



FACULTADE DE MATEMÁTICAS

Traballo Fin de Grao

# Estimación del soporte para reconstruir el *home range*

Sarela Folgar Varela

2019/2020

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



GRAO DE MATEMÁTICAS

Traballo Fin de Grao

Estimación del soporte para  
reconstruir el *home range*

Sarela Folgar Varela

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



# Trabajo propuesto

**Área de Coñecemento:** Estatística e Investigación Operativa.

**Título:** Estimación do soporte para reconstruír o *home range*

## Breve descrición do contido

Un problema que é de interese para biólogos e ecólogos é a reconstrución do *home range* a partir das localizacións xeográficas de especies animais. O *home range* defínese como a área ocupada por un individuo para realizar as súas actividades cotiás. Empregaremos a teoría da estimación do soporte para proporcionar diferentes estimadores do *home range*. Este enfoque non ten en conta a dependencia temporal das observacións. Aínda que non serán consideradas neste traballo, existen na literatura reconstrucións alternativas que solventan este problema. Como ilustración, empregaremos dous conxuntos de datos reais, un de lobos en Mongolia e outro de cabras salvaxes na illa de Hawaii para mostrar os distintos estimadores do *home range*. Finalmente, discutirase que criterios se empregan habitualmente para seleccionar a reconstrución do *home range* máis realista.



# Índice general

<b>Resumen</b>	<b>VII</b>
<b>1. Preliminares</b>	<b>1</b>
1.1. Datos reales . . . . .	4
1.1.1. Localizaciones geográficas de dos lobos en Mongolia. . . . .	4
1.1.2. Localizaciones geográficas de cabras salvajes en Hawaii . . . . .	6
1.2. Herramientas matemáticas para la estimación del <i>home range</i> . . . . .	8
1.2.1. Distancias entre conjuntos . . . . .	9
1.2.2. Propiedades geométricas . . . . .	10
<b>2. Métodos de estimación del soporte para reconstruir el <i>home range</i></b>	<b>15</b>
2.1. Sin restricciones de forma . . . . .	15
2.2. Con restricciones de forma . . . . .	21
<b>3. Ranking de <i>home ranges</i></b>	<b>29</b>
3.1. Ranking de <i>home ranges</i> para las ubicaciones del lobo en Mongolia. . . . .	30
3.2. Ranking de <i>home ranges</i> para las ubicaciones de las cabras salvajes en Hawaii. . . . .	33
<b>4. Conclusiones y discusión</b>	<b>39</b>
<b>Bibliografía</b>	<b>43</b>



## Resumen

En este trabajo revisaremos los principales estimadores propuestos en la literatura para la estimación del soporte, y como aplicación práctica trataremos de reconstruir el *home range* de uno o varios animales a partir de una muestra de sus localizaciones geográficas. El *home range* se define como la porción de área atravesada por un individuo para realizar sus actividades cotidianas. Para esto, haremos uso de dos conjuntos de datos que contienen ubicaciones geográficas de lobos y cabras salvajes.

La reconstrucción no paramétrica del soporte puede considerarse desde dos enfoques distintos: asumiendo que el soporte no satisface ninguna restricción de forma o suponiendo que verifica alguna propiedad geométrica concreta. En el segundo caso, es posible asumir que el soporte es convexo o, de forma más flexible,  $r$ -convexo para algún  $r > 0$ .

Ninguno de los estimadores del soporte tienen en cuenta la dependencia temporal existente entre las observaciones. Por tanto, una cuestión importante es cómo incorporar la dependencia temporal en estos estimadores, ya que podemos disponer de datos de ubicaciones demasiado cercanos en el tiempo. Aunque no serán consideradas en este trabajo, citaremos brevemente algunas de las propuestas disponibles en la literatura para solventar este problema.

Por último, realizaremos un ranking con los *home ranges* obtenidos comparando las áreas de los estimadores obtenidos a partir de la muestra completa y a partir la primera mitad cronológica de la misma.

## Abstract

In this work we will review the main estimators proposed in the literature for estimating the support, and as a practical application we will try to reconstruct the home range of one or more animals from a sample of their geographic locations. The home range is defined as the portion of the area crossed by an individual to carry out their daily activities. For

this, we will make use of two datasets that contain geographic locations of wolves and wild goats.

The non-parametric reconstruction of the support can be considered from two different approaches: assuming that the support does not satisfy any shape constraint or assuming that it verifies some specific geometric property. In the second case, it is possible to assume that the support is convex or, more flexibly,  $r$ -convex for some  $r > 0$ .

None of the support estimators take into account the temporal dependence between the observations. Therefore, an important question is how to incorporate the temporal dependence in these estimators, since we can have data from locations that are too close in time. Although they will not be considered in this work, we will briefly cite some of the proposals available in the literature to solve this problem.

