades normales en busca de alimento, apareamiento y cuidado de las crías. Las salidas ocasionales de esta área, quizás en jornadas de exploración, no se considerarían parte del home range.** Además, hizo especial hincapié en la importancia de distinguir entre *home range* y *territory* (territorio), pues el territorio sería la porción dentro del home range que el animal defiende.

A partir de esta definición, resulta inmediato pensar que el home range no permanece constante durante toda la vida del animal, sino que puede variar con el tiempo por distintas razones:

- **Edad:** en algunas especies el home range del animal cambia notablemente con la edad, pues individuos más jóvenes suelen tener un home range más variable y tiende a estabilizarse en la madurez debido a la presencia de crías.
- **Factores climáticos:** las especies migratorias tienen distinto home range según las diferentes estaciones del año y según la definición anterior, la ruta entre un home range y otro, no sería considerada parte del home range del animal.

En otras ocasiones, especies no migratorias adaptadas a unas condiciones climatológicas determinadas, se ven obligadas a hacer un cambio permanente de su área de actividad cuando estas condiciones se ven de alguna forma modificadas.

- **Presencia de depredadores o escasez de alimento:** el riesgo de depredación o la baja de disponibilidad de alimento ponen en riesgo la supervivencia de la especie, obligándola a cambiar su área de actividad para asegurar su vida.
- **Sexo:** en algunas especies, las hembras se separan del resto de la especie durante la época de cría, cambiando así su área de actividad.

Por estos motivos, conocer el home range de una especie puede facilitarnos datos sobre la edad de los individuos que la componen, las condiciones climáticas o medioambientales del entorno, la localización de los recursos, la aparición de nuevas especies... lo que evidencia el enorme interés biológico del concepto.

Los datos necesarios para estimar el home range de un animal se componen de un conjunto de puntos en el espacio que no son más que localizaciones del animal que han sido recogidas identificando su posición en los distintos instantes de tiempo. De este modo, cada observación será bidimensional (si sólo recoge los movimientos del animal en el plano) o tridimensional (cuando el animal no sólo cambia de posición en el plano, sino en el espacio).

Tradicionalmente los datos se recogían mediante observación directa o incluso valiéndose de trampas repartidas por la porción de terreno habitada por el animal en estudio. En la actualidad, el sistema de recolección de localizaciones ha mejorado enormemente gracias a los Sistemas de Posicionamiento Global (GPS) que transmiten la posición del animal en intervalos de tiempo regulares.

Este avance en la recolección muestral ha supuesto una evolución en los métodos de estimación del home range, pues originalmente se estimaba a partir de datos independientes pero con el tiempo, surgió la necesidad de adaptarse a la dependencia temporal. Esta característica además, da lugar a la existencia de autocorrelación entre las observaciones, lo que nos permite tomar ventaja de la información del pasado para predecir los futuros movimientos del animal, mejorando así la estimación del home range.

Por este motivo, el problema de estimación del home range no cuenta con una única solución, pudiendo afrontarse desde distintas perspectivas, desde la teoría de la probabilidad hasta la estadística no paramétrica pasando por la geometría y los procesos estocásticos, con métodos muy sencillos hasta otros más complejos.

Estos diferentes enfoques para la resolución del problema, han dado lugar a definiciones más técnicas del concepto. Por ejemplo, el home range puede verse como una área o como la utilización que un animal hace del espacio y puede describirse utilizando la *distribución de utilización*, es decir, distribución de probabilidad de las localizaciones del animal durante un periodo de tiempo, véase [van Winkle, 1975]. Este punto de vista, nos permite obtener el home range de un animal definiéndolo como un conjunto de nivel que contiene la densidad de la función de utilización con una probabilidad alta.

Como vemos, las posibilidades para la resolución del problema son numerosas y la elección de unas u otras dependerá del resultado que deseemos obtener, de la naturaleza de los datos disponibles para hacer la estimación e incluso del comportamiento que presente la especie en estudio dentro de su territorio.

La estructura que seguiremos en la exposición de las técnicas disponibles para la estimación del home range es la misma que se utiliza en [Baíllo and Chacón, 2018], aunque el contenido de cada sección ha sido ampliado y actualizado con los correspondientes artículos de referencia. Así mismo, el estudio del comportamiento de los distintos estimadores que se recoge en el Capítulo 4 y la estimación sobre un conjunto de datos reales del Capítulo 5, son originales.

## Capítulo 2

# Estimación del home range con datos independientes

En este capítulo se recoge una breve revisión bibliográfica de las técnicas disponibles para la estimación del área de actividad de un animal bajo la suposición de independencia de las observaciones. Expondremos por tanto las técnicas más tradicionales y veremos dos perspectivas diferenciadas para la resolución del problema: el enfoque geométrico y el enfoque estadístico (paramétrico y no paramétrico).

El primero se centra en la estimación del área mientras que el segundo permite además un conocimiento del uso que el animal hace de su área de actividad.

Los datos que se emplearán en este contexto para la estimación del home range serán datos bidimensionales, donde la primera coordenada indica la posición del animal con respecto a la horizontal y la segunda, con respecto a la vertical. Por tanto  $\mathbf{x} = (x, y)$ .

### 2.1. Métodos globales para datos independientes

En este apartado trataremos aquellas técnicas de estimación del home range que consideran que el conjunto muestral se compone de observaciones independientes. Además, los métodos expuestos serán **técnicas globales** en el sentido de que utilizan el conjunto de datos en su totalidad para realizar la estimación.

#### 2.1.1. Cierre convexo

Fue uno de los primeros métodos utilizados para la estimación del área de actividad y su uso se extendió rápidamente debido a su simplicidad, ya que es un procedimiento geométrico que no requiere ningún conocimiento estadístico y que incluso puede realizarse a mano.

Desde el punto de vista matemático, un **polígono convexo** es aquel cuyos ángulos internos no exceden los  $180^\circ$ . Se dice que es mínimo porque es la mínima área convexa del polígono que contiene a todos los puntos, véase [Worton, 1987].

Así, para calcular el cierre convexo o mínimo polígono convexo (MCP) de un conjunto de puntos, basta con unir los puntos periféricos de la muestra mediante algún algoritmo. Existen varias opciones en la literatura, pero a continuación expondremos brevemente el método Quickhull que puede consultarse en [Eddy, 1977], por ser este el algoritmo implementado en R debido a su eficiencia computacional y que por tanto, utilizaremos sobre nuestro conjunto de datos.

El Quickhull [Barber et al., 1996] determina qué puntos de un conjunto en el plano son los vértices de un cierre convexo, quedando así perfectamente delimitado el polígono convexo. Su funcionamiento básicamente consiste en dividir el conjunto de datos de entrada en dos subconjuntos con respecto a una recta de referencia en el plano. Uno de los subconjuntos estará formado por los puntos que han quedado por encima de la recta de referencia y el otro, por los puntos que han quedado por debajo. Iterando este procedimiento, se obtienen finalmente los vértices del polígono.

---

**Algorithm 1** Quickhull algorithm
 

---

**Paso 0:** Comprobar que la muestra tiene al menos 3 puntos, en caso contrario, no podremos calcular el mínimo polígono convexo.

**Paso 1:** Elegir de la muestra de datos dos puntos  $(x_1, y_1)$  y  $(x_2, y_2)$ , de forma que sean los puntos muestrales con las coordenadas  $x_i$  con  $i \in \{1, 2\}$  más alta y más baja respectivamente. Serán los dos primeros vértices del cierre convexo.

**Paso 2:** Trazar una recta que pase por los dos puntos escogidos en el paso anterior, se designará como  $L_1$  y se utilizará como pivote para separar los puntos de la muestra en dos subconjuntos  $A_1$  y  $B_1$  que englobarán a los puntos que han quedado, respectivamente, por encima y por debajo de la recta. Los puntos que queden sobre la recta empleada como pivote, no serán considerados en sucesivas iteraciones.

**Paso 3:** Buscar los puntos  $P_{A_1}$  y  $P_{B_1}$  en los subconjuntos  $A_1$  y  $B_1$  respectivamente, que más lejos estén de la recta  $L_1$ .  $P_{A_1}$  forma un triángulo con los puntos  $(x_1, y_1)$  y  $(x_2, y_2)$  que están sobre la recta  $L_1$ . Análogamente para  $P_{B_1}$ . Los puntos que queden dentro de los dos triángulos, claramente no podrán ser vértices del mínimo polígono convexo, por lo que serán ignorados en los pasos siguientes.

**Paso 4:** Repetir el paso anterior sobre los dos segmentos que forman el triángulo (obviando el lado formado por la recta inicial  $L_1$ ).

**Paso i:** Repetir los pasos 3 y 4 hasta que no quede ningún punto de la muestra. Los puntos que hayan quedado al final, serán los vértices del cierre convexo.

---

A pesar de la popularidad de este método en la literatura debido a su sencillez, resulta evidente que presenta algunas desventajas.

- Es un método afectado por el tamaño muestral, pues la estimación está fuertemente correlada con el número de observaciones usadas, especialmente para muestras pequeñas que pueden no ser suficientemente representativas del área de actividad. Por este mismo motivo, es también un método muy sensible a la presencia de atípicos.
- No está adaptado a las elevaciones del territorio y tiende a sobrestimar la porción de área debido a la restricción de convexidad, pues incluye porciones de tierra nunca visitadas y en algunos casos imposibles de visitar por impedimentos geográficos.
- Presupone que el área de estimación es un polígono convexo, suposición que puede ser muy restrictiva en la práctica.
- No hay medida de la frecuencia de uso del espacio interno, simplemente encierra a todos los puntos cuando conecta los dos puntos más externos.



Pero sin duda, la deficiencia más grave resulta ser la **sobrestimación**. Por este motivo, se han propuesto algunas modificaciones de este método en la literatura para intentar corregirla. Existen multitud de variantes pero las más utilizadas consisten en eliminar, según cierto criterio, algunos puntos de la muestra y luego estimar normalmente el MCP, o bien en eliminar la restricción de convexidad del polígono.

En [Bath et al., 2006] puede consultarse una comparativa entre los distintos criterios para eliminar localizaciones individuales de la muestra, con el objetivo de obtener un método robusto para la estimación del área de actividad. Concretamente, los tres criterios propuestos pueden resumirse de la siguiente forma.

- Eliminar los puntos más alejados en distancia euclídea del centroide de la muestra.  
Este algoritmo es muy sencillo.
  - Calculamos el cierre convexo, así como su área y el centroide de la muestra  $c = (\bar{x}, \bar{y})$ .
  - Seleccionamos el dato muestral más alejado del centroide y lo eliminamos.
  - Repetimos el paso anterior hasta eliminar el número deseado de puntos de la muestra.
- Eliminar los puntos más alejados en distancia euclídea de la media armónica de la muestra. Es un caso completamente análogo al anterior, pero en lugar de calcular el centroide de la muestra, se calcula la media armónica y se elimina el punto más alejado de ella. La ventaja de esta técnica con respecto a la anterior reside en que la media armónica es más robusta a la presencia de atípicos que la media aritmética.
- Eliminar los puntos que producen un mayor decrecimiento en el área del home range. Es decir, el objetivo es eliminar la observación  $i$  que provoca mayor diferencia entre el área del MCP y el área del MCP sin esa observación. Por tanto, si  $A_{MCP}$  representa el área del MPC con todos los datos y  $A_{MCP-i}$  el área del MCP con todos los puntos de la muestra excepto la observación  $i$ , en cada iteración el objetivo es encontrar la localización  $i$  que maximice  $A_{MCP} - A_{MCP-i}$ . Afortunadamente, no necesitamos hacer la prueba con todas las observaciones, pues por definición, los que producen mayor decrecimiento del área del mínimo polígono convexo, serán aquellos que están en los extremos del cierre convexo.

Una cuestión importante es determinar la proporción de puntos de la muestra a eliminar para conseguir un método suficientemente robusto, que no se vea tan afectado por la sobrestimación. Worton sostiene que la robustez se alcanza al eliminar el 50 % de las observaciones muestrales, pues así la mayor parte de los atípicos habrán desaparecido. En [Worton, 1995] puede consultarse un estudio de simulación que considera muestras de distintos tamaños y procedentes de distintas distribuciones, y se concluye que eliminar el 50 % de los datos muestrales es la opción que proporciona mayor robustez al método sin infraestimar el home range.

Según el estudio elaborado en [Bath et al., 2006], de los tres criterios anteriores, el que presentó resultados más consistentes fue el que elimina los puntos en función de la distancia de estos al centroide de la muestra. Quizá se deba a la tendencia de la media a permanecer estable en todas las muestras. Por otra parte, el método que elimina los más alejados de la media armónica fue el que más variabilidad presentó en función de la distribución subyacente a los datos, probablemente por la variabilidad espacial que presenta la media armónica. No obstante, los tres criterios anteriores mostraron dificultades a la hora de diferenciar entre distribuciones asimétricas y simétricas.

En cuanto a las variantes del método del MCP que eliminan la restricción de convexidad para evitar así la sobreestimación, es popular el **método del polígono cóncavo**, que puede consultarse en [Harvey and Barbour, 1965]. Este método consiste en calcular la distancia  $d$  entre los dos puntos de la muestra que están más alejados. Emplearemos  $\frac{1}{4}d$  para determinar los límites exteriores del home

range, es decir, conectamos las observaciones periféricas de la muestra que disten entre si menos de  $\frac{1}{4}d$  y si algún punto externo dista más de  $\frac{1}{4}d$  de cualquier otro punto de la muestra, entonces se conecta al punto más cercano a través de una línea recta, pues probablemente se trate de una observación atípica y de esta forma la incluimos sin modificar el área del home range. Así, el área de actividad resultante no es necesariamente convexa y se adapta mejor a la forma de la muestra.

### 2.1.2. Cierre $\alpha$ -convexo

Este método está a caballo entre los enfoques geométrico y estadístico del problema. En parte porque ahora ya no se intenta simplemente estimar una porción de área a partir de unos puntos dados, sino que se intenta estimar el home range viéndolo como la reconstrucción del soporte de un vector aleatorio absolutamente continuo  $X$  del que tenemos una muestra aleatoria  $\mathcal{X}_n = \{X_1, \dots, X_n\}$ .

Si la desventaja principal del MCP era la sobrestimación, causada sobre todo por la suposición de convexidad del área a estimar, el cierre  $\alpha$ -convexo intenta corregir esta deficiencia afinando más en el grado de convexidad del conjunto. Conviene entonces una definición formal de concepto.

**Definición 2.1.** Un conjunto  $A \subset \mathbb{R}^d$  se dice  $\alpha$ -convexo para algún  $\alpha > 0$  si  $A = C_\alpha(A)$ , donde:

$$C_\alpha(A) = \bigcap_{B_\alpha(\mathbf{x}): B_\alpha(\mathbf{x}) \cap A = \emptyset} (B_\alpha(\mathbf{x}))^C \quad (2.1)$$

denota el cierre  $\alpha$ -convexo de  $A$  y  $B_\alpha(\mathbf{x})$  la bola abierta con centro  $\mathbf{x}$  y radio  $\alpha$ .

Por tanto, el cierre  $\alpha$ -convexo de la muestra no es más que la intersección de los complementarios de las bolas abiertas de radio  $\alpha$  que no contienen ningún punto de la muestra. Así, un rasgo característico de este estimador es que su frontera ya no son líneas rectas como en el caso del MCP, sino que son arcos de circunferencia. Esta característica soluciona en parte el problema de sobrestimación, pues el  $\alpha$  controlará que no se incluya en la estimación aquellas regiones dentro del área de actividad del animal que nunca han sido visitadas.

En [Rodríguez-Casal, 2007], se introdujo por primera vez el método del cierre  $\alpha$ -convexo para estimar el soporte de una variable aleatoria. En dicha publicación pueden consultarse detalladamente las propiedades del estimador y además se demuestra que a pesar de ser menos restrictivo en las suposiciones sobre el conjunto a estimar que el MCP, ambos presentan las mismas tasas de convergencia al conjunto original.

Este estimador, ampliamente estudiado en geometría computacional (ver en [Edelsbrunner, 2010]), produce buenas reconstrucciones si los puntos de los que disponemos para hacer la estimación están uniformemente repartidos por la muestra, por eso esta será una característica a tener en cuenta, sobre todo a la hora de escoger el valor para el parámetro  $\alpha$ , que controla la forma del conjunto a estimar. De hecho, para el caso  $\alpha \rightarrow \infty$ , estaríamos considerando que el área a estimar es convexa y por tanto sería completamente análogo al MCP.

En general, valores muy pequeños del  $\alpha$  pueden dar lugar a un estimador fragmentado, con varios núcleos, mientras que un  $\alpha$  grande puede producir sobrestimación.

En [Rodríguez-Casal and Saavedra-Nieves, 2014] se propone un método de selección automática del  $\alpha$ , atendiendo a criterios de consistencia, convergencia y uniformidad en el reparto de los puntos en la muestra, para después construir un estimador para el home range a partir de la elección hecha del  $\alpha$ .

### 2.1.3. Estimación a través de la función de utilización

En este apartado afrontaremos el problema de estimación del home range desde una perspectiva probabilística, pues ahora no se trata simplemente de estimar el área de actividad de un animal a partir de un conjunto de localizaciones, sino de determinar la probabilidad de encontrar al animal en un punto concreto del espacio, por tanto nos da una estimación del área y además, de la utilización que se hace de la misma.

En este contexto, estrechamente ligado al concepto de home range aparece el término de **distri-bución de utilización** (UD) [van Winkle, 1975].

La UD es por tanto una función de distribución de dos variables que representa la probabilidad de encontrar un animal en un área definida como su home range, permitiendo así conocer el tiempo que pasa el animal en cualquier punto del espacio. Teniendo esto en cuenta, el home range puede redefinirse como el conjunto de nivel  $\{f \geq c\}$  que con un 95 % de probabilidad contiene a la densidad de utilización, de forma que:

$$0.95 = \int_{\{f \geq c_{0.95}\}} f(\mathbf{x}) d\mathbf{x}$$

Por tanto, el objetivo es hacer una estimación de una función de densidad 2-dimensional a partir de un conjunto de localizaciones, en principio independientes.

### Estimación paramétrica

En la literatura relativa a la estimación de la función de densidad de la distribución de utilización en el contexto de home range, se han utilizado inicialmente versiones paramétricas. En [Calhoun and Casby, 1958] puede consultarse una propuesta de estimación utilizando un modelo normal bidimensional o en [Don and Rennolls, 1983], con una mixtura de normales, que entre otras cosas permite la multimodalidad de la densidad estimada.

Pero los métodos paramétricos con frecuencia han demostrado ser inapropiados, pues presuponen que las localizaciones proceden de una distribución concreta con unas características determinadas que fácilmente pueden no corresponderse con la realidad.

### Estimación no paramétrica. Estimador tipo núcleo.

Los estimadores no paramétricos de la función de densidad unidimensional más frecuentes son: el histograma, el estimador naive y el estimador tipo núcleo. Un problema de los dos primeros estimadores es que son discontinuos, independientemente de que la función de densidad a estimar sea continua, por tanto, el estimador más habitual y que emplearemos será el estimador tipo núcleo, ya que remedia este problema.

En [Worton, 1989] se propuso por primera vez el estimador tipo núcleo de la función de densidad como una aproximación no paramétrica de la densidad de utilización.

La utilización del estimador tipo núcleo de la función de densidad para estimar el home range es, junto con el método del cierre convexo, la técnica más ampliamente utilizada en la literatura.

El estimador tipo núcleo de la función de densidad (KDE, kernel density estimator) se introdujo a finales de los años 50 por Parzen y Rosenblatt [Rosenblatt, 1956]. Para un conjunto de observaciones  $\{X_1, X_2, \dots, X_n\}$  y para el caso unidimensional se define como:

$$\hat{f}_{n,K}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i),$$

donde  $K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right)$ .

La función de densidad  $K$  representa a la función tipo núcleo que se utiliza para la estimación. Algunas de las funciones tipo núcleo más frecuentes son:

- Uniforme:  $\frac{1}{2\mathbb{I}(|x| < 1)}$ .
- Triangular:  $(1 - |x|)\mathbb{I}(|x| < 1)$ .
- Epanechnikov:  $\frac{3}{4(1-x^2)\mathbb{I}(|x| < 1)}$ .
- Biweight:  $\frac{15}{16(1-x^2)^2\mathbb{I}(|x| < 1)}$ .