Finally, we will perform a ranking with the home ranges obtained by comparing the areas of the estimators obtained from the complete sample and those obtained from the first chronological half of it.

# Capítulo 1

## Preliminares

Durante mucho tiempo, ha existido interés en identificar distintas características del uso del espacio de especies animales o individuos. Biólogos conservacionistas y ecologistas espaciales, tienen interés en estimar mapas de uso del espacio a partir de datos de seguimiento de animales. Por ejemplo, para monitorear especies amenazadas y la planificación de la conservación del medio. De la misma forma, el uso de smartphones con geolocalizador permite disponer de cantidades masivas de datos de seguimiento. El análisis del uso del espacio es, por lo tanto, más atractivo que nunca, no solo desde un punto de vista conservador (análisis de interacciones entre animales y humanos) sino también desde una perspectiva económica e incluso antropológica.

A lo largo del tiempo, la forma en la que se recogen las localizaciones de los animales ha cambiado radicalmente. Originalmente, se obtenían usando trampas o haciendo observaciones continuas al animal. En los años 60, estas técnicas fueron sustituidas por radio telemetría, que consiste en colocar un transmisor al animal y registrar su posición por transmisiones de señales de radio. Con los avances tecnológicos actuales, podemos obtener datos de localizaciones de animales con mayor frecuencia y precisión, por ejemplo, haciendo uso de *Global Positioning System* (GPS). Esto, junto con la mejora de los recursos informáticos, ha originado el uso de nuevas técnicas de estimación y un creciente interés en el análisis del uso del espacio basado en individuos rastreados.

*Baño y Chacón (2018)* han realizado una revisión exhaustiva sobre este tema presentando algunas reflexiones de autores sobre el concepto del *home range*. Como *Seton (1909)* señala, “Ningún animal salvaje vagabundea al azar por el campo, cada uno tiene un *home range* incluso si no es el actual hogar. El tamaño de esta región se corresponde de alguna forma con el tamaño del animal”. *Burt (1943)* se acredita con la primera formalización de la idea del *home range* como “ese área atravesada por el individuo en sus actividades normales de recolección de alimentos, apareamiento y cuidado de los jóvenes. Salidas oca-

sionales fuera del área, tal vez de naturaleza exploratoria, no deberían considerarse parte del *home range*". Sin embargo, las áreas de descanso y riego y las rutas entre estas ubicaciones estratégicas deben considerarse parte de él. Por tanto, es importante no confundir el *home range* de un animal con su territorio, que es la porción defendida del *home range*. En particular, los *home ranges* de diferentes individuos pueden solaparse (por ejemplo, es posible que compartan lugares de abastecimiento).

De acuerdo con los comentarios previos, la delimitación de la frontera del *home range* es una forma simple y popular de describir el uso del espacio de un individuo monitoreado. Matemáticamente, la reconstrucción de dicho borde puede verse como un problema clásico de estimación no paramétrica del soporte. En lo que sigue,  $S \subset \mathbb{R}^d$  denotará el soporte compacto y no vacío de un vector aleatorio continuo  $X$  con distribución de probabilidad  $P_X$ , verificando además que  $S$  es el conjunto cerrado más pequeño tal que  $P_X(S) = 1$ .

Dada una muestra independiente e idénticamente distribuida de  $X$ ,  $\mathfrak{N}_n = \{X_1, \dots, X_n\}$ , el objetivo de la teoría de estimación del soporte es reconstruir el soporte  $S$  de  $X$ . En este trabajo, emplearemos dos conjuntos de datos reales para ilustrar cómo funcionan los distintos métodos no paramétricos de estimación del soporte para reconstruir el *home range* utilizando el software libre R. Concretamente, una de las muestras se corresponde con las localizaciones de dos lobos en Mongolia registradas entre 2004 y 2005, y la otra, con las posiciones geográficas de 13 cabras en Hawái tomadas entre 2010 y 2011. En la *Figura 1.1* se muestran las coordenadas geográficas (latitud y longitud) de estos dos conjuntos de datos. Las ubicaciones de los dos lobos se muestran en la parte izquierda; las de las cabras, en la parte derecha.

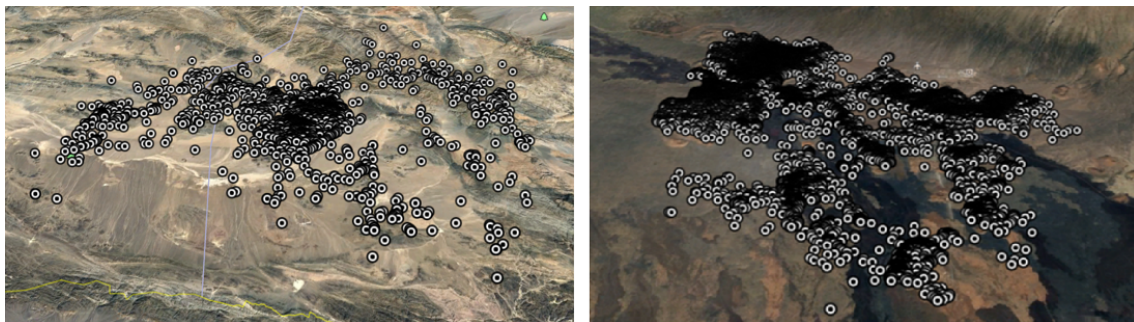


Figura 1.1: Datos de ubicaciones de los dos lobos (izquierda) y de las cabras (derecha) representados en Google Earth.

Una de las principales desventajas de considerar la estimación del soporte para reconstruir el *home range* es la dependencia temporal inherente al tipo de datos que manejamos. *Fleming et al. (2015)* resalta el hecho de que los datos de seguimiento estén, por natu-

raleza, autocorrelados. Esto es una consecuencia obvia de que el movimiento de animales es un proceso continuo. Por tanto, podemos pensar que datos próximos en el tiempo pueden ser dependientes, ya que con las muestras obtenidas a partir de tecnología GPS, podemos tener una gran cantidad de datos en un intervalo de tiempo pequeño. También considera que una muestra autocorrelada contiene menos información geométrica sobre la frontera del conjunto que una muestra de observaciones independientes del mismo tamaño. Sin embargo, la existencia de autocorrelación entre observaciones implica que podemos aprovechar la información del pasado para predecir los futuros movimientos del animal y, por tanto, debería usarse también en la estimación del *home range*. Tal y como comentaremos en las secciones siguientes, existen alternativas en la literatura que incorporan esta información.

En resumen, son dos los objetivos centrales de este trabajo. Primero, proporcionar una revisión de los principales métodos de estimación del soporte para reconstruir el *home range* de uno o varios animales en la literatura sin obviar el problema de la dependencia temporal. Finalmente, dada la gran cantidad de estimadores posibles, es fundamental compararlos para establecer cuál es la reconstrucción óptima. Para ello, discutiremos algunos de los criterios más usados en la literatura donde ésta es una cuestión todavía abierta.

El *Capítulo 1* se organiza como sigue. En la *Sección 1.1*, describiremos en detalle los dos conjuntos de datos reales ya introducidos y con los que vamos a ilustrar la reconstrucción del *home range*. Concretamente, en la *Sección 1.1.1*, presentaremos los datos de los lobos y en la *Sección 1.1.2*, de las cabras de Hawaii. Algunos de los estimadores del soporte que revisaremos en este trabajo asumen que el soporte satisface alguna condición de forma geométrica concreta. En la *Sección 1.2*, definiremos formalmente estas condiciones de forma. Además, varias distancias entre conjuntos serán consideradas. Si se mide la distancia entre el soporte teórico y una estimación del mismo, es posible cuantificar o medir el error de estimación cometido. De esta forma, distintos estimadores del *home range* podrían ser comparados.

En el *Capítulo 2*, revisaremos los métodos de estimación del soporte más empleados en la literatura para reconstruir el *home range*. Particularmente, en la *Sección 2.1*, revisaremos algunos estimadores para el soporte sin imponer sobre este ninguna restricción de forma, mientras que en la *Sección 2.2*, impondremos restricciones de forma sobre el soporte para tratar de obtener estimadores más realistas.

En el *Capítulo 3*, revisaremos las propuestas de algunos autores para seleccionar el estimador del *home range* más realista de entre todos los calculados. Concretamente, en la *Sección 3.1*, realizaremos un ranking sobre los estimadores del *home range* obtenidos para los datos de ubicaciones de los dos lobos y en la *Sección 3.2*, lo haremos para los datos de

localizaciones de las cabras salvajes en la isla de Hawaii.

Finalmente, en el *Capítulo 4*, expondremos las conclusiones obtenidas a través de los resultados anteriormente logrados y discutiremos las cuestiones presentadas a lo largo del trabajo.

## 1.1. Datos reales

En las siguientes secciones presentaremos dos conjuntos de datos de ubicaciones que emplearemos a lo largo del trabajo para analizar el comportamiento de los estimadores del soporte considerados. Específicamente, en la *Sección 1.1.1*, introducimos un conjunto de datos de ubicaciones de dos lobos en Mongolia, y en la *Sección 1.1.2*, las localizaciones de 13 cabras salvajes en la isla de Hawaii.