- Triweight:  $\frac{35}{32(1-x^2)^3 \mathbb{I}(|x| < 1)}$ .
- Gaussiana:  $\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$ .

Pero es sabido que la elección del núcleo tiene poco efecto en la precisión del estimador de la función de densidad en comparación con el efecto del parámetro de suavizado (o ventana)  $h$ , [Epanechnikov, 1969].

La ventana óptima  $h$  controla la cantidad de variabilidad en cada componente de la estimación. Es decir, si  $h$  toma un valor pequeño, podremos observar con detalle las características de los datos mientras que si  $h$  toma un valor grande, la función estimada será suave y podrá ocultar así alguna de sus características. Conviene tener en cuenta que un aumento de la precisión de la estimación con un  $h$  pequeño conlleva un notable incremento del sesgo.

Habitualmente,  $h$  se elige de forma que minimice algún criterio de error (local o global). El problema es que estas ventanas no pueden utilizarse en la práctica ya que dependen de la función de densidad que intentamos estimar y que a priori, es desconocida. A pesar de ello, nos proporcionan un punto de partida para construir selectores a partir de los datos.

Los métodos que utilizaremos son los siguientes:

- Normal scale rule (regla del pulgar) [Silverman, 1986]: escoge la ventana  $h$  de forma que minimice el error cuadrático medio integrado asintótico (AMISE), que no es más que la versión asintótica de la integral del error cuadrático medio que se define como

$$MSE(x) = (\mathbb{E}(\hat{f}(x)) - f(x))^2.$$

En este caso, la ventana  $h_{AMISE}$  adquiere la forma

$$h_{AMISE} = \left( \frac{R(K)}{n\mu_2^2(K)R(f'')} \right)^{\frac{1}{5}}$$

donde  $R(f'') = \int f''(x)dx$ . La propuesta de Silverman consiste en aproximar  $R(f'')$  por la integral correspondiente de una normal de media 0 y desviación típica  $\sigma$ , y en ese caso tendríamos que

$$h_{AMISE} = \left( \frac{8\sqrt{\pi}R(K)}{3\mu_2(K)^2n} \right)\sigma,$$

donde  $\sigma$  se reemplaza por un estimador que puede ser la cuasi-desviación típica, el rango intercuartílico estandarizado o, como recomienda Silverman, el mínimo de ambas.

- Plug-in rule [Sheather and Jones, 1991]: nuevamente elige la ventana  $h$  minimizando el AMISE. En esta ocasión, Sheather and Jones proponen estimar de manera no paramétrica  $R(f'')$  utilizando estimadores tipo núcleo. Denotemos por  $\psi_r = \mathbb{E}(f^{(r)}(X))$ , aplicando dos veces integración por partes, se tiene que

$$\int (f''(x))^2 dx = \int f^{(4)}(x)f(x)dx = \psi_4.$$

Por tanto, el problema radica en estimar  $\psi_4 = \mathbb{E}(f^{(4)}(x))$  que viene dada por la media muestral

$$\hat{\psi}_r = \frac{1}{n} \sum_{i=1}^n \hat{f}^{(r)}(X_i)$$

donde  $\hat{f}^{(r)}$  podría ser un estimador tipo núcleo de la densidad, con por ejemplo, núcleo  $L$  y ventana  $g$

$$\hat{f}^{(r)}(X_i) = \frac{1}{n} \sum_{j=1}^n L_g^{(r)}(X_i - X_j),$$

pero en este caso necesitaríamos una nueva ventana  $g$ . Más concretamente, para  $r = 4$  podríamos obtener  $g$  por argumentos similares pero necesitaríamos  $\hat{\psi}_6$  y para  $\hat{\psi}_6$  necesitaríamos  $\hat{\psi}_8$ . En este punto, Sheather y Jones proponen asumir que  $f$  es normal con desviación típica  $\sigma$  y en ese caso

$$\psi_r = \frac{(-1)^{\frac{r}{2}} r!}{(2\sigma)^{r+1} (\frac{r}{2})! \sqrt{\pi}}.$$

Se estima  $\psi_r$  utilizando  $\hat{\sigma}$  y se itera hacia atrás.

- Validación cruzada (CV, cross validation), [Bowman, 1984]: los métodos de validación cruzada afrontan el problema de selección de la ventana aproximando medidas de error, en lugar de basarse en expresiones de ventanas óptimas.

El error cuadrático integrado (ISE) de  $\hat{f}_{n,k}$  como estimador de  $f$  es:

$$ISE(X_1, \dots, X_n; h) = \int (\hat{f}(x) - f(x))^2 dx,$$

que puede expresarse de forma equivalente resolviendo el cuadrado como

$$ISE(h) = \int \hat{f}_{n,k}^2(x) dx - 2 \int \hat{f}_{n,k}(x) f(x) dx + \int f^2(x) dx = R(\hat{f}_{n,k}) - 2 \int \hat{f}_{n,k}(x) f(x) dx + R(f).$$

Obsérvese que este último sumando no depende de  $h$  y por tanto, la ventana óptima para el ISE es

$$\begin{aligned} h_{ISE} &= \arg \min_h R(\hat{f}_{n,k}) - 2 \int \hat{f}_{n,k}(x) f(x) dx = \arg \min_h \int \hat{f}_{n,k}^2(x) dx - 2 \int \hat{f}_{n,k}(x) f(x) dx = \\ &= \arg \min_h \frac{1}{n^2 h} \sum_{i,j} K * K \left( \frac{X_i - X_j}{h} \right) - 2 \mathbb{E}(\hat{f}_{n,k}(x)). \end{aligned}$$

Por tanto, si tuvieramos una muestra  $\tilde{X}_1, \dots, \tilde{X}_m$  de  $X$ ,  $\mathbb{E}(\hat{f}_{n,k}(x))$  podría aproximarse por  $\frac{1}{m} \sum_{k=1}^m \hat{f}_{n,k}(\tilde{X}_k)$ . Pero no disponemos de tal muestra, y es aquí donde entra la idea de validación cruzada, ya que utilizando  $X_1, \dots, X_n$  la cantidad anterior se puede aproximar por

$$\frac{1}{n} \sum_{i=1}^n \hat{f}_{n,k}^{-i}(X_i),$$

donde  $\hat{f}_{n,k}^{-i}$  es el KDE obtenido con toda la muestra excepto el dato  $i$ -ésimo.

El método de selección de  $h$  por validación cruzada se basa en obtener el  $h$  que minimiza la función de validación cruzada  $CV(h) = R(\hat{f}_{n,k}) - \frac{2}{n} \sum_{i=1}^n \hat{f}_{n,k}^{-i}(X_i)$ .

La ventana  $h_{CV}$  se denomina ventana de validación cruzada insesgada.

Una de las ventajas de las técnicas de validación cruzada es que se pueden exportar fácilmente a otros contextos, aunque son altamente variables y tienden a infrasuavizar. Una mejora puede obtenerse el método de validación cruzada sesgada, que en lugar de considerar el ISE, parte del AMISE.

- Selector bootstrap, [Taylor, 1989]: La idea de aproximar  $h$  utilizando bootstrap consiste en aproximar un criterio de error global, por ejemplo el error cuadrático medio integrado (MISE) mediante remuestreo. Denotemos por  $h_1$  la ventana seleccionada para obtener un KDE inicial con la muestra observada. El algoritmo será el siguiente:

**Algorithm 2** Selector bootstrap de  $h$ 

**Paso 0:** De la muestra inicial  $X_1, \dots, X_n$  obtener un KDE  $\hat{f}_{n,K}(x)$  que depende de  $h_1$ .

**Paso 1:** Considerar  $\hat{f}_{n,K}(x)$  como la verdadera densidad en el universo bootstrap.

**Paso 2:** La versión bootstrap del MISE es

$$MISE^*(h_2) = \mathbb{E}^* \left( \int (\hat{f}_{n,K}^*(x, h_2) - \hat{f}_{n,K}(x, h_1))^2 dx \right)$$

**Paso 3:**  $h_1$  es ventana para el KDE inicial  $h_2$  la ventana bootstrap.

**Paso 4:** El MISE se puede aproximar por Monte Carlo y después minimizar.

Pero el caso que nos ocupa es el de la estimación de una función de densidad bidimensional, así que, una vez expuesto el caso univariante, introducimos la generalización multivariante de la estimación no paramétrica de la función de densidad.

Consideremos ahora  $\mathbb{X}$  un vector aleatorio  $d$ -dimensional y sea  $\mathcal{X} = (\mathbb{X}_1, \dots, \mathbb{X}_n)$  una muestra aleatoria. El estimador tipo núcleo de la densidad puede generalizarse de la siguiente forma:

$$\hat{f}_{n,K,H}(\mathbf{x}) = \frac{1}{n\sqrt{\det(H)}} \sum_{i=1}^n \frac{K_d}{\sqrt{H(\mathbf{x} - \mathbb{X}_i)}},$$

con  $\mathbf{x} \in \mathbb{R}^d$  donde  $H$  es una matriz de ventanas y  $K_d$  es una densidad  $d$ -dimensional.

Para su construcción, debemos inicialmente fijar una función núcleo. Para ello, se podría considerar una densidad normal multivariante o bien tomar un núcleo producto:  $K_d(\mathbf{x}) = \prod_{j=1}^d K(\mathbf{x}_j)$  siendo  $K$  una densidad unidimensional.

Con respecto a la matriz de ventanas  $H$ , esta tiene  $\frac{d(d+1)}{2}$  elementos diferentes. Así, en dimensión 2, tendríamos que calcular 3 ventanas. Algunas formas simplificadas de la matriz  $H$  son:

- $H = hI$ , siendo  $I$  la matriz identidad  $d$ -dimensional.
- $H = \text{diag}(h_1, \dots, h_d)$ , permitiendo distintos grados de suavidad en distintas componentes.

Dependiendo de si los valores de  $H$  están fijados en todo el plano o si sus valores varían según las distintas zonas del plano (las regiones con una baja concentración de puntos tendrán valores de  $H$  más grandes que las áreas con mucha concentración de puntos), distinguimos entre **kernel fijo** y **kernel adaptativo** respectivamente.

Por tanto, el estimador de la UD con el método del kernel adaptativo sería de la forma:

$$\hat{f}_{n,K,H}(\mathbf{x}) = \frac{1}{n\sqrt{\det(H_i)}} \sum_{i=1}^n \frac{K_d}{\sqrt{H_i(\mathbf{x} - \mathbb{X}_i)}},$$

donde los parámetros de suavizado  $H_i$  varían de una observación a otra.

En los últimos años, han surgido algunas modificaciones del KDE como pueden ser el **KDE que incorpora información del animal en estudio** propuesto por [Matthiopoulos, 2003], o el **estimador permisible del área de actividad de un animal** (PHRE-permissible home range estimator), [Tarjan and Tinker, 2016].

Matthiopoulos nos propone la utilización de un KDE modificado (MSKS-model supervised kernel smoothing), calculado como promedio del KDE usual y un modelo que refleje los hábitos del animal, como pueden ser: limitaciones geográficas, comportamiento, o hábitos alimenticios. Si el modelo de los hábitos de los animales es exhaustivo, la estimación habitualmente mejora a la proporcionada por el KDE usual.

Este modelo no sólo tiene el tamaño de la ventana  $h$  como parámetro, sino también el peso  $w$  que se da a la estimación usual (estimación con KDE), siendo  $1 - w$  el peso de la estimación auxiliar (la que incorpora la información del animal).

A continuación en el Algoritmo 3 se expone el método para el cálculo del MSKS propuesto en [Matthiopoulos, 2003].

---

**Algorithm 3** Algoritmo del MSKS

---

**Paso 0:** Construimos el home range del animal en base a la información que tenemos de la especie, ya sea relativa a estudios anteriores o a sus hábitos y costumbres. Esta superficie la denotaremos como  $u_x$ .

**Paso 1:** Construimos el home range con el procedimiento expuesto anteriormente utilizando el KDE con todas las observaciones menos una, y la denotamos por  $v_x(h)$ , siendo  $h$  el coeficiente de suavizado empleado en el estimador tipo núcleo.

**Paso 2:** Combinamos las dos superficies calculadas anteriormente  $u_x$  y  $v_x(h)$  haciendo un promedio ponderado,

$$\hat{s}_x(h, w) = wu_x + (1 - w)v_x(h),$$

donde  $0 \leq w \leq 1$  representa el peso que se da a cada una de las estimaciones del home range,  $u_x$  y  $v_x(h)$ .

**Paso 3:** Obtenemos el log-verosimilitud bajo la superficie  $\hat{s}_x(h, w)$  de la observación descartada en la estimación hecha con KDE. Es decir, calculamos el logaritmo de los valores de la superficie en las coordenadas de la observación faltante.

**Paso 4:** Repetimos los pasos del 1 al 3 para todas las observaciones del conjunto de datos y construimos así un log-verosimilitud que dependa del valor de los parámetros  $h$  y  $w$ .

**Paso 5:** Repetimos el Paso 4 con diferentes valores del parámetro de suavizado  $h$  y diferentes pesos  $w$  y seleccionamos los dos valores que proporcionen un log-verosimilitud más alto.

**Paso 6:** Generamos la superficie final realizando los pasos del 0 al 2 usando los valores de  $h$  y  $w$  obtenidos en el Paso 5.

---

El resultado de la estimación con MSKS estará por tanto afectado por el tamaño de la muestra y la calidad del modelo auxiliar que incorpora información sobre el comportamiento y las costumbres del animal en estudio. En el peor de los casos, si el modelo de comportamiento no es suficientemente informativo, el resultado del algoritmo será muy similar a la salida obtenida con el KDE usual.

Por otro lado, las modificaciones que introduce Tarjan, aunque conceptualmente son similares, se basan en la incorporación de información adicional al modelo mediante una transformación previa de los datos en función de la elevación del terreno. Posteriormente, se realiza normalmente la estimación con el KDE. En general, el problema del método es que no está claro que en todos los contextos, las características del terreno puedan dar lugar a un nuevo sistema de coordenadas.

Veamos brevemente una descripción del PHRE.

---

**Algorithm 4** Algoritmo del PHRE

---

**Paso 0:** Identificar los elementos del hábitat que condicionan el uso que el animal hace de su entorno.

**Paso 1:** Transformar las observaciones en un nuevo sistema de coordenadas que sea representativo del comportamiento de la especie sobre la superficie disponible.

**Paso 2:** Hacer una estimación del KDE bidimensional sobre las observaciones transformadas en el Paso 1.

**Paso 3:** Deshacer la transformación hecha en el Paso 1, de la función de densidad estimada en el Paso 2 para que sus valores estén en el sistema de coordenadas habitual.

---

Este método presenta buenos resultados en cuanto a que excluye el área no visitada, predice con exactitud y requiere pocos datos para la estimación, como puede consultarse en [Tarjan and Tinker, 2016]. Pero al igual que el MSKS tienen el inconveniente de que exigen un conocimiento previo de las condiciones del terreno.

## 2.2. Métodos localizados para datos independientes

En esta sección expondremos métodos que de nuevo asumen la independencia de las observaciones utilizadas en la estimación del home range, pero al contrario que en la sección anterior, las técnicas de las que hablaremos serán **métodos localizados**, pues la estimación no se hará con la totalidad de los datos disponibles, sino agrupándolos en diferentes categorías para posteriormente hacer una estimación sobre cada una de ellas.

Por tanto, los métodos localizados para la estimación del área de actividad de una animal que estudiaremos a continuación, siguen el siguiente esquema:

- Agrupación de las observaciones en entornos locales de acuerdo a algún criterio (con técnicas de clustering, con la técnica del vecino más próximo o puntos que pertenecen a una bola de un radio prefijado).
- Utilización de alguno de los métodos globales expuestos en la subsección anterior para obtener la estimación del home range en cada uno de los entornos obtenidos en el paso anterior.
- Unión de todos los home range obtenidos localmente para cada entorno.

Este procedimiento puede sufrir variaciones dependiendo de qué criterios se utilicen para contruir los entornos y para unirlos posteriormente. Definimos a continuación las técnicas locales para datos independientes más empleadas en la estimación del home range.

### 2.2.1. Cierre convexo local (LoCoH) o k-vecino más próximo del cierre convexo (k-NNCH)

En [Getz and Wilmsers, 2004] se propone una aplicación local del método del cierre convexo expuesto en la Subsección 2.1.1 dando lugar al procedimiento que describimos a continuación.

El método comienza construyendo el cierre convexo asociado a cada observación  $\mathbf{x} = (x_i, y_i)$  y sus  $k - 1$  vecinos más próximos (k-NNCH, k-nearest neighbor convex hulls) en distancia euclídea hasta haber incluido todos los puntos.

Inmediatamente después, se calcula el área cubierta por cada k-NNCH, criterio que se utiliza para reordenarlos por tamaño. El home range estimado será el resultado de la unión progresiva de los cierres



calculados hasta incluir el porcentaje de puntos deseado, desde el más pequeño hasta el más grande, donde los cierres más pequeños indican áreas visitadas con más frecuencia.

En la práctica, escoger un número entero  $k$  de puntos a incluir en la estimación local puede no ser un criterio suficientemente flexible, por este motivo, en [Getz et al., 2007] se proponen dos extensiones del  $k$ -LoCoH que solucionan este problema y que explicamos brevemente:

- **$r$ -LoCoH** o *estimación de cierre convexo local con radio fijo*: se realiza un cierre convexo de la observación de referencia y todos los puntos que estén a una distancia menor o igual que  $r$  del mismo. Ahora los cierres serán aproximadamente del mismo tamaño, por lo que los ordenamos de mayor a menor según el número de puntos que contienen. En caso de empate, añadiremos antes el cierre de menor área.

Si  $r$  es muy pequeño puede ocurrir que algunos puntos solo tengan un vecino o ninguno. En ese caso, el punto se uniría con una línea al cierre convexo más próximo de forma que no añade área a la estimación.

- **$a$ -LoCoH** o *estimación de cierre convexo local con radio adaptativo*: en este caso, se agrupan todos los puntos cuya suma de la distancia al punto de referencia es menor o igual que  $a$ . Este método simplemente ajusta el radio del círculo que circunscribe cada cierre local, de forma que el cierre convexo local más pequeño contendrá muchos puntos y corresponderá a áreas visitadas con mucha frecuencia. De nuevo, la estimación del home range resultará de la unión progresiva de los cierres convexos locales.

### Reglas para seleccionar $k$ , $r$ o $a$

Para valores pequeños de  $k$ ,  $r$  o  $a$  la estimación puede contener agujeros que desaparecen al incrementar ese valor. En [Getz and Wilmers, 2004] se propone el método MSHC (minimum spurious hole covering) que consiste en seleccionar el mínimo valor de  $k$ ,  $r$  o  $a$  que produzca una estimación final con la misma topología que el conjunto dado.

En la práctica, el conjunto que se pretende estimar no es conocido y por tanto, su topología tampoco. En estos casos, el MSHC propone inferir el número de agujeros que tendrá el conjunto a estimar, identificando características del terreno como lagos, montañas, o zonas inaccesibles para la especie en estudio, obviando aquellos que tengan un tamaño muy pequeño en comparación con el conjunto que se pretende estimar.

La idea de incorporar información topológica, aunque no ha sido muy estudiada en la literatura, se presenta como una prometedora área de investigación estadística en el futuro que bajo la denominación de Topological Data Analysis (TDA) abarca una colección de métodos para análisis de datos mediante el estudio de la estructura de los mismos. En [Wasserman, 2018] puede consultarse una pequeña revisión de la literatura existente sobre TDA.

### 2.2.2. Cierre poligonal característico (CHP)

Es de nuevo una versión local del mínimo polígono convexo, con dos ventajas importantes: permite la unión de polígonos cóncavos y la existencia de regiones disjuntas en la estimación del home range.

Fue propuesto en [Downs and Horner, 2009] y su implementación es relativamente sencilla, pues consiste simplemente en obtener la triangulación de Delaunay, [Delaunay, 1934], a partir del conjunto de observaciones y posteriormente, construir el estimador del área de actividad eliminando de la triangulación el 5% de los triángulos de mayor perímetro. El área resultante, será el estimador del home range.

La idea subyacente de la triangulación de Delaunay consiste simplemente en conectar puntos cercanos hasta formar triángulos de forma que se crea una red de triángulos conexa y convexa que cumple la **condición de Delaunay**: *la circunferencia circunscrita en cada triángulo de la red no debe contener el vértice de ningún otro triángulo.*

Existen numerosos algoritmos para llevar a cabo la triangulación de Delaunay, puede consultarse una breve comparativa de su funcionamiento en [Su and Drysdale, 1997].