### 1.1.1. Localizaciones geográficas de dos lobos en Mongolia.

*Baíllo y Chacón (2018)* revisan distintas técnicas de estimación del *home range* usando como ilustración un conjunto de datos de lobos en Mongolia, ver *Figura 1.2*. En este

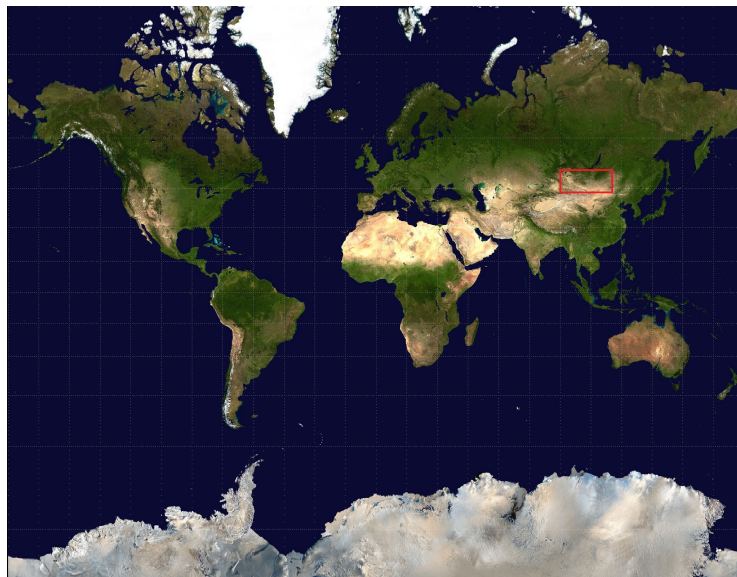


Figura 1.2: Mongolia enmarcada con un rectángulo rojo en un mapamundi.

trabajo, usaremos el mismo conjunto de datos reales. Se corresponden con ubicaciones con identificador 14291019 que pueden descargarse en la web de Movebank<sup>1</sup>. Esta base de datos

---

<sup>1</sup>[www.movebank.org](http://www.movebank.org)

contiene información sobre movimientos de animales y está coordinada por el instituto Max Plank de ornitología, el museo de ciencias naturales de Carolina del Norte y la Universidad de Konstanz. Específicamente, nuestro conjunto de datos contiene las localizaciones de dos lobos de Mongolia: Boroo (hembra) y Zimzik (macho), observados durante un periodo que comienza en mayo de 2003 para Boroo y marzo de 2004 para Zimzik, terminando en septiembre de 2005, ambos recogidos con tecnología GPS, disponiendo de 2226 y 1455 localizaciones, respectivamente. La transmisión de la señal se realizaba de una a tres veces por día, en momentos irregulares. Entre ambos lobos, hemos seleccionado las ubicaciones de Zimzik, ya que su trayectoria parece corresponder a un *home range* con características matemáticas interesantes. En la *Figura 1.3* y *Figura 1.4*, hemos representado las posiciones de los dos lobos Boroo y Zimzik. El mapa comprende parte de las provincias de Hovd y

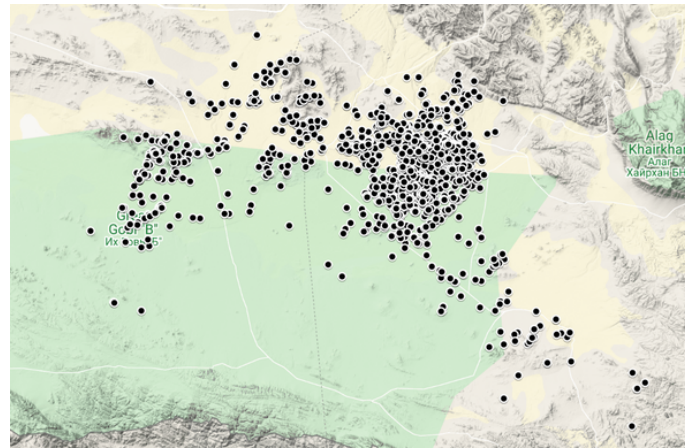


Figura 1.3: Ubicaciones de Boroo en Google Maps.