### 2.2.3. Single-linkage cluster

Es otra variante del MCP que sólo se diferencia del LoCoH en como se agrupan inicialmente las observaciones, pues en este caso se utiliza una técnica de clúster jerárquico con single linkage.

El algoritmo fue propuesto por [Kenward et al., 2001] como una alternativa a las técnicas de estimación globales, que permitía al estimador ser multinuclear para reducir así el sesgo. El algoritmo comienza agrupando las observaciones en clústers. Se establece un tamaño mínimo del clúster (tres observaciones), es decir, las tres localizaciones con la mínima distancia entre ellas, formarán el primer clúster. Cada clúster, ganará una nueva observación minimizando la distancia a la observación más cercana contenida en el clúster (criterio del single linkage). En caso de que una observación esté exactamente a la misma distancia de la observación más próxima de dos clústers, la observación irá finalmente al clúster más cercano en términos de distancia euclídea entre la observación y el centroide del clúster.

Una vez formados los grupos, se procede como en el LoCoH: se hace una estimación del MCP en cada grupo y la unión de todos ellos será el estimador buscado.

## Capítulo 3

# Estimación del home range con datos dependientes

Las nuevas técnicas de recolección de datos (sistemas de posicionamiento global - GPS) nos permiten obtener más observaciones y con mayor frecuencia.

Aunque de esta forma se ha simplificado la recogida de datos y supone un recurso muy potente de información sobre el comportamiento del sujeto en estudio, también ha cambiado las características de los datos, pues las observaciones que están próximas entre si tienen dependencia temporal.

Por este motivo, es necesario abordar la estimación con otras técnicas que tengan en cuenta esta característica de los datos y que aprovechen toda la información para hacer un seguimiento más detallado de los movimientos del animal. Así, a continuación trataremos de nuevo métodos globales y locales, especialmente diseñados para trabajar con observaciones correlacionadas.

### 3.1. Métodos globales para datos dependientes

La dependencia temporal de las observaciones nace de la naturaleza continua del movimiento y por la repetición de ciertos movimientos a lo largo del tiempo. A pesar de que los datos autocorrelados contienen menos información posicional que los datos independientes, razón por la cual las técnicas de estimación habituales empleadas sobre datos dependientes infraestiman el home range, también es cierto que esta dependencia nos permite relacionar pasado y futuro facilitándonos así la elaboración de predicciones.

En esta sección expondremos un método global para la estimación del home range con datos dependientes, basado en la función de utilización, así como las variaciones que se han propuesto del mismo hasta la actualidad. También veremos otros dos métodos que hacen la estimación tratando las observaciones como trayectorias de un proceso estocástico.

#### 3.1.1. Estimación de la densidad geográfica-temporal (TGDE)

El método TGDE (Time-Geography Density Estimation) surge como una fusión de las técnicas de geografía temporal y la estimación tipo núcleo de la función de densidad (KDE).

La geografía temporal fue definida por primera vez en [Hägerstrand, 1970] y es un campo de la demografía que proporciona herramientas para modelizar el comportamiento humano y las iteraciones a partir de observaciones espaciales e información de la actividad o capacidad de movimiento de los sujetos. Un concepto importante en este contexto es el de **prisma espacio-temporal**, que delimita todas las posibles localizaciones entre dos observaciones  $\mathbf{x}_i$  y  $\mathbf{x}_{i+1}$  en instantes de tiempo consecutivos donde podría haber estado el sujeto teniendo en cuenta las restricciones temporales y espaciales y su velocidad máxima. La proyección del prisma espacio-temporal en el plano, da lugar a la **geoelipse** o

**PPA** (Potential Path Area), que para dos observaciones  $\mathbf{x}_i$  y  $\mathbf{x}_{i+1}$  en un intervalo de tiempo  $[t_i, t_{i+1}]$  se define como:

$$g_i = \{\mathbf{x} : \|\mathbf{x} - \mathbf{x}_i\| + \|\mathbf{x} - \mathbf{x}_{i+1}\| \leq (t_{i+1} - t_i)v_i\},$$

donde  $v_i$  denota la velocidad máxima en el intervalo de tiempo considerado. Así, la geolipse no es más que una elipse cuyos focos son dos observaciones consecutivas y que tiene la forma determinada por la amplitud del intervalo de tiempo, la distancia entre las observaciones consideradas y la velocidad máxima.

Por tanto, en [Downs et al., 2011] se propone la técnica de TGDE como una fusión del KDE y las PPA para construir la estimación del home range en datos que presentan dependencia temporal.

La velocidad máxima se supone constante en todos los intervalos, por tanto  $v_i = v$  para todo  $i \in \{1, \dots, n\}$ , siendo  $n$  el tamaño muestral. Así, podemos estimar la densidad espacio-temporal en un punto  $\mathbf{x}$  como:

$$\hat{f}_{TG}(\mathbf{x}) = \frac{1}{(n-1)[(t_n - t_1)v]^2} \sum_{i=1}^{n-1} G\left(\frac{\|\mathbf{x} - \mathbf{x}_i\| + \|\mathbf{x} - \mathbf{x}_{i+1}\|}{(t_{i+1} - t_i)v}\right), \quad (3.1)$$

donde  $G$  es una función tipo núcleo en la que el numerador es la suma de la distancia entre una observación dada y las observaciones  $\mathbf{x}_i$  y  $\mathbf{x}_j$  tomadas en los instantes de tiempo  $t_i$  y  $t_j$ . En cuanto al denominador, calcula la distancia máxima a la que puede moverse el sujeto en estudio en un intervalo de tiempo dado considerando que lo hace a la máxima velocidad.

El TGDE del home range será por tanto el conjunto de nivel al 95 % de la estimación de la densidad hecha en (3.1),

$$0.95 = \int_{\{\hat{f}_{TG} \geq \hat{c}_{0.05}\}} \hat{f}_{TG}(\mathbf{x}) d\mathbf{x}.$$

Puesto que esta técnica es relativamente novedosa, el efecto de los parámetros sobre la estimación está todavía en estudio, pero los factores principales que afectan a la estimación son:

- El intervalo de tiempo transcurrido entre dos observaciones consecutivas: cuanto menor es el intervalo de tiempo entre cada par de observaciones, existe menos incertidumbre en el movimiento del animal y por tanto, tenemos información más detallada sobre su posición, lo que se traduce en estimaciones del home range más precisas. Si por el contrario la toma de observaciones se realiza con poca frecuencia, existirá más incertidumbre sobre la posición del animal en los instantes temporales en los que no se ha tomado muestra y por tanto, en este caso el estimador sobrestimará el home range.
- La función tipo núcleo  $G$ : aunque al igual que en el KDE, la elección de la función tipo núcleo no es un factor que influya en gran medida la estimación, sí debemos tener en cuenta que esta función controla la distribución de los datos. Mientras que un núcleo uniforme presupone que las observaciones se reparten de forma igual por toda la muestra, uno no uniforme puede asignar mayor probabilidad a una región concreta.
- La velocidad máxima  $v$ : juega el mismo papel que el parámetro de ventana  $h$  en el KDE. Un valor grande de la velocidad máxima producirá geolipses grandes y por tanto, se traducirá en una sobrestimación del home range y por el contrario, un valor demasiado pequeño, infraestimaré el home range.

Por estos motivos, aunque el TGDE se presenta como una alternativa para lidiar con la autocorrelación presente en los datos debido a la dependencia temporal de las observaciones, lo cierto es que para su correcto funcionamiento es necesaria una buena elección del parámetro  $v$  de velocidad máxima y una muestra en la que las observaciones hayan sido tomadas en intervalos de tiempo pequeños.

Un caso particular del TGDE es el **PPA home range**, que no es más que la unión de las  $n - 1$  geoelipses obtenidas de cada par de observaciones consecutivas de la muestra, es decir, el TGDE con núcleo uniforme:

$$PPA_{HR} = \cup_{i=1}^{n-1} g_i,$$

considerando la velocidad máxima constante en todos los intervalos, es decir,  $v_i = v$  para todo  $i \in \{1, \dots, n - 1\}$ .

En [Long and Nelson, 2015], se propone una transformación del PPA que permite definir el home range en función de la tasa de movimiento existente entre cada par de observaciones, es decir, un PPA en el que la velocidad máxima ya no es constante, resolviendo así el problema de la elección de este parámetro. Este método recibe el nombre de **PPA dinámico**. El objetivo de esta pequeña modificación es principalmente, eliminar la sobrestimación que el PPA produce en las zonas donde el sujeto en estudio se mueve muy despacio.

La construcción por tanto es totalmente análoga a la anterior, con la diferencia de que la velocidad máxima no es constante en todos los intervalos. Para estimar la velocidad máxima, Long y Nelson proponen considerar  $p$  instantes temporales y en cada uno de ellos, calcular la velocidad como un ratio de la distancia recorrida entre el tiempo empleado en recorrerla. Posteriormente, se ordenan las velocidades obtenidas para cada instante temporal  $v_1 < \dots < v_{p-1} < v_p$  y en cada uno de ellos se calcula la velocidad máxima como:

$$v_{max} = \left( \frac{k+2}{k+1} \right) v_m - \left( \frac{1}{k+1} \right) v_{m-k},$$

con  $1 < k < m \leq p$ . De esta forma, se obtendría el estimador PPA dinámico del home range.

Otros métodos que se han propuesto para lidiar con la dependencia de las observaciones parten del KDE y le añaden pequeñas modificaciones. Exponemos a continuación cuatro de ellas.

### Estimación de la función de densidad 4-dimensional

La definición de distribución de utilización (UD) que manejamos es la dada por [van Winkle, 1975], que la expresa como una función de distribución de dos variables que representa la probabilidad de encontrar un animal en el plano. De esta forma, estamos presuponiendo que la elevación del terreno y los instantes temporales no aportan información relevante para incorporar a la estimación del home range.

Por este motivo, [Keating and Cherry, 2009] propusieron una reformulación de la definición de UD, considerándola como una función de distribución de cuatro variables: tres dadas por el espacio y la cuarta dada por el tiempo. En este caso, la densidad que intentamos estimar tendría 4-dimensiones, por tanto propone utilizar una generalización del KDE a 4-dimensiones considerando simplemente núcleos multiplicativos.

En la práctica, esta aproximación no parece adecuada puesto que trata el tiempo como una variable de la que nos interesa conocer la frecuencia, cuando no es el caso.

### Movement-based KDE (MKDE)

En [Benhamou and Cornelis, 2010] se propone un KDE móvil (MKDE), que consiste en dividir cada intervalo de tiempo  $[t_i, t_{i+1}]$  en subintervalos añadiendo puntos a intervalos regulares. Así, el KDE se calcula sobre las observaciones y sobre los puntos añadidos por interpolación a intervalos regulares en cada paso.

El parámetro de suavizado  $h$  no es constante sino que es una función del intervalo de tiempo que va desde  $t_i$  hasta  $t_{i+1}$ , alcanzando su valor más pequeño ( $h_{min}$ ) en los puntos finales y el más grande ( $h_{max}$ ) en el punto medio.

Por tanto, el MKDE depende de la elección de tres parámetros:  $h_{min}$ ,  $h_{max}$  y la longitud del intervalo de tiempo. La selección óptima de forma simultánea de estos tres parámetros es computacionalmente inabordable incluso para tamaños de muestra no muy grandes, lo que supone una gran desventaja del método. Además, tampoco está claro que el MKDE sea considerablemente diferente del KDE para observaciones independientes.

### Scaled-line KDE (SLKDE)

Este estimador fue propuesto en [Steiniger and Hunter, 2013] con el objetivo de retener la dependencia temporal de los datos en la estimación y es muy similar al MKDE anterior, pero ahora el KDE se utiliza sobre los segmentos que unen dos observaciones consecutivas del conjunto de datos.

En primer lugar, se unen las observaciones consecutivas mediante un segmento y a continuación, se rasterizan los segmentos, creando un mapa de bits con celdas de tamaño  $s$ , que será un parámetro con un papel análogo al  $h$  del KDE usual.

De esta forma, el segmento que une dos localizaciones queda caracterizado por un conjunto de puntos localizados en el interior de las celdas sobre los que se hace una estimación tipo núcleo usual. El tamaño óptimo de  $h$  se elige por validación cruzada y el tamaño de la celda  $s$  se escoge de forma que  $s < h$ .

Para terminar, se suman todas las estimaciones obtenidas sobre cada uno de los segmentos para conseguir la estimación final de la UD que no necesariamente integrará 1.

Este procedimiento se caracteriza porque los valores que toma el estimador de la UD tienden a decrecer hacia el centro del segmento, produciendo así una forma de hueso entre dos observaciones.

### KDE autocorrelado (AKDE)

El AKDE no es más que una generalización del KDE, de hecho, en ausencia de dependencia temporal en los datos, el AKDE converge al KDE usual.

Este método fue propuesto por [Fleming et al., 2015] y estima la densidad de la UD utilizando un núcleo gaussiano bidimensional. La novedad está en que se considera que el movimiento del animal está representado por un proceso no estacionario que permite la variación a lo largo del tiempo de la velocidad media y la función de autocorrelación (ambos síntomas de dependencia temporal).

Para calcular la matriz de parámetros de suavizado  $H$ , supone que las observaciones proceden de un proceso estocástico gaussiano no estacionario y la obtiene como resultado de minimización del MISE. Pueden consultarse los cálculos relativos a la obtención de  $H$  en el Apéndice B de [Fleming et al., 2015].

#### 3.1.2. Modelo de movimiento del puente browniano (BBMM)

El BBMM (Brownian Bridge Movement Model) fue propuesto por primera vez por [Bullard, 1999] y supone un nuevo enfoque del problema de estimación de home range. Hasta el momento, las técnicas expuestas suponen independencia y las que son capaces de tratar con datos dependientes son sencillamente una modificación de las anteriores que de alguna forma, incorporan la coordenada temporal.

Pero en este caso, se propone aprovechar la naturaleza continua en espacio y tiempo del movimiento del animal para realizar la estimación utilizando un proceso estocástico, que nos dará una idea más realista del home range.

De entre los procesos estocásticos, nos interesarán los **paseos aleatorios**, pues son procesos estocásticos que describen el desplazamiento aleatorio de un móvil en  $\mathbb{R}^d$ . Supongamos que nuestro proceso estocástico  $\{X_t\}_{t \in \{0, \dots, T\}}$  está formado por las observaciones tomadas en cada uno de los instantes temporales  $t \in \{0, \dots, T\}$ . De este modo, si pensamos en un caso unidimensional, la suma de las

posiciones nos daría la trayectoria del móvil desde el inicio hasta el instante  $t$ , es decir  $Z_n = \sum_{i=0}^t X_i$  sería un paseo aleatorio.

Utilizando el Principio de Invarianza de Donsker [Karatzas and Shreve, 1991], no resulta complicado probar que, bajo ciertas condiciones, una sucesión de paseos aleatorios converge en distribución a un **movimiento browniano**. Por este motivo, el movimiento browniano será el proceso estocástico que utilizemos para modelizar la trayectoria del animal. En concreto, esta propuesta de Bullard pretende estimar la trayectoria del animal con una modificación del movimiento browniano: el **punte browniano**.

Para explicar en qué consiste el puente browniano, conviene introducir antes una pequeña definición del movimiento browniano.

**Definición 3.1.** [Billingsley, 1995] Un movimiento browniano es un proceso estocástico que denotaremos por  $\{B_t\}_{t \geq 0}$  sobre un espacio de probabilidad  $(\Omega, F, \mathbb{P})$  cumpliendo las siguientes propiedades:

- El proceso comienza en 0, es decir,  $\mathbb{P}(B_0 = 0) = 1$ .
- Los incrementos son independientes, es decir, para  $0 \leq t_1 < t_2 < \dots < t_k$  los incrementos  $B_{t_k} - B_{t_{k-1}}, B_{t_{k-1}} - B_{t_{k-2}}, \dots, B_{t_2} - B_{t_1}$  son variables aleatorias independientes.
- Para  $0 \leq s < t$ , el incremento  $B_t - B_s$  sigue una distribución normal de media 0 y varianza  $t - s$ .

Además, los incrementos del movimiento browniano  $B_t - B_s$  son estacionarios porque su distribución depende solamente de la diferencia  $t - s$ .

Teniendo esto en cuenta, el proceso estocástico de puente browniano se define como

$$X_t = B_t - tB_1, 0 \leq t \leq 1.$$

De esta forma, el proceso comienza y acaba en el mismo punto ya que  $X_0 = 0$  y  $X_1 = 0$ . Teniendo en cuenta que los incrementos de procesos brownianos siguen una distribución normal, resulta inmediato ver que la función de distribución de  $X$  también es normal.

Por tanto, la elección del puente browniano parece adecuada para la modelización de la trayectoria de un animal puesto que es un proceso continuo en el que la probabilidad de estar en un lugar determinado está condicionada por los puntos de inicio y final.

Sea  $Z_t^{\mathbf{a}, \mathbf{b}, T}$  la posición de un animal en  $t \in [0, T]$  siguiendo un paseo aleatorio desde  $\mathbf{a}$  hasta  $\mathbf{b}$ , con valores conocidos  $Z_0^{\mathbf{a}, \mathbf{b}, T} = \mathbf{a}$  y  $Z_T^{\mathbf{a}, \mathbf{b}, T} = \mathbf{b}$  en  $\mathbb{R}^2$ . Este proceso tiene distribución normal  $Z_t^{\mathbf{a}, \mathbf{b}, T} \sim N_2(\boldsymbol{\mu}(t), \sigma_t^2 I)$  donde

$$\boldsymbol{\mu}(t) = \mathbf{a} + \frac{t}{T}(\mathbf{b} - \mathbf{a}),$$

$$\sigma^2(t) = \frac{t(T-t)}{T} \sigma_m^2,$$

donde  $\sigma_m^2$  es el coeficiente de difusión relativo al movimiento del animal, que se estima empíricamente por verosimilitud, e  $I$  es la matriz identidad. Puede consultarse el procedimiento en [Horne et al., 2007].

Así, la posición de un animal que sigue una trayectoria aleatoria entre  $\mathbf{a}$  y  $\mathbf{b}$  en cualquier intervalo de tiempo desde  $t = 0$  hasta  $t = T$  puede estimarse con una distribución normal. Por tanto, la probabilidad de encontrar un animal en una región  $A$  del espacio en tiempo  $t \in [0, T]$  es:

$$\mathbb{P}(Z_t^T \in A) = \int \int \mathbb{P}(Z_t^{x,y} \in A) f_{\mathbf{a}}(x) f_{\mathbf{b}}(y) dx dy,$$

siendo  $f_{\mathbf{a}}(x)$  y  $f_{\mathbf{b}}(y)$  los puntos de inicio y final de la trayectoria respectivamente, que se consideran aleatorias y que son función de las variables  $x$  e  $y$ .

Tomemos ahora la función indicadora  $1_A(\mathbf{x})$ , que toma el valor 1 si  $\mathbf{x} \in A$  y el valor 0 en cualquier otro caso. Esta función nos da por tanto el tiempo de ocupación de la región  $A$ . Si dividimos por  $T$  (es decir, el tiempo total que ha durado la observación del animal) y tomamos la esperanza de la función indicadora, obtendremos la fracción de tiempo que el animal ha pasado en  $A$ , así podremos estimar la UD del animal con la densidad  $h(\mathbf{z})$  tal que:

$$\mathbb{E}\left[\frac{1}{T} \int_0^T 1_A(Z_t^T) dt\right] = \int_A h(\mathbf{z}) d\mathbf{z}.$$

Haciendo cálculos, que pueden consultarse detalladamente en [Horne et al., 2007], tenemos que:

$$h(\mathbf{z}) = \frac{1}{T} \int_0^T \int \int \hat{p}_t^{x,y,T}(x, z) f_{\mathbf{a}}(x) f_{\mathbf{b}}(x) dx dy dt,$$

donde  $\hat{p}_t^{x,y,T}(x, z)$  es la densidad de una normal con media  $\boldsymbol{\mu}(t)$  y varianza  $\sigma^2(t)I$ .

Así, ya tendríamos una estimación de la UD que podríamos utilizar directamente para la estimación del home range.

### 3.1.3. Modelo de movimiento browniano reflejado (MRBM)

En [Cholaquidis et al., 2016] se propone un método de estimación de un conjunto compacto  $S \subset \mathbb{R}^d$  a partir de una trayectoria de un proceso estocástico contenido en él.

Si pensamos el home range que queremos estimar como un conjunto compacto  $S$  y la trayectoria descrita por el animal como el proceso estocástico contenido en su interior, vemos que este método es fácilmente aplicable a la estimación del área de actividad de un animal.

En concreto, este método está a medio camino entre el cierre  $\alpha$ -convexo y el BBMM, pues intenta hacer una estimación de un conjunto compacto mediante la unión de bolas de radio  $\alpha$  al igual que en el cierre  $\alpha$ -convexo, inspirándose en un proceso estocástico como hacía el BBMM. Por tanto, el MRBM es una modificación del cierre  $\alpha$ -convexo que tiene en cuenta la dependencia temporal de las observaciones, pues modelizar la actividad del animal como un proceso estocástico es una forma más realista de tratar con datos correlados.

El proceso estocástico que se utiliza para describir la trayectoria es un movimiento browniano reflejado (RBM-reflected brownian motion), que tiene la particularidad de comportarse como un movimiento browniano usual en el interior del conjunto  $S$ , que se refleja en la frontera, estableciendo así los límites del conjunto. Se define del siguiente modo:

$$Z_t = B_t + RY_t,$$

donde  $B_t$  es un movimiento browniano,  $R \in \mathcal{M}(\mathbb{R})_{d \times d}$  es la matriz de reflexión e  $Y_t$  es un vector  $d$ -dimensional, continuo y no decreciente, tal que  $Y_0 = 0$ .

En [Cholaquidis et al., 2016] se puede consultar un estudio detallado de las relaciones entre las condiciones probabilísticas para la existencia del RBM y las restricciones geométricas sobre el soporte.

Además, en el documento, se proponen esencialmente dos métodos para la estimación del home range: **cierre  $\alpha$ -convexo de la trayectoria** y el **RBM sausage**.

En el primer caso, el cierre  $\alpha$ -convexo de la trayectoria no es más que la generalización de la Definición 2.1 para una trayectoria descrita por un RBM,  $C_\alpha(W_t)$  donde  $W_T = \{Z_t : 0 \leq t \leq T\}$  es la trayectoria de un RBM.

En cuanto a la RBM sausage, su filosofía está próxima a la del estimador tipo núcleo. El conjunto que estima al home range se define como:

$$D_T = B_{\epsilon_T}(W_T) = \{x \in \mathbb{R}^d : \exists t \in [0, T], \|\mathbf{x} - X_t\| \leq \epsilon_T\}.$$

A menudo, suponer que la trayectoria viene dada por un RBM puede traer limitaciones, como por ejemplo que la distribución estacionaria es necesariamente uniforme en el dominio, que puede no ser



realista para el movimiento del animal ya que con frecuencia habrá regiones del home range donde el animal pase más tiempo y por tanto, el uso que hace de la misma no será uniforme. Para corregir esta deficiencia, Cholaquidis propone la utilización de un RBM con deriva o más generalmente, de una clase de procesos estocásticos más amplia que engloban al RBM con deriva y que reciben el nombre de **difusiones reflejadas**, y no son más que una solución a una ecuación estocástica en la que variando el valor que toman los parámetros podemos obtener multitud de procesos estocásticos con distribuciones estacionarias no necesariamente uniformes.

## 3.2. Métodos localizados para datos dependientes

Al igual que en el caso de datos independientes, también en este contexto de dependencia temporal se han desarrollado técnicas de estimación local. Concretamente, expondremos una técnica que no es más que una modificación del LoCoH para datos dependientes.

### 3.2.1. T-LoCoH

Si recordamos brevemente, los métodos locales de estimación del home range comprenden fundamentalmente tres pasos:

- Agrupación de las observaciones en un número  $k$  de entornos.
- Estimación del home range en cada uno de los  $k$  entornos.
- Unión de las estimaciones del home range en cada uno de los entornos para obtener la estimación final.

En este caso, el T-LoCoH (Time-Local Convex Hull) propuesto en [Lyons et al., 2013], incorpora el tiempo en el paso 1 para tratar con observaciones dependientes.

En primer lugar, la agrupación de las observaciones se lleva a cabo mediante la técnica del vecino más próximo en cualquiera de las tres variantes explicadas en la Subsección 2.2.1, con la diferencia de que la métrica utilizada ya no es la euclídea, sino una métrica que incorpora la coordenada temporal que recibe el nombre de **TSD** (Time Scaled Distance) y se define del siguiente modo:

$$TSD_{ij}^2 = \|\mathbf{x}_i - \mathbf{x}_j\|^2 + s^2 v_{max}^2 (t_i - t_j)^2,$$

donde  $\mathbf{x}_i$  con  $i \in \{1, \dots, n\}$ , siendo  $n$  el número de observaciones,  $s \geq 0$  es un factor de escalado adimensional de la velocidad máxima teórica  $v_{max}$ . Cuando  $s = 0$  la TSD es equivalente a la distancia euclídea en  $\mathbb{R}^2$  y a medida que  $s$  crece, el tiempo va tomando mayor protagonismo en la métrica.

Tanto  $k$  como  $s$  se escogen por validación cruzada y  $v_{max}$  en base a estudios biológicos o por algún proceso estadístico.

Esta métrica transforma un intervalo de tiempo entre dos puntos cualesquiera en un tercer eje del espacio euclídeo. El efecto de este eje consiste en alejar los puntos lejanos en el tiempo pero próximos en distancia euclídea en el plano. De esta forma, los cierres producidos por este método son locales no sólo en términos espaciales, sino también en términos temporales ya que reflejan el movimiento del individuo en un lugar y tiempo determinados. Debido a esta característica, podemos deducir el uso que el sujeto en estudio hace de una determinada zona viendo puntos que están próximos en distancia euclídea sobre el plano pero que no han sido agrupados en un mismo entorno con el método de los vecinos más próximos con métrica  $TSD$ . Así, queda reflejado el uso que el animal hace de cada zona y la importancia de cada conjunto del home range.

Una vez agrupados los datos en  $k$  entornos, se hace la estimación local del home range en cada uno de ellos, siguiendo las instrucciones dadas en la Subsección 2.2.1 y finalmente, para unir los  $k$  estimadores locales, [Lyons et al., 2013] propone ordenarlos de menor a mayor siguiendo alguno de los siguientes criterios dependiendo del objetivo que se persiga: área encerrada por el estimador local, número de puntos encerrados o ratio perímetro-área (PAR).



## Capítulo 4

# Estimación del home range con datos simulados

En este capítulo analizaremos el funcionamiento de las técnicas expuestas en el Capítulo 2 para la estimación del área de actividad de un animal utilizando datos simulados y suponiendo independencia.

En primer lugar escogeremos un conjunto en el espacio bidimensional  $S \subset \mathbb{R}^2$  del cual conocemos el área y utilizando técnicas Monte Carlo, generaremos observaciones independientes e idénticamente distribuidas en su interior con las que estimaremos el home range empleando cada una de las técnicas descritas.

Puesto que en este caso el área original del conjunto es conocida, podremos hacer una comparación entre la estimación y la realidad, valorando así los pros y contras de cada método.

Dividiremos el capítulo en dos secciones, en la primera se evaluará la calidad de cada estimador para captar la geometría subyacente de los datos, suponiendo que el comportamiento del animal en la porción de área ha sido uniforme. La segunda, valorará el comportamiento del estimador sobre un conjunto de observaciones independientes e idénticamente distribuidas pero no uniformes en un disco de radio 1.

El estudio que aparece en este capítulo se ha realizado utilizando el software estadístico R, [core Team, 2017].

### 4.1. Simulación de observaciones uniformes en distintos conjuntos

Hemos generado muestras bidimensionales de tamaño  $n = 500$ , aleatorias, independientes y con una distribución uniforme en un conjunto  $S \subset \mathbb{R}^2$ . Se ha seleccionado este tamaño como un tamaño intermedio en que los estimadores son razonables. Por coste computacional no se han referido otros tamaños muestrales.

Se han escogido tres conjuntos distintos  $S \subset \mathbb{R}^2$  para evaluar la calidad de cada estimador dependiendo de la forma original del home range, intentando imitar aquellos que contienen zonas no visitadas en su interior.

Los conjuntos elegidos son los que aparecen en la Figura 4.1 y podemos definirlos formalmente como  $O = \{(x, y) \in \mathbb{R}^2 / 0.25^2 \leq (x-0.5)^2 + (y-0.5)^2 \leq 0.5^2\}$ ,  $S = \{(x, y) \in \mathbb{R}^2 / O \setminus \{(x > 0.5) \cap \{y < 0.5\}\} \cup \{0.25^2 \leq (x-0.5)^2 + (y+0.25)^2 \leq 0.5^2 \setminus \{(x < 0.5) \cap \{y > -0.25\}\}\}\}$  y  $C = \{(x, y) \in O \cap \{x \leq 0.5\}\}$ .

Con estos datos, es sencillo hallar el área de cada uno de los conjuntos  $O$ ,  $S$  y  $C$ , de esta forma, disponemos del área real del home range.

Para calcular el área estimada por cada una de las técnicas estudiadas, hemos realizado un estudio Monte Carlo con 300 repeticiones. En cada una de ellas, se genera una muestra aleatoria independiente

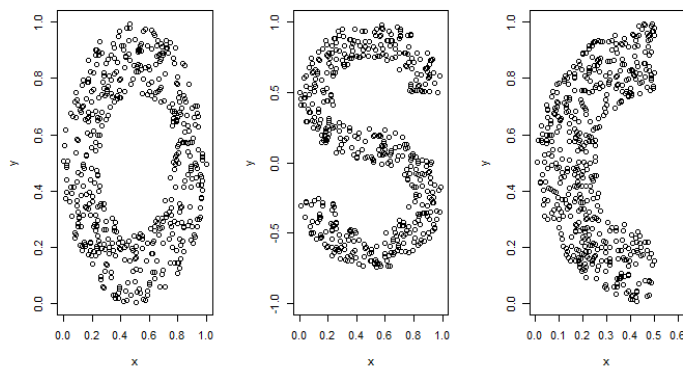


Figura 4.1: De izquierda a derecha, representación gráfica de 500 observaciones simuladas de una distribución uniforme en los conjuntos O, S y C respectivamente.

y con distribución uniforme de tamaño  $n = 500$  en el conjunto deseado y a partir de ellas, se estima el área del home range. Para ello se han utilizado funciones ya programadas en R. Una vez realizadas las 300 repeticiones, se hace una media del área del home range estimada por cada una de las técnicas.

Concretamente las técnicas que hemos utilizado son:

- **Cierre convexo (MCP):** programado en el paquete `adehabitatHR` [Calenge, 2006], en la función `mcp`. Esta función emplea efectivamente el Algoritmo 1 descrito en la Subsección 2.1.1, pero no utiliza la totalidad de puntos de la muestra, sino que elimina el 5% de las observaciones más extremas por considerarlas atípicas y que por tanto, no entran en la definición del home range dada por Burt, [Burt, 1943]. El criterio que sigue para eliminar este 5% es el de eliminar los puntos más alejados del centroide en distancia euclídea y el objetivo es reducir la sobrestimación inherente al método.
- **Estimación a través de la función de utilización (UD):** recordemos que estas técnicas intentan estimar la UD a partir de la muestra. La función utilizada para su cálculo de nuevo puede encontrarse en el paquete `adehabitatHR` con el nombre de `kernelUD`. El núcleo que hemos elegido ha sido un núcleo gaussiano bidimensional y otra vez, la técnica elimina el 5% de las observaciones más extremas. El parámetro de suavizado  $H$  se ha escogido utilizando distintos criterios ya descritos en la Subsección 2.1.3, concretamente: ventana plug-in, normal scale rule y validación cruzada. El paquete utilizado para la selección de la ventana óptima con los diferentes criterios ha sido `sm`, [Bowman and Azzalini, 1999].
- **Single-linkage cluster:** también está programada en el paquete `adehabitatHR` en la función `clusthr`. La técnica utilizada es exactamente la explicada en la Subsección 2.2.3, propuesta en [Kenward et al., 2001].

En cuanto a las técnicas que precisan de la elección de un parámetro para hacer la estimación, hemos analizado el funcionamiento del cierre convexo local (LoCoH), que como hemos comentado en el Capítulo 2, requiere de la elección de un  $k$ , un  $r$  o una  $a$ . También se ha estudiado la **técnica de estimación global** de cierre  $\alpha$ -convexo, que precisa de la elección del parámetro  $\alpha$ . Para estos casos, hemos decidido optar por una representación gráfica de la variación del área en función de la elección del parámetro y así observar su evolución.

Para el cálculo del área estimada del home range con el cierre convexo local (LoCoH), hemos utilizado la función `LoCoH` del paquete `adehabitatHR`, que permite escoger los parámetros  $k$ ,  $r$  o  $a$ . En cuanto al cierre  $\alpha$ -convexo, la función empleada ha sido `ahull` del paquete `alphahull`, [Pateiro-López and Rodríguez-Casal, 2010].

En el caso de la selección de  $k$ , el rango de valores oscila entre  $k = 50$  y  $k = 500$ , es decir, se crearán clusters cuyo tamaño varía entre los 50 puntos y los 500. Como la muestra es de tamaño  $n = 500$ , para  $k = 500$  tendríamos un único clúster, por tanto sería un caso particular del cierre convexo, aunque obsérvese que en este caso el cierre convexo estima el área con el 95 % de la muestra mientras el LoCoH lo hace con la totalidad, por tanto la estimación del área dada por MCP será ligeramente menor que la dada por LoCoH con el método del  $k$ -vecino más próximo para  $k = 500$ .

En cuanto al  $r$ , sería el radio de la circunferencia que forma cada clúster. El rango de valores escogido para este parámetro es una secuencia desde  $r = 0.1$  hasta  $r = 1$ , avanzando de 0.1 en 0.1.

El parámetro  $a$ , si recordamos representa el radio adaptativo de la circunferencia y es por tanto la suma de la distancia de cada punto que pertenece al clúster al punto de referencia. Los valores escogidos para  $a$  van desde  $a = 5$  hasta  $a = 40$  avanzando de 5 en 5.

Por último, los valores del parámetro  $\alpha$  han sido escogidos entre  $\alpha = 0.1$  y  $\alpha = 0.5$  avanzando de 0.05 en 0.05.

A continuación, se muestran los resultados obtenidos para la estimación del área del home range con las distintas técnicas en las Tablas 4.1, 4.2 y 4.3.

■ Figura  $O$

En la Tabla 4.1 podemos observar las estimaciones del área del home range con cada una de las técnicas para la figura  $O$  que tiene un área real de 0.5890 unidades cuadradas. En todos los casos la estimación obtenida es mayor, como cabía esperar, pues la sobrestimación es generalmente el problema que presentan los métodos globales de estimación para datos independientes. Además, en este caso concreto, el conjunto que se intenta estimar tiene un agujero en medio que estas técnicas no han sido capaces de detectar. Si es cierto que este problema se soluciona con el método localizado del *single-linkage cluster*, puesto que la superficie que estima en este caso no tiene por qué ser conexa, lo que disminuye el área de la misma.

| Técnica                | Área media estimada |
|------------------------|---------------------|
| MCP                    | 0.720               |
| UD (plug-in rule)      | 0.727               |
| UD (normal scale rule) | 1.100               |
| UD (cross validation)  | 0.743               |
| Single-linkage cluster | 0.517               |

Tabla 4.1: Área media estimada del home range para cada una de las técnicas indicadas, a partir de observaciones generadas en el conjunto  $O \in \mathbb{R}^2$ , con un área real de 0.5890 unidades.

En cuanto a las técnicas de cierre convexo local LoCoH, podemos ver en la Figura 4.2 que incluso partiendo de un clúster con un radio muy pequeño o con un número pequeño de observaciones en su interior, se sobrestima el área original del conjunto.

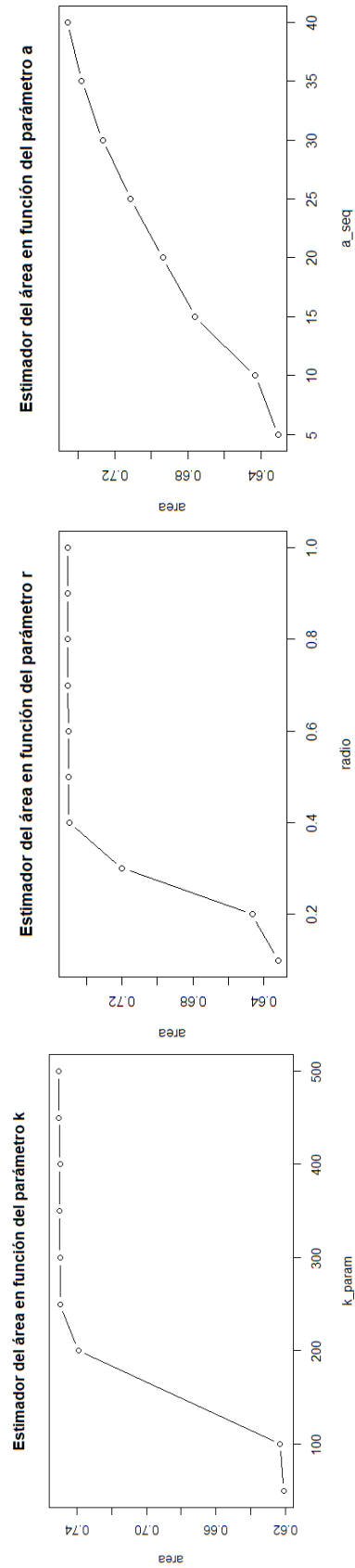


Figura 4.2: De izquierda a derecha, evolución del área media estimada del home range en función de la elección de los parámetros  $k$ ,  $r$  y  $a$  en el cierre convexo local (LoCoH) para el conjunto  $O$ , área real 0.5890.

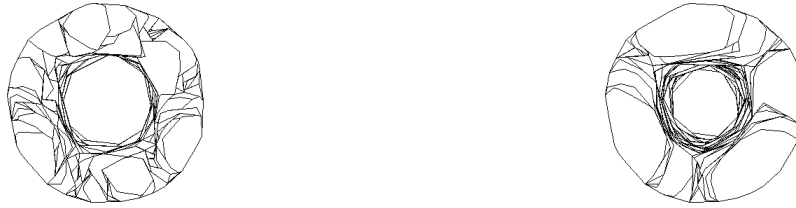


Figura 4.3: Cierre convexo local (LoCoH) con el método del k-vecino más próximo para el conjunto  $O$ , área real 0.5890. De izquierda a derecha los valores escogidos de  $k$  son  $k = 50$ ,  $k = 100$ .



Figura 4.4: Cierre convexo local (LoCoH) con el método del k-vecino más próximo para el conjunto  $O$ , área real 0.5890. De izquierda a derecha los valores escogidos de  $k$  son  $k = 150$ ,  $k = 450$ .

Podemos observar en más detalle en las Figuras 4.3 y 4.4 que a medida que aumenta el valor de  $k$ , al método le cuesta más distinguir las zonas no visitadas del interior del conjunto. Si nos fijamos por ejemplo en el caso  $k = 50$ , los clusters formados son pequeños porque contienen pocas observaciones y esto se traduce en una mayor precisión, puesto que ha sido capaz de distinguir que en el centro efectivamente, no existen al menos 50 observaciones para formar un cluster. En cambio, para  $k = 450$ , el método crea un gran clúster que prácticamente engloba a todas las observaciones de la muestra, que recordemos, era de tamaño  $n = 500$ , y por tanto no es capaz de distinguir un agujero en el centro.

Por último, en el cierre  $\alpha$ -convexo, en la Figura 4.5 observamos que se produce un rápido incremento del área media estimada del home range principalmente cuando pasamos de  $\alpha = 0.25$  a  $\alpha = 0.3$ . Es precisamente en ese intervalo donde se alcanza el área real. A partir de ese punto, el área media estimada se estabiliza entorno a valores próximos a 0.75 que de nuevo, sobrestiman el área del conjunto  $O$ , sobrepasando ligeramente la estimación obtenida con la técnica del MCP.

En la Figura 4.6, se incluye una representación gráfica del área media estimada para el conjunto  $O$  con la técnica del cierre  $\alpha$ -convexo para los valores de  $\alpha = 0.1$  y  $\alpha = 0.28$ . Podemos apreciar que para el primer caso, el método es capaz de detectar el agujero en el interior del home range, en cambio en el segundo, estima el área directamente como un círculo, de ahí la sobrestimación.