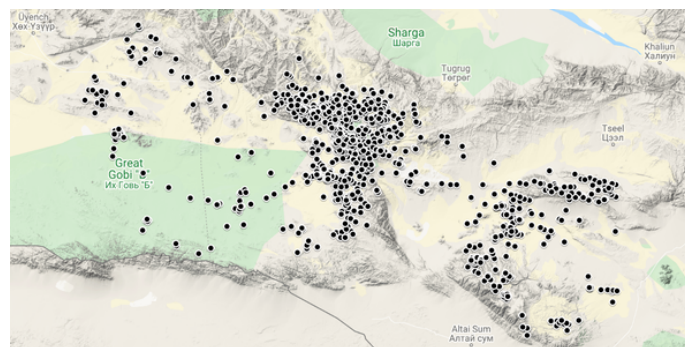


Figura 1.4: Ubicaciones de Zimzik en Google Maps.

Govi-Altay en el suroeste de Mongolia y China (la frontera está indicada con una línea

sólida gris). También podemos ver partes de la zona estrictamente protegida del Gran Gobi B, una Reserva Internacional de la Biosfera en el desierto de Gobi y otras reservas naturales. A partir de la *Figura 1.4* se observa que Zimzik prefería los terrenos montañosos (pendiente  $>5^\circ$ ), ya que dificultan la caza de lobos, sobre la estepa. En consecuencia, en este caso, los contornos de elevación contribuyen a la forma del *home range*. En general, parece razonable creer que hay variables explicativas relacionadas (hasta cierto punto) con la distribución de utilización y el *home range* de un animal. El problema de cómo incorporar esta información en el estimador de *home range* ha sido poco analizado todavía en la literatura.

### 1.1.2. Localizaciones geográficas de cabras salvajes en Hawaii

El comportamiento de los métodos considerados en este trabajo para reconstruir el *home range*, será ilustrado empleando otro conjunto de datos disponible también en la web de Movebank<sup>2</sup>. Este conjunto de datos contiene las ubicaciones de cabras salvajes tomadas entre julio de 2010 y julio de 2011 en Pōhakuloa Training Area (PTA), en la isla de Hawaii, ver *Figura 1.5*. PTA es una instalación militar activa de  $438 \text{ km}^2$  situada entre



Figura 1.5: De izquierda a derecha, el archipiélago de Hawaii enmarcado con un rectángulo rojo en un mapamundi, la isla a la que pertenecen las ubicaciones de las cabras salvajes (Hawaii) y Pōhakuloa Training Area (PTA).

tres volcanes, Mauna Kea (4205 m), Mauna Loa (4169 m) y Hualālai (2521 m) caracterizada por una elevada variabilidad climática, con temperaturas que oscilan entre 10 y 22 grados centígrados durante al menos 4 meses al año. El setenta por ciento de las precipitaciones anuales ( $561 \text{ l/m}^2$  de promedio anual) ocurren típicamente entre noviembre y marzo, y el mes de verano más seco (agosto) recibe  $<30 \text{ l/m}^2$  de lluvia. La isla de Hawaii se compone de un complejo mosaico de comunidades de plantas y hay secciones de los últimos bosques secos tropicales que quedan en Hawaii, incluidas las siguientes comunidades de plantas

---

<sup>2</sup>[www.movebank.org](http://www.movebank.org)

principales: bosques de *Metrosideros*, matorrales de *Dodonaea* y bosques de *Myoporum-Sophora*. Aunque las cabras salvajes se encuentran en cinco de las ocho islas hawaianas más grandes en prácticamente todos los tipos de hábitat, una densidad particularmente alta de estos animales ocupa paisajes de montaña seca como la PTA.

En los paisajes de las islas, las cabras salvajes no nativas (*Capra hircus*) tienen un gran efecto y representan una amenaza significativa para la conservación y restauración de los ecosistemas nativos, ya que han degradado estos ecosistemas en las islas de todo el Océano Pacífico y han sido particularmente perjudiciales en los paisajes secos montañosos de Hawaii desde su introducción a finales del siglo XVIII. Como un agente extremo, las cabras salvajes modifican el paisaje directamente a través del consumo y el pisoteo de plantas nativas y no nativas que pueden conducir a efectos indirectos, como la modificación de la estructura y función del ecosistema. Estos efectos se intensifican en ecosistemas, como las islas hawaianas, que han evolucionado en ausencia de grandes herbívoros mamíferos. Así, comprender el alcance del hogar, el uso del espacio y los patrones de movimiento ayudará a priorizar la conservación del paisaje y esfuerzos de restauración en paisajes secos montañosos. Por lo tanto, la estimación del *home range* presenta un gran interés en este caso.

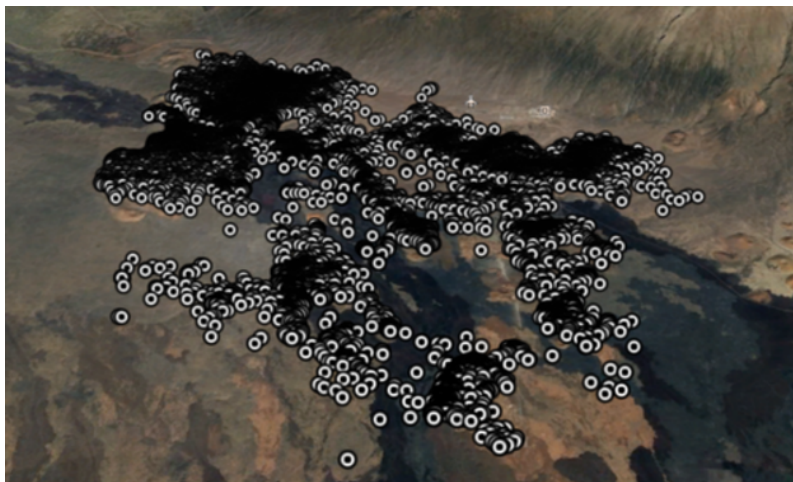


Figura 1.6: Ubicaciones de las cabras salvajes en Google Earth.

En julio de 2010, 13 cabras salvajes adultas fueron capturadas para seguimiento en la parte norte de la PTA, 6 machos y 7 hembras. Para lograr una muestra representativa, los individuos se seleccionaron de 12 grupos distintos, así como de sexo y edad diferentes. El seguimiento se llevó a cabo a través de collares con tecnología GPS de aproximadamente 450 gr. Estos estaban equipados con dos transmisores separados, uno de ellos para localizar a los animales, y el otro un transmisor para la descarga remota de datos vía satélite. Se

programaron para registrar la temperatura ambiente cada dos horas durante un año y descargar los datos de ubicaciones cada cinco días, obteniendo al final del estudio un total de 30246 datos de ubicaciones. Nueve de los trece collares duraron todo el periodo de estudio, dos de ellos fallaron por razones desconocidas y ocurrieron dos eventos de mortalidad. Presentamos en la *Figura 1.6* los datos de ubicaciones de las cabras importados a Google Earth.

En este caso, disponemos de un número muy elevado de datos, por lo que el cálculo y la visualización de algunos estimadores pueden ser computacionalmente muy costosas. Para solucionar esto, teniendo en cuenta que las ubicaciones fueron escogidas de varias cabras pertenecientes a grupos distintos, disminuirémos el tamaño de la muestra original seleccionando tres individuos representativos. Específicamente, dos machos y una hembra,

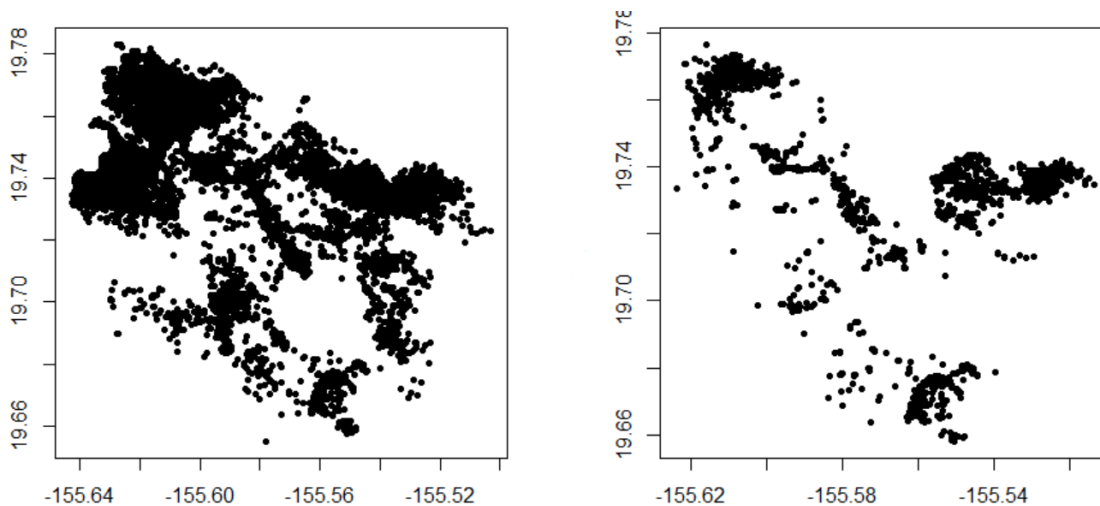


Figura 1.7: Muestra total de ubicaciones de las cabras (izquierda) frente a la submuestra reducida que usaremos (derecha).

y tomamos solamente las ubicaciones obtenidas en un periodo concreto, teniendo ahora un nuevo tamaño de muestra de 3932 ubicaciones. En la *Figura 1.7*, se representa la muestra total (izquierda) y la submuestra seleccionada (derecha).