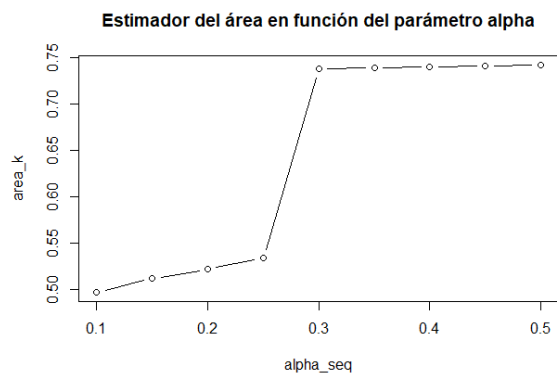


Figura 4.5: Evolución del área media estimada del home range en función de la elección del parámetro  $\alpha$  en el cierre  $\alpha$ -convexo para el conjunto  $O$ , área real 0.5890.

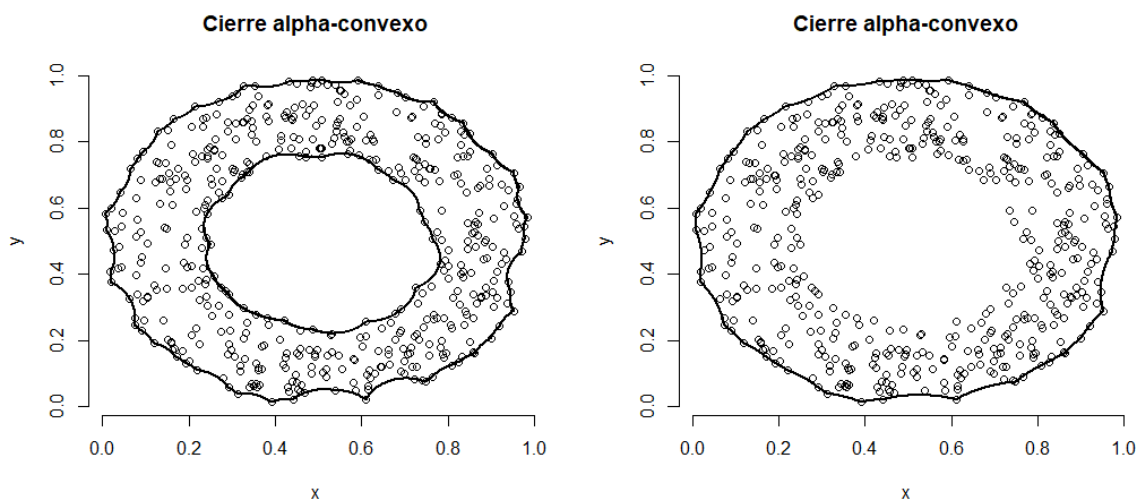


Figura 4.6: De izquierda a derecha, cierre  $\alpha$ -convexo para el conjunto  $O$ , área real 0.5890, tomando los valores de  $\alpha = 0.1$  y  $\alpha = 0.28$ , respectivamente.



- Figura  $S$

Para el conjunto  $S$ , con un área real de 0.8835 unidades cuadradas, podemos extraer conclusiones similares. En la Tabla 4.2, vemos que de nuevo todas las técnicas han sobrestimado el área real del home range, a excepción del single-linkage clúster.

| Técnica                | Área media estimada |
|------------------------|---------------------|
| MCP                    | 1.426               |
| UD (plug-in rule)      | 1.057               |
| UD (normal scale rule) | 1.753               |
| UD (cross validation)  | 1.015               |
| Single-linkage cluster | 0.743               |

Tabla 4.2: Área media estimada del home range para cada una de las técnicas indicadas, a partir de observaciones generadas en el conjunto  $S \in \mathbb{R}^2$ , con un área real de 0.8835 unidades.

En la Figura 4.7 observamos una representación gráfica de la estimación del home range realizada con single-linkage clúster, donde apreciamos que efectivamente el área media estimada no es conexas, lo que produce una infraestimación del área real del home range.



Figura 4.7: Estimación del home range para el conjunto  $S \in \mathbb{R}^2$  utilizando la técnica de single-linkage clúster. El área media estimada es de 0.743 mientras que el área real del conjunto es 0.8875.

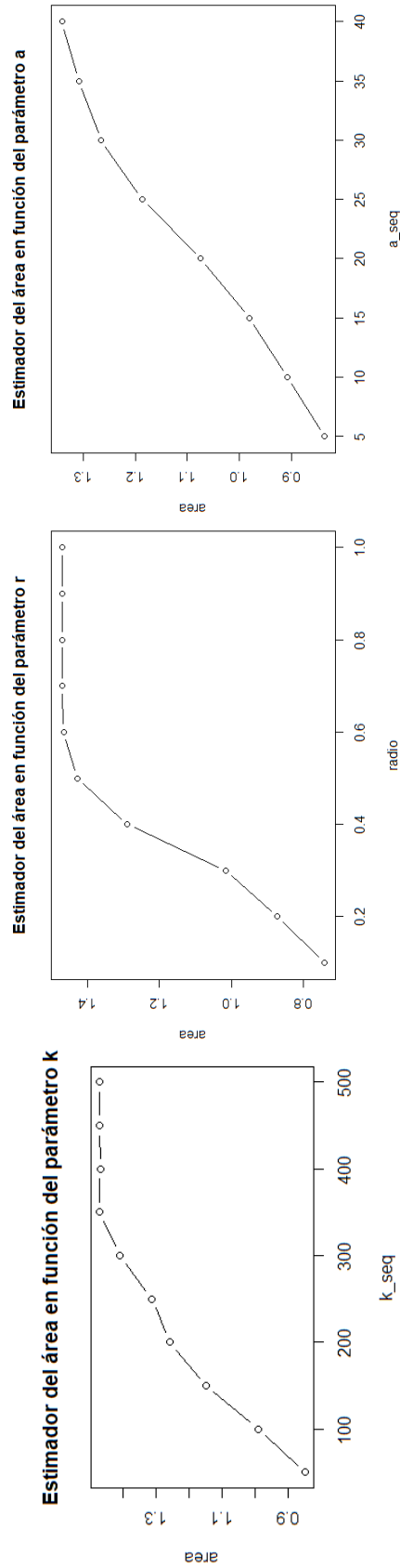


Figura 4.8: De izquierda a derecha, evolución del área media estimada del home range en función de la elección de los parámetros  $k$ ,  $r$  y  $a$  en el cierre convexo local (LoCoH) para el conjunto  $S$ , área real 0.8835.

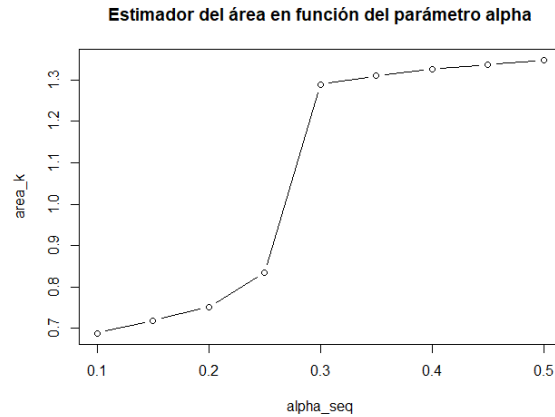


Figura 4.9: Evolución del área media estimada del home range en función de la elección del parámetro  $\alpha$  en el cierre  $\alpha$ -convexo para el conjunto  $S$ , área real 0.8875.

En el LoCoH, Figura 4.8a, para  $k = 50$  se hace una estimación muy cercana a la original. A partir de ese valor se sobrestima el área, estabilizándose en  $k = 350$  en torno a 1.5 unidades cuadradas, que sobrepasan notablemente a las 0.8835 originales.

Para la elección de  $r$  y  $a$ , Figuras 4.8b y 4.8c respectivamente, igualmente las estimaciones más próximas al valor real se obtienen para valores muy pequeños del parámetro.

En cuanto al cierre  $\alpha$ -convexo, en la Figura 4.9, vemos que de nuevo se produce un incremento notable del área estima del conjunto entre  $\alpha = 0.25$  y  $\alpha = 0.3$ , siendo en este intervalo donde se obtienen las mejores aproximaciones del área real.

■ Figura  $C$

Para terminar, evaluamos el estudio realizado sobre el conjunto  $C \in \mathbb{R}^2$ , con un área de 0.2945 unidades cuadradas. De nuevo los resultados obtenidos en la Tabla 4.3 son los esperados, pues son análogos a los anteriores. Las técnicas de estimación globales para datos independientes que aparecen en la Tabla 4.3, continúan sobrestimando el área del home range, a excepción, como cabría esperar, del single-linkage clúster.

| Técnica                | Área media estimada |
|------------------------|---------------------|
| MCP                    | 0.362               |
| UD (plug-in rule)      | 0.388               |
| UD (normal scale rule) | 0.675               |
| UD (cross validation)  | 0.425               |
| Single-linkage cluster | 0.239               |

Tabla 4.3: Área media estimada del home range para cada una de las técnicas indicadas, a partir de observaciones generadas en el conjunto  $C \in \mathbb{R}^2$ , con un área real de 0.2945 unidades.

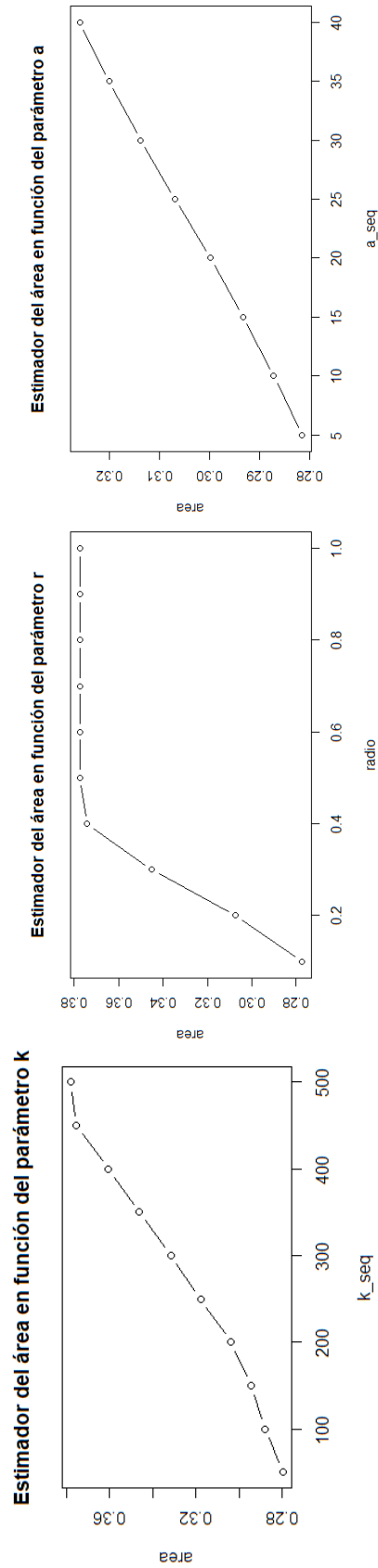


Figura 4.10: De izquierda a derecha, evolución del área media estimada del home range en función de la elección de los parámetros  $k$ ,  $r$  y  $a$  en el cierre convexo local (LoCoH) para el conjunto  $C$ , área real 0.2945.

Para el LoCoH con el método del  $k$ -vecino más próximo observamos en la Figura 4.10a que para los valores más pequeños de  $k$  se producen las estimaciones del área más precisas. A medida que aumentamos el valor del parámetro, aumenta la sobrestimación del área.

En el LoCoH con radio fijo  $r$ , vemos en la Figura 4.10b que de nuevo la estimación más precisa se obtiene entre  $r = 0.1$  y  $r = 0.2$ . Además, a partir de  $r = 0.4$  la estimación del área se estabiliza. En la Figura 4.11 observamos que esta estabilización se debe a que a partir de  $r = 0.4$  se produce la máxima sobrestimación y esta se mantiene a pesar de aumentar el radio.

En cuanto al LoCoH con radio adaptativo de la Figura 4.10c, se observa que el área aumenta de forma constante a medida que aumenta el valor de  $a$ . En este caso no se estabiliza el área porque no se ha llegado al punto de sobrestimación máxima como ocurría en la Figura 4.10b, como podemos apreciar en las Figuras 4.11 y 4.12.

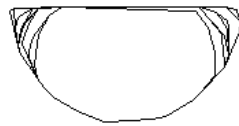
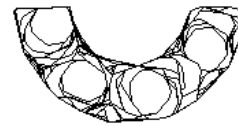
LoCoH con  $r=0.4$ LoCoH con  $r=0.1$ 

Figura 4.11: LoCoH con el método de radio fijo para el conjunto  $C$ , área real 0.2945, tomando los valores de  $r = 0.1$  y  $r = 0.4$  respectivamente.

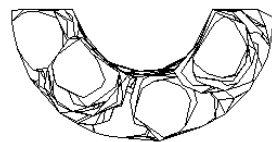
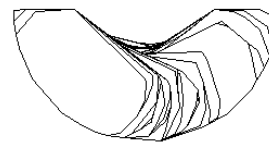
LoCoH con  $a=5$ LoCoH con  $a=40$ 

Figura 4.12: LoCoH con el método de radio adaptativo para el conjunto  $C$ , área real 0.2945, tomando los valores de  $a = 5$  y  $a = 40$  respectivamente.

Por último, para el cierre  $\alpha$ -convexo, en la Figura 4.13 vemos que entre  $\alpha = 0.25$  y  $\alpha = 0.3$  se produce un incremento más pronunciado del área, siendo de nuevo el intervalo donde se alcanza el área real.

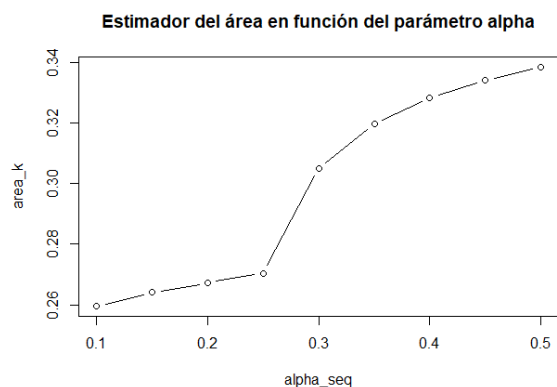


Figura 4.13: Evolución del área media estimada del home range en función de la elección del parámetro  $\alpha$  en el cierre  $\alpha$ -convexo para el conjunto  $C$ , área real 0.2945.

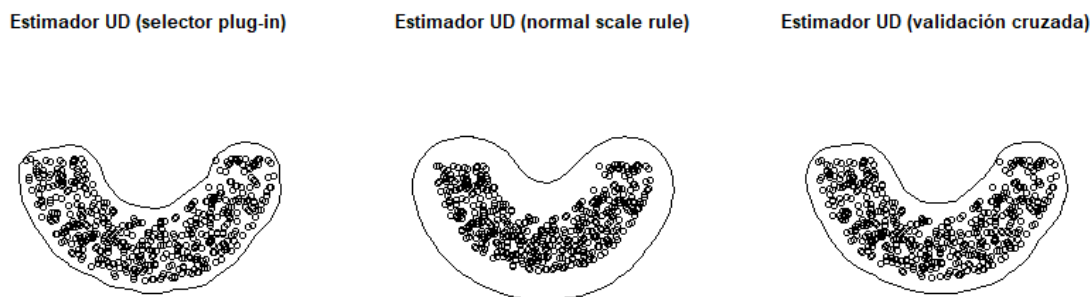


Figura 4.14: Estimación del home range para el conjunto  $C \in \mathbb{R}^2$  a través de la función de utilización empleando distintos criterios para la selección del parámetro de ventana.

En términos generales, se observa que las técnicas de estimación global para datos independientes sobreestiman el área del home range, principalmente porque no son capaces de distinguir los agujeros existentes en el interior de los conjuntos a estimar. Es especialmente llamativo el caso de la estimación de la función de utilización escogiendo el parámetro de ventana con el criterio de *normal scale rule*, pues en todos los casos estima un área de prácticamente el doble que el área real. En cambio, la estimación de la función de utilización utilizando como criterio para seleccionar el parámetro de ventana la regla plug-in, presenta un funcionamiento aceptable, con menor sobreestimación. Este hecho pone en evidencia la principal flaqueza de este método, que es la elección del parámetro de suavizado.

En la Figura 4.14, podemos observar la estimación del home range para el conjunto  $C \in \mathbb{R}^2$  mediante la función de utilización y empleando distintos criterios para elegir el parámetro de ventana. Efectivamente el selector plug-in presenta la menor sobreestimación.

En cuanto al cierre convexo y cierre  $\alpha$ -convexo, aunque también están afectados por la sobreestimación, en el segundo caso, una elección adecuada del  $\alpha$  minimiza este problema.

Por último, las técnicas locales, principalmente el single-linkage cluster, se han demostrado efectivas a la hora de detectar regiones no visitadas en el interior del home range. Si pensamos en el LoCoH, hemos observado que valores pequeños de los parámetros son capaces de reproducir con bastante precisión la forma del home range aunque de nuevo, la elección correcta del parámetro puede no ser trivial. Teniendo esto en cuenta, de entre los tres tipos de LoCoH estudiado, el LoCoH con radio adaptativo parece ser el que menos sufre por una mala elección del parámetro.

## 4.2. Simulación de observaciones aleatorias en un disco

En esta sección evaluaremos el comportamiento de cada estimador cuando el home range es un disco de radio 1 pero ahora, a diferencia del caso anterior, el comportamiento del animal no tiene por qué ser uniforme.

Esta es una situación más realista, pues los movimientos de los animales no solo dependen de barreras geográficas, sino también de sus comportamientos o necesidades, como pueden ser la disponibilidad de alimento o la presencia de depredadores. Por tanto es esperable que presenten preferencia por zonas concretas del home range y que no hagan un uso uniforme del mismo.

En este caso tendremos en cuenta tres distribuciones generadas en un disco de radio  $r = 1$ , siguiendo el ejemplo de [Bath et al., 2006]: normal bidimensional restringida al disco de radio unidad, normal bidimensional invertida y uniforme.

- **Normal bidimensional restringida al disco de radio unidad:** algunos animales tienden a pasar el mayor tiempo posible en una área central o núcleo, por lo que la probabilidad de encontrarlo disminuye a medida que aumenta la distancia al núcleo. Este caso vendría representado por tanto por una normal bidimensional restringida al disco unidad. Para simularla hemos generado una normal bidimensional estándar quedándonos solo con aquellas observaciones que caían dentro del disco de radio 1.
- **Normal bidimensional invertida:** otros animales, en cambio, presentan un comportamiento fronterizo, pues pasan en la frontera del home range la mayor parte del tiempo. Este será el comportamiento que representa la normal bidimensional invertida.

Su expresión matemática es la siguiente:

$$f(x, y) = \frac{1}{c} \left( 1 - \exp - \frac{1}{2}(x^2 + y^2) \right),$$

con  $c = 2\pi \left( \exp - \frac{\sigma^2}{2} + \frac{\sigma^2}{2} - 1 \right)$  y siendo  $\sigma$  el radio del disco donde queremos generar los datos, en este caso  $\sigma = 1$ .

- **Uniforme:** es el escenario más sencillo y representa un uso uniforme del área disponible. Hemos decidido tener este caso también en cuenta para compararlo con los otros dos considerados.

En la Figura 4.15, puede observarse una representación de los comportamientos considerados.

En este caso, de nuevo, disponemos del área real del home range, pues al ser un disco de radio 1, su área será  $\pi$ .

El estudio se ha llevado a cabo en las mismas condiciones que el realizado en la Subsección 4.1, mediante técnicas Monte Carlo, para muestras de tamaño  $n = 500$  y con  $B = 300$  repeticiones. De nuevo, las funciones utilizadas para la estimación del área del home range con cada técnica son las programadas en R, explicadas en la Subsección 4.1. Además, presentaremos los datos siguiendo la misma estructura.