## 1.2. Herramientas matemáticas para la estimación del *home range*

Con el fin de evaluar la calidad de los estimadores del soporte que revisaremos en este trabajo para reconstruir el *home range*, se emplean habitualmente en la literatura

las distancias entre conjuntos, que presentaremos en la *Sección 1.2.1*. Además, algunos de estos estimadores asumen que el soporte satisface alguna condición de forma concreta, las cuales serán presentadas en la *Sección 1.2.2*.

### 1.2.1. Distancias entre conjuntos

En la estimación de conjuntos, los resultados teóricos hacen especial énfasis en las propiedades asintóticas, especialmente la consistencia (el error del estimador tiende a cero cuando el tamaño de la muestra tiende a infinito). Dos de las distancias más empleadas en este contexto son la distancia en medida y la distancia Hausdorff. Ambas serán introducidas en las Definiciones 1.1 y 1.2, respectivamente.

**Definición 1.1.** Sean  $A$  y  $C$  dos subconjuntos de  $\mathbb{R}^d$  acotados y borelianos, la distancia en medida entre  $A$  y  $C$  se define como

$$d_\mu(A, C) = \mu(A \Delta C),$$

donde  $\Delta$  denota la diferencia simétrica habitual (*i.e.*  $x \in A \Delta B \Leftrightarrow x \in A$  o  $x \in B$ ), y  $\mu$  denota la medida de Lebesgue. En la *Figura 1.8* se representa en azul la diferencia simétrica entre los conjuntos  $A$  y  $B$ .

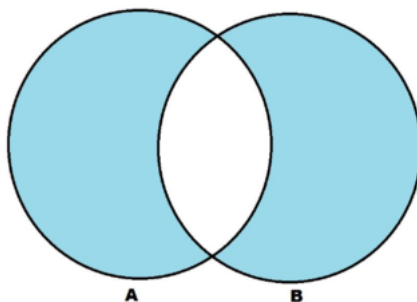


Figura 1.8: Diferencia simétrica entre los conjuntos  $A$  y  $B$ .

**Definición 1.2.** Dados dos conjuntos  $A, C \subset \mathbb{R}^d$  no vacíos, cerrados y acotados, se define la distancia Hausdorff entre  $A$  y  $C$  como

$$d_H(A, C) = \max \left\{ \sup_{x \in A} \inf_{y \in C} \|x - y\|, \sup_{y \in C} \inf_{x \in A} \|x - y\| \right\}$$

donde  $\| \cdot \|$  denota la norma euclídea, definida como  $\|x\| = \sqrt{x_1^2 + x_2^2 + \dots + x_d^2}$  con  $x = (x_1, x_2, \dots, x_d)$  un vector de  $\mathbb{R}^d$ .

Para simplificar, podemos expresar la distancia de Hausdorff de la forma siguiente:

$$d_H(A, C) = \max \left\{ \sup_{x \in A} d(x, C), \sup_{y \in C} d(y, A) \right\}$$

con  $d(x, C) = \inf_{y \in C} \{ \|x - y\| : y \in C \}$  y  $d(y, A) = \inf_{x \in A} \{ \|x - y\| : x \in A \}$ . Observamos

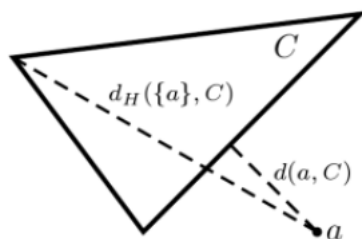


Figura 1.9: Distancia de Hausdorff entre  $\{a\}$  a  $C$ . Observamos que  $d_H(\{a\}, C)$  y  $d(\{a\}, C)$  son distintas.

en la *Figura 1.9* que la distancia de Hausdorff no tiene por qué coincidir con la distancia usual. Esta distancia refleja una noción de cercanía diferente y más visual entre conjuntos.

### 1.2.2. Propiedades geométricas

Algunos de los estimadores del soporte que emplearemos en este trabajo asumen que el soporte a reconstruir satisface alguna condición de forma. Revisaremos brevemente alguna de estas propiedades.

#### Convexidad

Dado un conjunto  $A \subset \mathbb{R}^d$ , se dice convexo si para cada par de puntos  $a, b \in A$  se tiene  $a(1 - t) + tb \in A$  para todo  $t \in [0, 1]$ .

De acuerdo con los conjuntos presentados en la *Figura 1.10*, en un conjunto no convexo, cada segmento que no cumpla esta hipótesis corta la frontera del conjunto por lo menos dos veces. Por tanto, la convexidad es una propiedad que depende esencialmente de la forma del borde del conjunto.