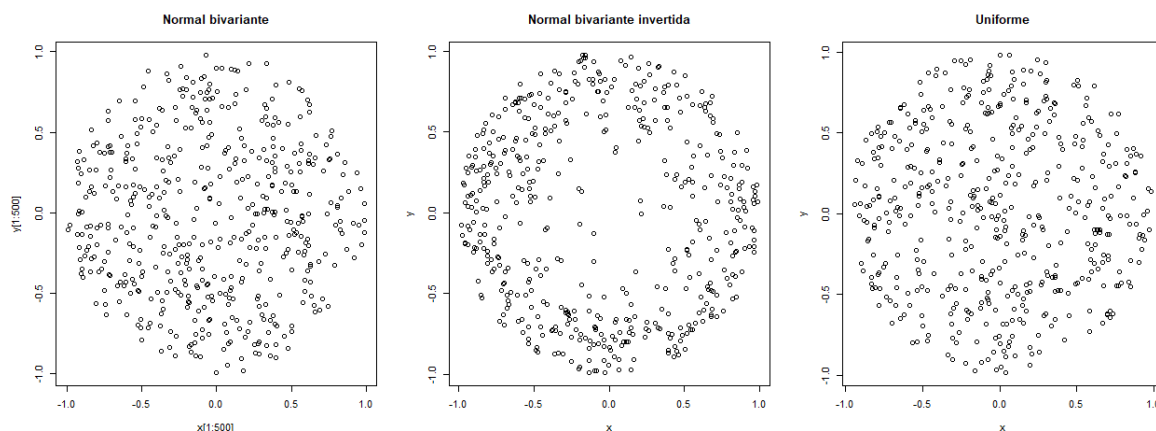


Figura 4.15: Representación gráfica de 500 observaciones independientes en el interior de un disco de radio 1, considerando, de izquierda a derecha, distribución normal bivalente, distribución normal bivalente invertida y distribución uniforme.

- Distribución normal bidimensional

Suponiendo que las observaciones de los movimientos del animal siguen una distribución normal bidimensional en un disco de radio 1, hemos obtenido los siguientes resultados para **técnicas globales con datos independientes** que pueden consultarse en la Tabla 4.4. Observamos que las técnicas basadas de la función de utilización continúan sobrestimando el área real, y de nuevo, el criterio plug-in para la selección del parámetro óptimo de suavizado es el que presenta los resultados más próximos al valor real.

| Técnica                | Área media estimada |
|------------------------|---------------------|
| MCP                    | 2.741               |
| UD (plug-in rule)      | 3.392               |
| UD (normal scale rule) | 3.780               |
| UD (cross validation)  | 3.610               |
| Single-linkage cluster | 2.517               |

Tabla 4.4: Área media estimada del home range para cada una de las técnicas indicadas, a partir de observaciones generadas con una distribución normal bivalente en un disco de radio 1 con una área real de 3.1415.

El método del single-linkage cluster infraestima ligeramente el home range y, del mismo modo, al contrario que lo que ocurría en la sección anterior, la técnica del cierre convexo (MCP) también infraestima la porción de área utilizada, porque ahora el conjunto a estimar ya no tiene agujeros y además el hecho de que se elimine el 5% de las localizaciones más alejadas del centroide por considerarlas atípicas, justifica la infraestimación.



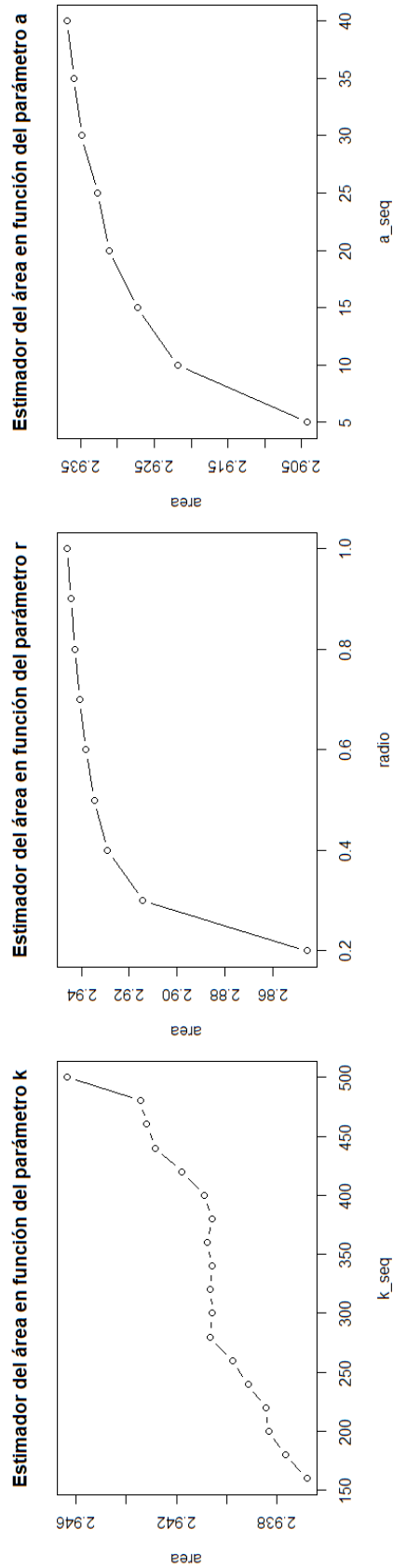


Figura 4.16: De izquierda a derecha, evolución del área media estimada del home range en función de la elección de los parámetros  $k$ ,  $r$  y  $a$  en el cierre convexo local (LoCoH) para  $n = 500$  observaciones normales bidimensionales en un disco de radio 1 con área real de 3.1415.

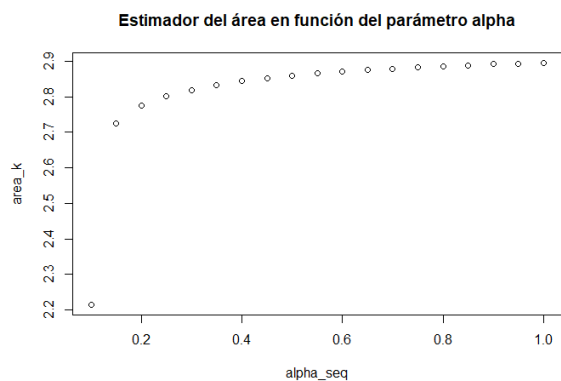


Figura 4.17: Evolución del área media estimada del home range en función de la elección del parámetro  $\alpha$  en el cierre  $\alpha$ -convexo para  $n = 500$  observaciones normales bidimensionales en un disco de radio 1 con área real de 3.1415.

En cuanto a las **técnicas de estimación local para datos independientes**, en el caso del cierre convexo local (LoCoH), Figura 4.16, resulta llamativo que tanto para el método del  $k$ -vecino más próximo, como para el de radio fijo  $r$  o el de radio adaptativo  $a$  se infraestima el área real del home range. Probablemente porque el conjunto es convexo.

Análogamente, en la Figura 4.17 vemos que para el cierre  $\alpha$ -convexo de nuevo se infraestima el área real, quedando la estimación máxima en torno a los mismos valores que el LoCoH, a pesar de ser este un método de estimación global.

- Distribución normal bidimensional invertida

Cuando el comportamiento del animal es fronterizo, cabe esperar que la estimación del área sea más próxima al área real o por lo menos, sean mayores que en el caso anterior.

Así lo observamos en la Tabla 4.5, en la que la estimación dada por el MCP, la distribución de utilización con ventana seleccionada por la *normal scale rule* y el single-linkage clúster son ligeramente mayores que en el caso anterior.

| Técnica                | Área media estimada |
|------------------------|---------------------|
| MCP                    | 2.936               |
| UD (plug-in rule)      | 3.260               |
| UD (normal scale rule) | 4.665               |
| UD (cross validation)  | 3.225               |
| Single-linkage cluster | 2.591               |

Tabla 4.5: Área media estimada del home range para cada una de las técnicas indicadas, a partir de observaciones generadas con una distribución normal bivalente invertida en un disco de radio 1, con una área real de 3.1415.

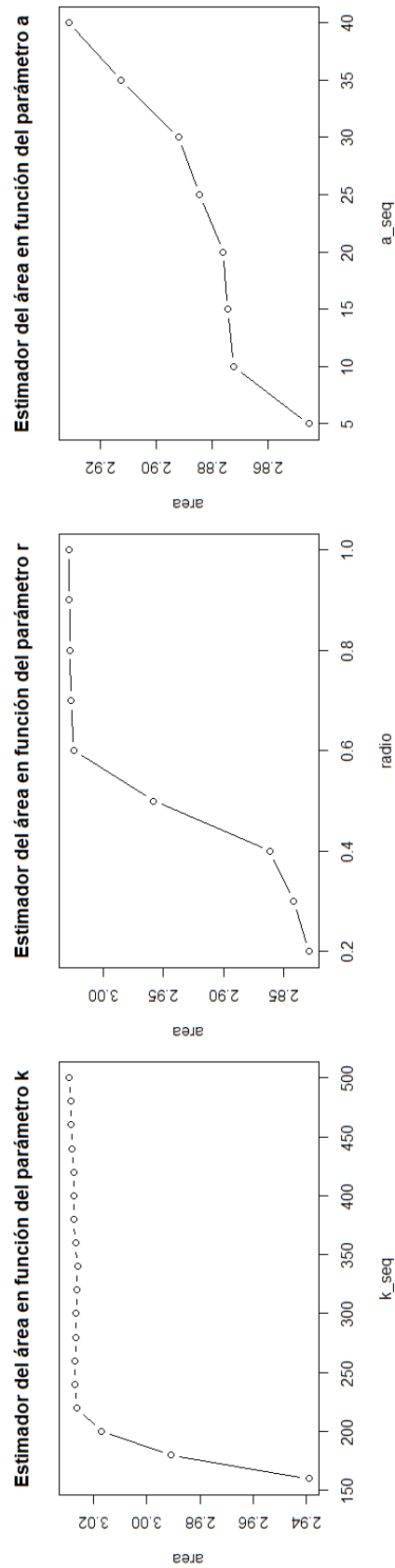


Figura 4.18: De izquierda a derecha, evolución del área media estimada del home range en función de la elección de los parámetros  $k$ ,  $r$  y  $a$  en el cierre convexo local (LoCoH) para  $n = 500$  observaciones de una normal invertida bidimensionales en un disco de radio 1 con área real de 3.1415.

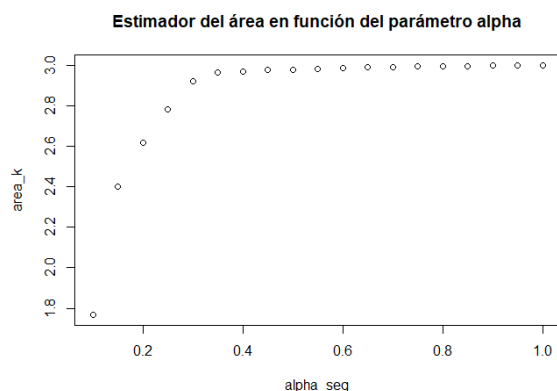


Figura 4.19: Evolución del área media estimada del home range en función de la elección del parámetro  $\alpha$  en el cierre  $\alpha$ -convexo para  $n = 500$  observaciones de una normal invertida bidimensional en un disco de radio 1 con área real de 3.1415.

En la Figura 4.18 observamos que para las distintas variantes del LoCoH, se alcanza la área real para valores grandes de los parámetros  $k$ ,  $r$  y  $a$ . Además como era de esperar, incluso para valores pequeños de los parámetros, la estimación que se obtiene del área del home range supera los valores obtenidos en el apartado anterior para observaciones generadas a partir de una normal bidimensional.

En cuanto al método de estimación global de cierre  $\alpha$ -convexo, en la Figura 4.19, vemos la evolución de la estimación del área en función del parámetro  $\alpha$ . La estimación es muy pequeña para valores pequeños del parámetro pero finalmente se estabiliza en torno a valores próximos aunque ligeramente menores que el área real y de nuevo mayores a los obtenidos en el mismo contexto para la distribución normal bidimensional.

- Distribución uniforme

Analizaremos por último la estimación del área del home range cuando la distribución de observaciones es uniforme en el disco de radio 1.

| Técnica                | Área media estimada |
|------------------------|---------------------|
| MCP                    | 2.811               |
| UD (plug-in rule)      | 3.459               |
| UD (normal scale rule) | 4.001               |
| UD (cross validation)  | 3.566               |
| Single-linkage cluster | 2.537               |

Tabla 4.6: Área media estimada del home range para cada una de las técnicas indicadas, a partir de observaciones generadas con una distribución uniforme en un disco de radio 1, con una área real de 3.1415.

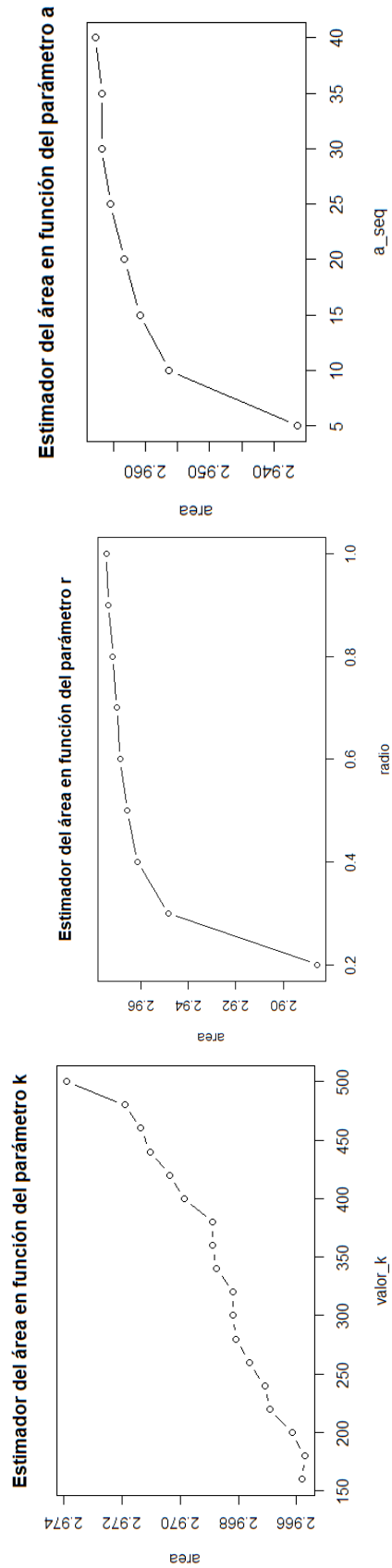


Figura 4.20: De izquierda a derecha, evolución del área media estimada del home range en función de la elección de los parámetros  $k$ ,  $r$  y  $a$  en el cierre convexo local (LoCoH) para  $n = 500$  observaciones de una distribución uniforme en un disco de radio 1 con área real de 3.1415.

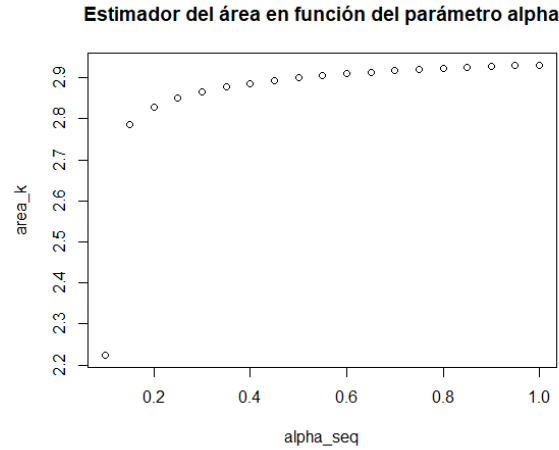


Figura 4.21: Evolución del área media estimada del home range en función de la elección del parámetro  $\alpha$  en el cierre  $\alpha$ -convexo para  $n = 500$  observaciones de una distribución uniforme en un disco de radio 1 con área real de 3.1415.

Al igual que en los casos anteriores, el MCP y el single-linkage infraestiman el área del home range, como puede verse en la Tabla 4.6. En cuanto a los métodos basados en la estimación de la función de utilización con la elección de distintos parámetros de ventana, continúan sobrestimando el home range, obteniéndose resultados similares a los observados en la Tabla 4.5 para observaciones obtenidas a partir de una normal invertida bidimensional.

Las tres técnicas utilizadas para la estimación con el LoCoH infraestiman el área real, quedando incluso por debajo de las estimaciones realizadas con la misma técnica para las observaciones de una normal invertida bidimensional, como puede verse en la Figura 4.20.

Por último, los comentarios hechos para el caso de LoCoH, son igualmente válidos para la estimación realizada con el cierre  $\alpha$ -convexo, de nuevo se infraestima el área real y es también menor que la obtenida cuando las observaciones proceden de una normal invertida bidimensional, como puede apreciarse en la Figura 4.21.

En términos generales, observamos que las técnicas efectivamente han sabido captar la distribución subyacente de los datos, pues cuando el comportamiento es nuclear, las estimaciones fueron más pequeñas. Así mismo, cuando el comportamiento es territorial, la estimación ha resultado ser incluso superior que cuando es uniforme.

Analicemos ahora el funcionamiento de las técnicas una a una.

El MCP tiene dificultades para distinguir los agujeros en el interior del home range, es por ello que siempre ha sobrestimado el área del home range cuando la simulación considerada era uniforme sobre las distintas figuras:  $O$ ,  $C$  y  $S$ . En cambio, para la estimación del home range en un disco de radio 1, considerando distintas distribuciones, el MCP infraestima el home range e incluso es capaz de distinguir entre los distintos comportamientos de los animales.

Las técnicas que emplean la densidad de utilización para realizar la estimación del home range sobrestiman en todos los casos, por lo que hemos observado, además de no identificar claramente las zonas no visitadas en el interior del home range, tampoco se presentan particularmente efectivas a la hora de captar la distribución de los datos.

En cuanto al single-linkage cluster, hemos visto que en todos los casos infraestima el área real porque no estima un home range conexo. Probablemente este problema se solventaría a medida que se aumente el tamaño muestral.

El LoCoH, al ser una técnica local, presenta resultados bastante aceptables, aunque la elección del parámetro  $k$ ,  $r$  o  $a$  no es trivial y mientras en la primera mitad de este Capítulo el área real se alcanzaba para valores pequeños de los parámetros, en la segunda mitad se alcanza para valores más grandes, estabilizando la estimación en torno al área real y reduciendo la sobrestimación que se había observado anteriormente. Por tanto, se demuestra una técnica efectiva para distinguir agujeros en el interior del home range y captar la distribución subyacente de los datos siempre y cuando se haga una adecuada elección de los parámetros.

Por último, los comentarios relativos al LoCoH son también válidos para el cierre  $\alpha$ -convexo. Mientras que en la Subsección 4.1 se estabilizaba en valores que sobrestimaban el área, en la Subsección 4.2 ha resuelto este problema, lo cual parece lógico teniendo en cuenta que en este segundo caso el home range carecía de agujeros en su interior.





## Capítulo 5

# Aplicación sobre un conjunto de datos real

En el capítulo anterior hemos analizado el funcionamiento de las técnicas locales y globales suponiendo independencia en los datos, pues eran datos simulados.

Pero en la práctica, debido a los avances en las técnicas de recolección de datos, lo más habitual es que las observaciones utilizadas para la estimación del home range presenten dependencia temporal. Por este motivo, destinaremos este capítulo al estudio del funcionamiento de las distintas técnicas, tanto para datos dependientes como para datos independientes (en ambos casos locales y globales) sobre un conjunto de datos reales.

Concretamente, los datos que vamos a utilizar son observaciones bidimensionales, correspondientes a la latitud y longitud de las localizaciones de un león de la reserva de Tsavo (Kenia), llamado Kiboche, perteneciente a la especie *Panthera leo*. Se dispone de 1465 observaciones tomadas entre el 22 de abril de 2002 a las 22:04:00:00 y el 21 de julio de 2007 a las 00:01:58:00.

Los datos han sido recogidos con un sistema GPS y las observaciones fueron tomadas en instantes aleatorios, sin guardar siempre el mismo intervalo de tiempo entre cada par. Los detalles del estudio así como las observaciones pueden consultarse en <https://www.movebank.org/>. Movebank es una base de datos online que da libre acceso a observaciones sobre los movimientos de los animales, tomadas por distintos investigadores y empleando diversas técnicas. En nuestro caso, el investigador que ha recogido los datos ha sido Bruce Patterson, biólogo e investigador en el Museo de Historia Natural de Chicago, experto en el estudio y conservación de los leones.



Figura 5.1: Imagen tomada de dos leones en el Parque Natural de Tsavo, en Kenya [Rhee, 2013] // Wikimedia Commons.



Figura 5.2: Imagen del Parque Natural de Tsavo Este [Ciesielski, 2012] // Wikimedia Commons.

A continuación dedicaremos un breve apartado al estudio de la especie y su entorno, pues para hacer una estimación adecuada del home range, se deben recoger datos sobre el hábitat, los recursos y las características del territorio, con la finalidad de entender en profundidad el comportamiento de los animales. Así podremos evaluar el funcionamiento de las distintas técnicas de estimación del home range, ya que ahora no disponemos del área real para hacer comparación.

El resto del capítulo estará dedicado a las estimaciones del home range con distintas técnicas.

## 5.1. Estudio de la especie y su hábitat

Los *Panthera leo* son grandes mamíferos con alimentación carnívora que viven en herbazales de sabana. Su dieta se compone de ñus, impalas, cebras, búfalos y gacelas, entre otros.

Aunque llegó a ser el segundo mamífero, después del ser humano, que más área ocupaba en la superficie terrestre, el censo de leones en 1950 rondaba los 400 mil ejemplares pero en el año 2004 se contabilizaban tan solo entre 16500-47500 ejemplares. En la actualidad, su superviviencia sólo es viable en parques naturales y reservas, [Bauer et al., 2016].

Se desconocen las causas del declive poblacional, que ha reducido la población entre el 30 % y el 50 % en un plazo de 20 años. Por tanto, no se sabe si estas condiciones son o no reversibles. No obstante, la pérdida de hábitat con respecto a los humanos se considera una amenaza grave para la especie.

En cuanto a sus hábitos y costumbres, un león en libertad vive de media 8 años si es macho y entre 12 y 14 si es hembra. En cautiverio pueden superar los 20 años. En general, pasan la mayor parte del tiempo descansando, permanecen inactivos unas 20 horas al día, 2 horas al día caminando y 50 minutos comiendo. Estos hábitos también pueden variar en función de si el león es **residente**, es decir, vive en manada y tiende a permanecer en los límites de su territorio, o si es **nómada**, es decir vive en solitario y desplazándose continuamente.

Resulta curioso que a diferencia de otros felinos, los leones no se aparean en ninguna época particular del año.

Por tanto, disponemos de observaciones de los movimientos de un mamífero de gran tamaño, que habitualmente vive en manada, tiene una esperanza de vida de aproximadamente una década, no tiene época concreta de apareamiento y además, permanece la mayor parte del tiempo inactivo.

El medio en el que se han tomado los datos es la Reserva Natural del río Tsavo en Kenia, entre las ciudades de Nairobi y Mombasa. Es el parque nacional más grande de Kenia, con una extensión próxima a los 22 mil kilómetros cuadrados y está dividido en dos zonas: Tsavo Este y Tsavo Oeste. La mayor parte del parque consiste en pastizales y sabana.

Además, es un paraje famoso por la gran cantidad de especies que alberga: elefantes, leopardos, búfalos, rinocerontes y unas 500 especies distintas de aves. En la actualidad, en esta región se calcula que existen unos 675 leones, Figura 5.1, aunque cada año unos 20 son capturados o asesinados por furtivos.

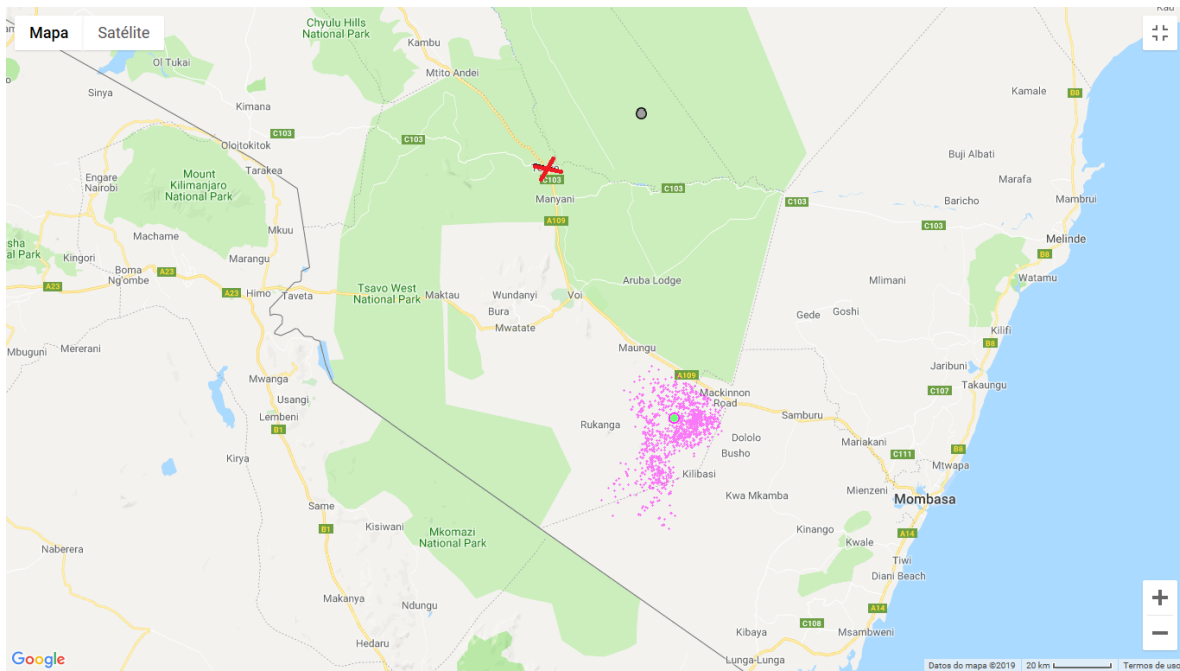


Figura 5.3: Mapa del condado de Taita-Taveta en Kenia, donde pueden observarse en color morado las localizaciones del león en estudio y con una cruz roja señalado el lugar donde se encuentra el puente sobre el río Tsavo por el que pasa el tren que une Kenia y Uganda. // Wikimedia Commons.

En la Figura 5.3, puede verse una fotografía de la región en la que se han monitorizado los movimientos de Kiboche, que está entre las dos zonas de la reserva natural. Desde el siglo XIX ha existido interés por los leones que habitaban en las orillas del río Tsavo, en lo que hoy en día forma parte de la reserva natural de Tsavo Este, Figura 5.2, puesto que en 1898 durante la construcción de un puente ferroviario sobre el río Tsavo para unir Kenia y Uganda, se produjeron numerosos ataques de leones a humanos, que a pesar de ser carnívoros, no presentan antropofagia. Finalmente se dio caza a estos leones e investigaciones posteriores desvelaron que eran más grandes de lo habitual y además, tenían problemas en la mandíbula, lo que les impedía cazar y masticar con normalidad. Este hecho junto con la reducción de presas debido a la dominación humana han hecho que la antropofagia deje de ser un hecho aislado para convertirse en un problema aún en la actualidad en algunas reservas.

Durante los años en los que se ha llevado a cabo el estudio de los movimientos de Kiboche, observamos que el home range se ha desplazado desde el siglo XIX hacia el sureste, hasta quedar fuera de la Reserva Natural Tsavo Este, pues en los años que duró la construcción de la vía de ferrocarril que une Kenia y Uganda, su hábitat estaba a las orillas del río Tsavo, en el lugar marcado con una cruz roja sobre la Figura 5.3. En cambio, las localizaciones registradas en el estudio, indican un notable desplazamiento hasta la región que contiene los puntos coloreados en morado.

Hemos comprobado las características de la zona y es una porción de sabana en la que no hay montañas, ni grandes lagos o ríos. Quizás lo único relevante es la proximidad a la carretera de Mackinon que une las ciudades de Voi y Mombasa, y sobre la cual se encuentra el aeropuerto de Mackinon, un aeropuerto civil sin vuelos regulares y con una única pista de aterrizaje no asfaltada.

Teniendo estas características en cuenta, pasamos a estimar el home range con cada una de las técnicas estudiadas.

## 5.2. Estimación del home range

En este contexto en el que las observaciones han sido tomadas mediante GPS, es esperable que exista correlación temporal, por tanto las técnicas que tienen en cuenta esta característica de los datos presentarán mejores resultados y en cambio, las técnicas para datos independientes lo infraestimarán, aunque si la correlación temporal no es muy fuerte, podrían funcionar bien.

Concretamente las técnicas que vamos a utilizar de entre las que tienen en cuenta la dependencia temporal son PPA home range, KDE autocorrelado (AKDE), modelo de movimiento del puente browniano (BBMM), RMB sausage y cierre  $\alpha$ -convexo de la trayectoria.

No obstante, aplicaremos también técnicas que presuponen independencia para hacer una comparación con las anteriores, concretamente las técnicas de single-linkage cluster, el cierre  $\alpha$ -convexo y la estimación no paramétrica de la función de utilización con selector plug-in del parámetro de suavizado.

De nuevo utilizaremos funciones ya programadas en R que explicaremos detalladamente más adelante. Como comentábamos al principio, los datos de los que disponemos son latitud y longitud medidas en grados, pero algunas de las técnicas que utilizaremos no admiten datos en estas unidades, por tanto en esos casos, haremos una proyección que recibe el nombre de **proyección equidistante entre dos puntos**. Es una proyección planar modificada, con una distorsión mínima en porciones no muy grandes, que tiene la particularidad de que mantiene la distancia entre los puntos, es decir, la distancia entre cualesquiera dos puntos en el mapa a otro, se mantiene antes y después de la proyección. Ha sido además la proyección escogida por ser la más utilizada en estudios sobre los movimientos de los animales. La función utilizada para hacer la proyección es `as.telemetry` del paquete `ctmm`, [Fleming and Calabrese, 2019].

A continuación, expondremos brevemente cada técnica utilizada y su forma de aplicación. Comenzamos por las técnicas que suponen dependencia temporal en los datos.

### ■ PPA (Potential Path Area) dinámico

Recordamos que es un caso particular de la estimación de la densidad geográfica temporal (TG-DE), en la que el núcleo es uniforme y la velocidad máxima se estima en cada intervalo de tiempo.

La función utilizada para la estimación del home range con esta técnica se puede encontrar en el paquete `wildlifeTG`, [Long, 2018]. Para el cálculo de la velocidad máxima en cada intervalo se ha tomado la velocidad media en dicho intervalo de tiempo. En la Figura 5.4, observamos que a pesar de que la velocidad no es constante en toda la estimación, la sobrestimación sigue siendo notable, probablemente debido a que el intervalo de tiempo entre dos observaciones es bastante grande, hasta doce horas en algunas ocasiones, lo que se traduce en incertidumbre y consecuentemente poca precisión en la estimación.

### ■ Estimador tipo núcleo autocorrelado (AKDE)

Como hemos comentado al inicio de la sección, nuestros datos presentan correlación temporal, por tanto esperamos que la técnica del AKDE ajuste mejor el home range que el estimador tipo núcleo usual.

Para este caso, hemos empleado la función `akde` del paquete `ctmm`.

En la Figura 5.5 vemos una estimación del home range de Kiboche utilizando un estimador tipo núcleo autocorrelado. El núcleo utilizado para la estimación ha sido un núcleo gaussiano, y las diferentes franjas que se observan en la figura representan incertidumbre en la estimación del área.

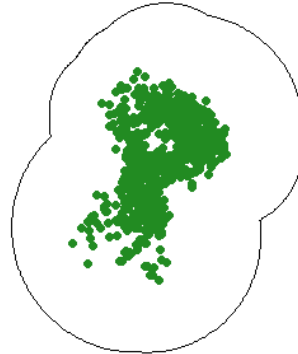
**Estimated home range using dynamic PPA**

Figura 5.4: Representación gráfica de la estimación del home range de Kiboche con la técnica de PPA dinámico. Realizada con RStudio.

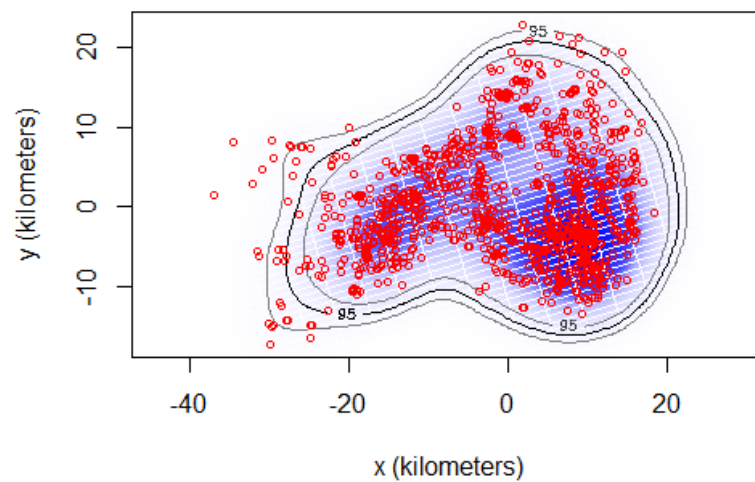


Figura 5.5: Representación gráfica de la estimación del home range de Kiboche con la técnica de AKDE. Realizada con RStudio.

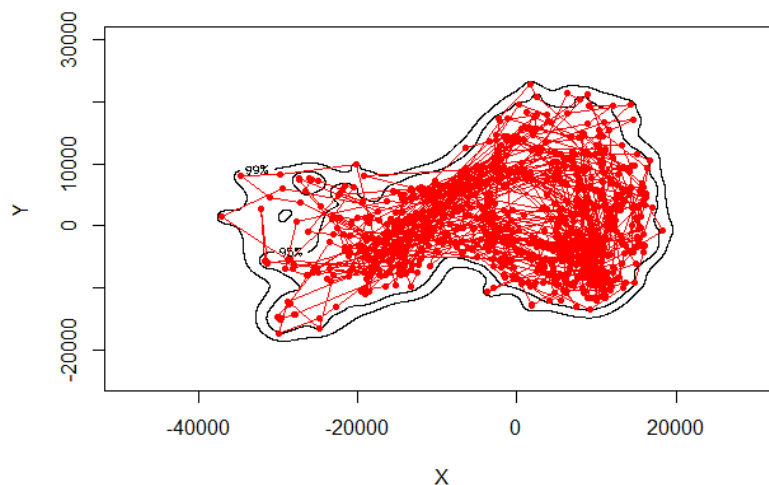


Figura 5.6: Representación gráfica de la estimación del home range de Kiboche con la técnica de BBMM. Realizada con RStudio.

#### ■ Modelo de movimiento del puente browniano (BBMM)

Esta técnica trata las observaciones como un proceso estocástico, concretamente como un puente browniano y estima la función de densidad para posteriormente calcular el home range como el conjunto de nivel que contiene a esa función de densidad con un determinado nivel de confianza.

En este caso, la técnica requiere que las localizaciones sean introducidas en metros, por lo que se ha realizado la proyección de las mismas. Las funciones utilizadas han sido `brownian.bridge` y `bbmm.contour` del paquete `BBMM`, [Nielson et al., 2013].

El resultado obtenido puede observarse en la Figura 5.6. Vemos que se han representado los conjuntos de nivel al 95 % y al 99 %. Este último, engloba a todas las observaciones disponibles y lo hace de una forma muy ajustada.

#### ■ Modelo de movimiento browniano reflejado (MRBM)

En este apartado incluiremos las dos estimaciones estudiadas: **RBM-sausage** y **cierre  $\alpha$ -convexo de la trayectoria**. Ambas programadas en el paquete `alphahull`, [Pateiro-López and Rodríguez-Casal, 2010], en las funciones `dw_track` y `ahull_track`.

En ambos casos, los datos de entrada no necesitan ser proyectados.

En la Figura 5.7 observamos la estimación realizada con la técnica de `RBM-sausage` y en la Figura 5.8 la estimación utilizando la técnica del cierre  $\alpha$ -convexo de la trayectoria, para dos valores distintos de  $\alpha$ .

El valor de los parámetros escogidos, en el primer caso ha sido  $\varepsilon = 0.025$  y en el segundo  $\alpha = 0.02$  y  $\alpha = 0.08$ . La elección se ha realizado empíricamente, probando una secuencia de valores y quedándonos con los que presentaban menor sobrestimación y al mismo tiempo, se ajustaban bien a la trayectoria sin producir un home range muy fragmentado. En el caso de la Figura 5.8a, observamos que el método ha sido capaz de identificar una pequeña porción en el interior del home range en la que no se ha registrado ninguna observación.

Por tanto, se pone de nuevo en evidencia la importancia de seleccionar adecuadamente el parámetro  $\varepsilon$  y  $\alpha$  respectivamente, pues controlan la precisión con la que se hace la estimación.



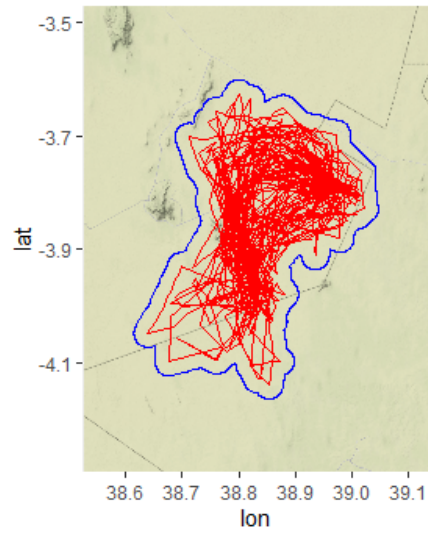


Figura 5.7: Representación gráfica de la estimación del home range de Kiboche con la técnica de RBM-sausage, tomando el valor de  $\varepsilon = 0.025$ . Realizada con RStudio.

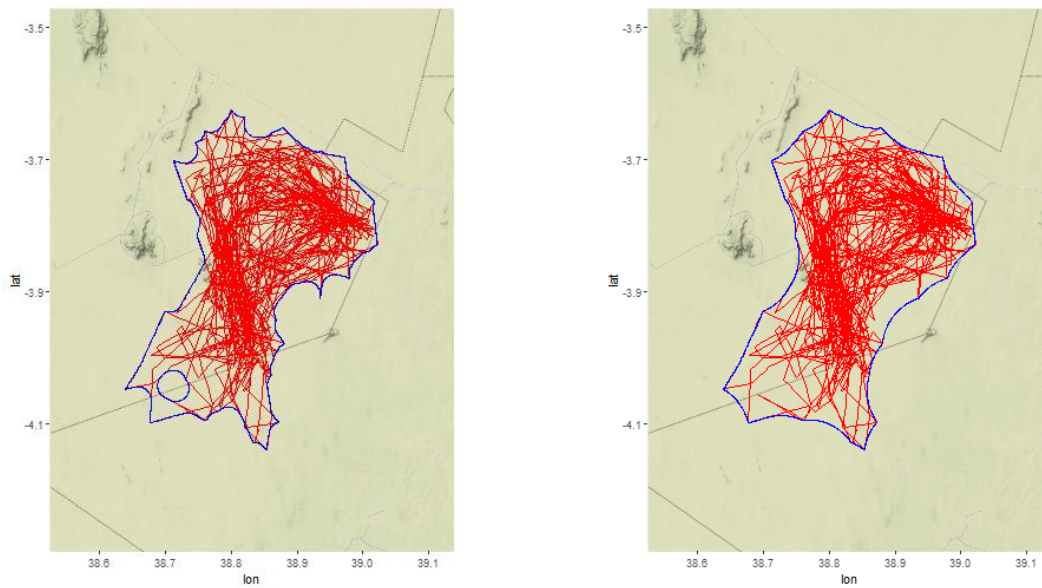


Figura 5.8: De izquierda a derecha, representación gráfica de la estimación del home range de Kiboche con la técnica del cierre  $\alpha$ -convexo de la trayectoria, tomando los valores de  $\alpha = 0.02$  y  $\alpha = 0.08$ , respectivamente. Realizada con RStudio.

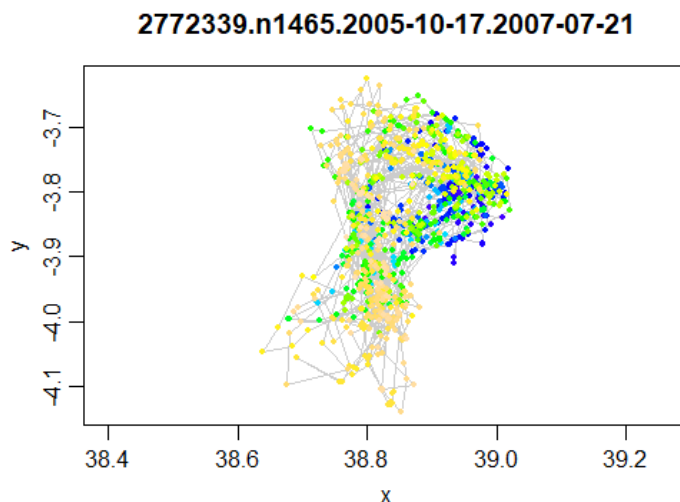


Figura 5.9: Representación gráfica de los movimientos de Kiboche. Aparecen en el mismo color las observaciones que están próximas en el tiempo. Realizada con RStudio.

Estudemos ahora el funcionamiento de la única técnica localizada expuesta para lidiar con la dependencia temporal en los datos: T-LoCoH.

- **Time-Local Convex Hull (T-LoCoH):**

Esta técnica es similar al LoCoH, de hecho también tiene las mismas tres variantes:  $k$ -vecino más próximo, radio fijo y radio adaptativo. Pero es capaz de tratar con datos dependientes porque incorpora la coordenada temporal a la hora de crear los grupos de observaciones, en lugar de hacerlo basándose únicamente en la distancia euclídea como ocurría con el LoCoH usual.

La dificultad ahora está en decidir cuánta importancia le damos a la coordenada temporal, es decir, en elegir el valor del parámetro  $s$ . Afortunadamente, la librería `tlocoh`, [Lyons and Getz, 2019], además de incorporar una función para realizar la estimación con esta técnica, también incorpora una función para elegir el valor del parámetro  $s$ , mediante técnicas de validación cruzada, que mejor se ajuste a nuestros datos. En este caso el valor escogido de  $s$  ha sido  $s = 0.13$ .

En la Figura 5.9, puede verse una representación de las localizaciones de Kiboche por colores, donde cada color es indicativo de los instantes temporales en los que han sido tomadas, es decir, cada color reúne a observaciones que se han tomado en instantes de tiempo próximos.

Finalmente la técnica escogida para realizar la estimación ha sido la del  $k$  vecino más próximo, tomando un valor de  $k = 25$ , es decir, los grupos están formados por 25 observaciones próximas entre sí y en distancia euclídea. No ha sido un valor elegido al azar, sino que de nuevo la librería incorpora la opción de elegir el parámetro  $k$ ,  $r$  o  $a$  en función del valor que toma  $s$ . En la Figura 5.10, podemos observar una representación gráfica de la estimación realizada. Vemos que los distintos colores representan el porcentaje de puntos que engloba la estimación.

Por último, incluimos tres técnicas globales para estimación del home range con datos independientes, para observar su funcionamiento a la hora de hacer estimaciones con datos que presentan una fuerte correlación temporal.

- **Single-linkage cluster**

En los estudios de simulación del Capítulo 4 observamos que el single-linkage clúster era capaz de representar con bastante exactitud la geometría del soporte, infraestimando el área original pero con la ventaja de que no requiere de la elección de ningún parámetro.



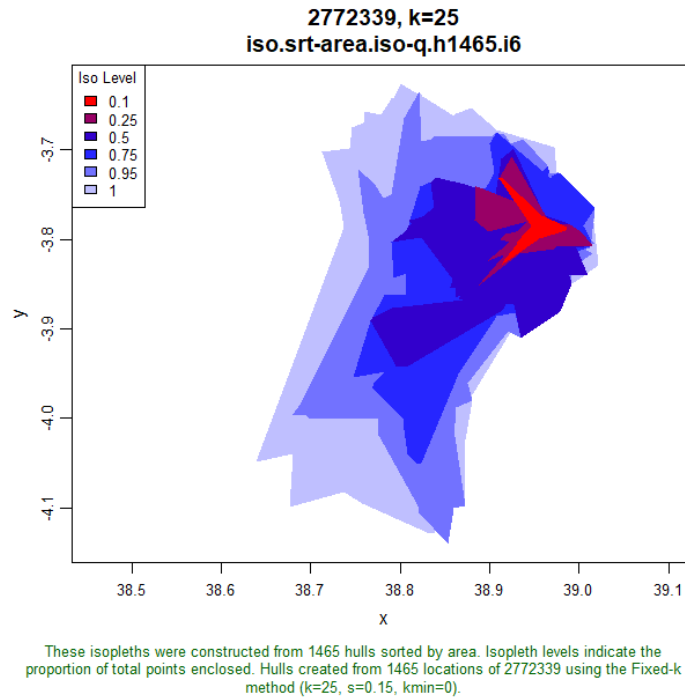


Figura 5.10: Representación gráfica del home range de Kiboche con la técnica de TLoCoH con la variante del  $k$ -vecino más próximo, tomando un valor de  $k = 25$  y  $s = 0.13$ . Realizada con RStudio.

Hemos realizado por tanto la estimación con esta técnica para comprobar su funcionamiento sobre un conjunto de datos real. La función empleada ha sido la misma que en los casos anteriores, `clusthr` del paquete `adehabitatHR`. En la Figura 5.11 podemos ver el resultado. Al contrario de lo que ocurría con las simulaciones de datos independientes, ahora el estimador no capta con claridad la geometría del conjunto a estimar.

#### ■ Cierre $\alpha$ -convexo

Para estimar el home range con la técnica del cierre  $\alpha$ -convexo, hemos utilizado de nuevo la función `ahull` del paquete `alphahull`. El valor del parámetro escogido para  $\alpha$  ha sido  $\alpha = 0.05$ . En este caso la elección también ha sido empírica, escogiendo de entre una secuencia de valores comprendidos entre 0.01 y 1, el que presentaba una estimación más fiel a la trayectoria.

En la Figura 5.12 podemos ver la representación gráfica de la estimación, que como era de esperar, ha unido los puntos que están en la frontera del home range con arcos de circunferencia, reduciendo así la sobrestimación que produce la técnica del cierre convexo (MCP).

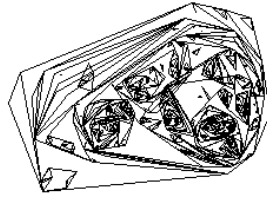


Figura 5.11: Representación gráfica de la estimación del home range de Kiboche con la técnica de estimación global para datos independientes de single-linkage cluster. Realizada con RStudio.

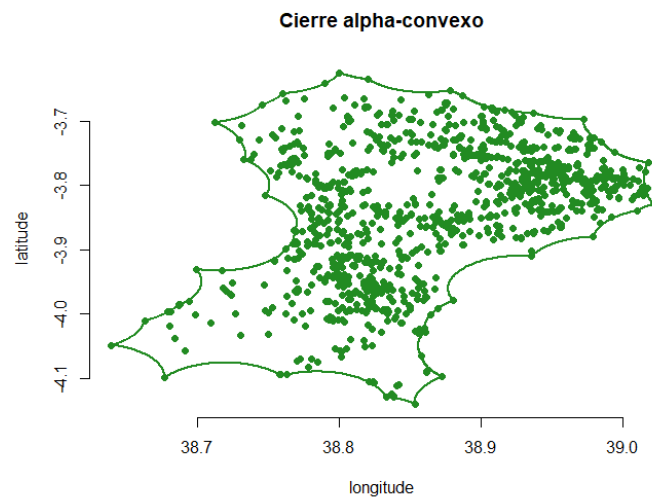


Figura 5.12: Representación gráfica de la estimación del home range de Kiboche con la técnica de estimación global para datos independientes de cierre  $\alpha$ -convexo, para un valor de  $\alpha = 0.05$ . Realizada con RStudio.

### Estimador UD (selector plug-in)



Figura 5.13: Representación gráfica de la estimación del home range de Kiboche con la técnica de estimación global para datos independientes que estima la función de utilización con selector plug-in. Realizada con RStudio.

#### ■ Función de utilización con selector plug-in

La función utilizada para la estimación es la misma que ya habíamos utilizado con anterioridad y que forma parte del paquete `adehabitathR`.

Así, en la Figura 5.13 observamos la estimación hecha con este método, observando que produce una notable sobrestimación, sobre todo si la comparamos con la estimación realizada con la técnica del AKDE, en parte porque no tiene en cuenta la dependencia temporal presente en los datos.

No obstante, presenta resultados similares al PPA dinámico. Teniendo en cuenta que estas dos técnicas requieren de la elección de parámetros de suavizado y que de entre todas las técnicas de estudiadas en esta sección, han sido las que han presentado una mayor sobrestimación, se pone de manifiesto que a menudo es tan importante la elección de la técnica como la de los parámetros de suavizado.

En términos generales, observamos que las técnicas de PPA dinámico y estimador UD con selector plug-in son las que más han sobrestimado el home range. En el primer caso porque el intervalo de tiempo entre cada par de observaciones es muy grande y en el segundo probablemente por una elección de un  $h$  demasiado grande.

Por otro lado el AKDE ha producido una estimación muy ajustada, pero con bastante incertidumbre y que además no engloba a todas las observaciones.

En cambio, la técnica local del TLoCoH y las técnicas globales basadas en procesos estocásticos han presentado resultados muy ajustados y por tanto satisfactorios, si bien hay que tener en cuenta que los casos del RBM-sausage y el cierre  $\alpha$ -convexo de la trayectoria requieren una elección adecuada del  $\varepsilon$  y  $\alpha$  que a menudo puede no ser trivial.



## Capítulo 6

# Conclusiones

A lo largo del presente documento hemos observado que la estimación del área de actividad de un animal no es un problema con una única solución sino que puede afrontarse desde distintas perspectivas y entre ellas, no existe una técnica vencedora sino que su buen funcionamiento depende de muchos factores, entre ellos la naturaleza de los datos (si presentan o no dependencia temporal), el tamaño muestral, el reparto uniforme de las observaciones en el espacio o la existencia de zonas no visitadas en el home range.

Desde luego, para datos independientes el cierre  $\alpha$ -convexo y los métodos locales de single-linkage cluster y LoCoH se han demostrado efectivos, a pesar de que el primero y tercero requieran de la elección de parámetros  $\alpha$ ,  $k$ ,  $r$ ,  $a$  respectivamente. Sin duda una línea de investigación futura sería el diseño e implementación de una técnica capaz de elegir los valores óptimos de estos parámetros.

En cuanto al single-linkage, podría resultar útil evaluar la calidad del estimador con otros tipos de linkage.

Por último, en las técnicas para datos dependientes, las que tratan a las observaciones como trayectorias de procesos estocásticos han presentado mejores resultados que las modificaciones del KDE que incorporan la coordenada temporal.

Pero sin duda este estudio no está ni mucho menos cerrado, ya que en la actualidad continúan desarrollándose nuevas técnicas como pueden ser el TDA (Topological Data Analysis) o la estimación de la función de densidad geográfico-temporal, que pueden mejorar la estimación del área de actividad de un animal al mismo tiempo que se presentan como prometedoras áreas de investigación que serán capaces de lidiar la gran cantidad de datos que acumulan los móviles y dispositivos de geolocalización sobre los movimientos de los humanos, con aplicaciones tan diversas como la predicción de la evolución de las epidemias, evolución del tráfico de una ciudad...[Meekan et al., 2017].

Así, el estudio de la estimación del área de actividad es un campo con un prometedor futuro, tanto por sus aplicaciones en la vida real como por todo el desarrollo matemático que vendrá parejo.



# Bibliografía

- [Baíllo and Chacón, 2018] Baíllo, A. and Chacón, J. (2018). A survey and a new selection criterion for statistical home range estimation.
- [Barber et al., 1996] Barber, B., Dobkin, D., and Huhdanpaa, H. (1996). The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software*.
- [Bath et al., 2006] Bath, S. K., Hayter, A. J., Cairns, D., and Anderson, C. (2006). Characterization of home range using point peeling algorithms. *Journal of Wildlife Management*.
- [Bauer et al., 2016] Bauer, H., Packer, C., Funston, P., Henschel, P., and Nowell, K. (2016). Panthera leo. *The IUCN Red List of Threatened Species*.
- [Benhamou and Cornelis, 2010] Benhamou, S. and Cornelis, D. (2010). Incorporating movement behavior and barriers to improve biological relevance of kernel home range space use estimates. *Journal of Wildlife Management*.
- [Billingsley, 1995] Billingsley, P. (1995). *Probability and Measure*. Wiley.
- [Bivand and Rundel, 2019] Bivand, R. and Rundel, C. (2019). *rgeos: Interface to Geometry Engine - Open Source ('GEOS')*. R package version 0.4-3.
- [Bowman, 1984] Bowman, A. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, pages 353–360.
- [Bowman and Azzalini, 1999] Bowman, A. and Azzalini, A. (1999). *Smoothing Methods for Nonparametric Regression and Density Estimation*.
- [Bullard, 1999] Bullard, F. (1999). *Estimating the home range of an animal: a Brownian bridge approach*. PhD thesis, University of North Carolina.
- [Burt, 1943] Burt, H. (1943). Territoriality and home range concepts as applied to mammals. *Journal of Mammalogy*.
- [Calenge, 2006] Calenge, C. (2006). The package adehabitat for the r software: tool for the analysis of space and habitat use by animals. *Ecological Modelling*.
- [Calhoun and Casby, 1958] Calhoun, J. and Casby, J. (1958). Calculation of home range and density of small mammals. *Public Health Monograph*.
- [Cholaquidis et al., 2016] Cholaquidis, A., Fraiman, R., Lugosi, G., and Pateiro-López, B. (2016). Set estimation from reflected brownian motion. *Journal of the Royal Statistical Society B*.
- [Ciesielski, 2012] Ciesielski, L. (2012).
- [core Team, 2017] core Team, R. (2017). A language and environment for statistical computing.
- [Darwin, 1861] Darwin, C. (1861). *On the origin of species*. Murray, London.

- [Delaunay, 1934] Delaunay, B. (1934). Sur la sphere vide. *Otdelenie Matematicheskii i Estestvennyka Nauk*, 7:793–800.
- [Don and Rennolls, 1983] Don, B. and Rennolls, K. (1983). A home range model incorporating biological attraction points. *Journal of Animal Ecology*.
- [Downs and Horner, 2009] Downs, J. and Horner, M. (2009). A characteristic-hull based method for home range estimation. *Transactions in GIS*, 13.
- [Downs et al., 2011] Downs, J., Horner, M., and Tucker, A. (2011). Time-geographic density estimation for home range analysis. *Annals of GIS*, 17.
- [Eddy, 1977] Eddy, W. F. (1977). A new convex hull algorithm for planar sets. *ACM Transactions on Mathematical Software*.
- [Edelsbrunner, 2010] Edelsbrunner, H. (2010). Alpha shapes. a survey. *Tessellations in the Sciences*, 27:1–25.
- [Epanechnikov, 1969] Epanechnikov, V. (1969). Non parametric estimation of a multivariate probability density. *Theory of probability and its applications*.
- [Fleming et al., 2015] Fleming, C., Fagar, W., Muller, T., Olson, K., Leimgruber, P., and Calabrese, J. (2015). Rigorous home range estimation with movement data: a new autocorrelated kernel density estimator. *Ecology*.
- [Fleming and Calabrese, 2019] Fleming, C. H. and Calabrese, J. M. (2019). *ctmm: Continuous-Time Movement Modeling*. R package version 0.5.6.
- [Getz et al., 2007] Getz, W., Fortman-Roe, S., Cross, P., Lyons, A., Ryan, S., and Wilmers, C. (2007). Locoh: Nonparametric kernel methods for constructing home ranges and utilization distributions. *PLOS ONE*, 2.
- [Getz and Wilmers, 2004] Getz, W. and Wilmers, C. (2004). A local nearest-neighbor convex-hull construction of home ranges and utilization distributions. *Ecography*, 4.
- [Harvey and Barbour, 1965] Harvey, M. J. and Barbour, R. W. (1965). Home range of microtus ochrogaster as determined by a modified minimum area method. *Journal of Mammalogy*.
- [Hägerstrand, 1970] Hägerstrand, T. (1970). What about people in regional science? *Papers of the Regional Science Association*.
- [Horne et al., 2007] Horne, J., Garton, E., Krone, S., and Lewis, J. (2007). Analyzing animal movements using brownian bridges. *Ecology*, 9.
- [Karatzas and Shreve, 1991] Karatzas, I. and Shreve, S. (1991). *Brownian motion and Stochastic Calculus*. Springer.
- [Keating and Cherry, 2009] Keating, K. and Cherry, S. (2009). Modeling utilization distributions in space and time. *Ecology*.
- [Kenward et al., 2001] Kenward, R. E., Clarke, R. T., Hodder, K. H., and Walls, S. S. (2001). Density and linkage estimators of home range: nearest-neighbor clustering defines multinuclear cores. *Ecology*.
- [Long, 2018] Long, J. (2018). *wildlifeTG: Time Geographic Analysis of Wildlife Telemetry Data*. R package version 0.1.
- [Long and Nelson, 2015] Long, J. and Nelson, T. (2015). Home range and habitat analysis using dynamic time geography. *The Journal of Wildlife Management*, 79:481–490.



- [Lyons and Getz, 2019] Lyons, A. and Getz, W. (2019). *T-LoCoH: Time Local Convex Hull Homorange and Time Use Analysis*. R package version 1.40.07.
- [Lyons et al., 2013] Lyons, A., Turner, W., and Getz, W. (2013). Home range plus: a space-time characterization of movement over real landscapes. *Movement ecology*, 2.
- [Matthiopoulos, 2003] Matthiopoulos, J. (2003). Model supervised kernel smoothing for the estimation of spacial usage. *Oikos*.
- [Meekan et al., 2017] Meekan, M., Duarte, C., Fernández-García, J., Thums, M., Sequeira, A., Harcourt, R., and Eguíluz, V. (2017). The ecology of human mobility. *Trends in ecology and evolution*.
- [Nielson et al., 2013] Nielson, M., Sawyer, H., and McDonald, T. L. (2013). *BBMM: Brownian bridge movement model*. R package version 3.0.
- [Pateiro-López and Rodríguez-Casal, 2010] Pateiro-López, B. and Rodríguez-Casal, A. (2010). Generalizing the convex hull of a sample: the r package alphahull.
- [Rhee, 2013] Rhee, R. (2013).
- [Rodríguez-Casal, 2007] Rodríguez-Casal, A. (2007). Set estimation underconvexity type assumptions. *Annales de l Institut Henri Poincaré Probabilités et Statistiques*, 43:763–774.
- [Rodríguez-Casal and Saavedra-Nieves, 2014] Rodríguez-Casal, A. and Saavedra-Nieves, P. (2014). A fully data-driven method for estimating the shape of a point cloud. *ESAIM: Probability and Statistics*, 20.
- [Rosenblatt, 1956] Rosenblatt, M. (1956). Remarks on some non-parametric estimates of a density function. *The Annals Mathematical Statistics*, 3.
- [Sheather and Jones, 1991] Sheather, S. J. and Jones, M. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society Series B*, 53:683–690.
- [Silverman, 1986] Silverman, B. (1986). *Density estimation for statistics and data analysis*. Chapman and Hall, London.
- [Steiniger and Hunter, 2013] Steiniger, S. and Hunter, A. (2013). A scaled line-based density estimator for the retrieval of utilization distributions and home ranges from gps movement tracks. *Ecological Informatics*.
- [Su and Drysdale, 1997] Su, P. and Drysdale, R. (1997). A comparison of sequential delaunay triangulation algorithms. *Computational Geometry*, 7.
- [Tarjan and Tinker, 2016] Tarjan, L. and Tinker, M. (2016). Permissible home range estimation (phre) in restricted habitats: A new algorithm and an evaluation for sea otters. *PLOS ONE*, 11.
- [Taylor, 1989] Taylor, C. (1989). Bootstrap choice of the smoothing parameter in kernel density estimation. *Biometrika*.
- [van Winkle, 1975] van Winkle, W. (1975). Comparison of several probabilistic home-range models. *Ecological Modelling*, 38:277–298.
- [Wasserman, 2018] Wasserman, L. (2018). Topological data analysis. *Annual Review of Statistics and Its Application*, 5.
- [Worton, 1987] Worton, B. J. (1987). A review of models of home range for animal movement. *Ecological Modelling*.

[Worton, 1989] Worton, B. J. (1989). Kernel methods for estimating the utilization distribution in home range studies. *Ecology*.

[Worton, 1995] Worton, B. J. (1995). A convex hull-based estimator of home-range size. *Biometrics*.