La envoltura convexa de un conjunto  $A$  es el menor conjunto convexo que contiene a  $A$ . Se denota por  $\text{conv}(A)$ . Además, coincide con la intersección de todos los hiperplanos

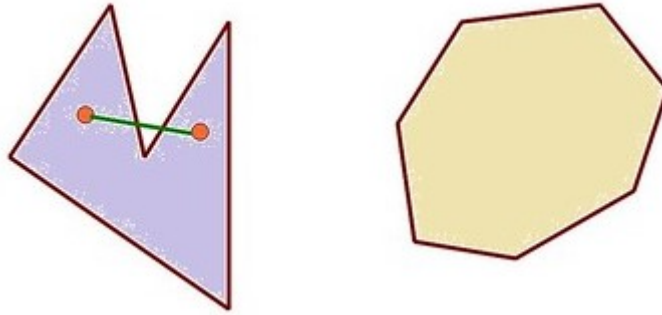


Figura 1.10: Conjunto no convexo (izquierda) y conjunto convexo (derecha).

que contienen al conjunto  $A$ .

**$r$ -convexidad:**

La  $r$ -convexidad es una propiedad geométrica que generaliza la convexidad. Así, la envoltura  $r$ -convexa de un conjunto  $A$  se define como la intersección de los complementarios de todas las bolas abiertas de radio  $r > 0$  que no intersecan al conjunto  $A$ , ver Definición 1.3.

**Definición 1.3.** Un conjunto cerrado  $A$  se dice  $r$ -convexo si  $A = C_r(A)$  donde

$$C_r(A) = \bigcap_{\text{int}(B(x,r)) \cap A = \emptyset} (\text{int}(B(x,r)))^c$$

se dice la *envoltura  $r$ -convexa* de  $A$  y siendo  $B(x,r)$  la bola cerrada de centro  $x$  y radio  $r$

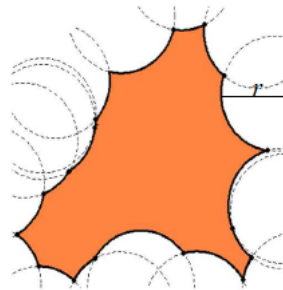


Figura 1.11: Sea  $A$  el conjunto representado en color naranja, se cumple que  $A = C_r(A)$  para  $r > 0$ . Por tanto,  $A$  es  $r$ -convexo.

con  $\text{int}$  denotando su interior, ver *Figura 1.11*. Para más detalles, ver *Edelsbrunner (1983)*.

Se puede ver que un conjunto cerrado convexo es  $r$ -convexo para todo  $r < 0$ . El recíproco es cierto cuando  $A$  tiene interior no vacío. Además, si un conjunto es convexo solo tendrá una componente conexa, mientras que los  $r$ -convexos pueden tener varias.

### Propiedad de rodamiento libre:

Algunos de los estimadores del soporte que revisaremos en este trabajo asumen que la frontera del conjunto a estimar es suave. Esta condición de regularidad está íntimamente ligada a la propiedad de rodamiento libre.

Un conjunto cerrado  $A$  se dice que cumple la propiedad de  $r$ -rodamiento libre exterior si “una bola puede rodar libremente a lo largo de la frontera, desde el exterior, teniendo contacto con todos los puntos de dicha frontera”. Formalmente, se dice que  $A$  satisface la propiedad de  $r$ -rodamiento libre exterior para  $r > 0$  si para todo  $s \in \partial A$  hay una bola cerrada  $B(x, r)$  de radio  $r$  tal que  $s \in B(x, r)$  y cuyo interior no interseca a  $A$ . Análogamente, se define el rodamiento libre interior sobre  $A^c$ . En la *Figura 1.12*, se representa esta propiedad. presentando un conjunto que no cumple la propiedad de rodamiento libre

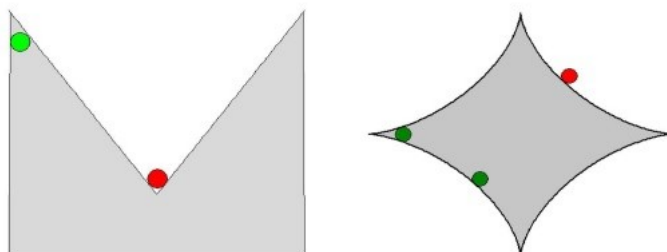


Figura 1.12: El conjunto de la izquierda no cumple la propiedad de rodamiento libre exterior ni interior, el de la derecha cumple el rodamiento libre exterior pero no interior.

interior ni exterior (izquierda), y otro que cumple la propiedad de rodamiento libre exterior pero no la interior (derecha). *Rodríguez-Casal y Saavedra-Nieves (2016)* demostraron que la propiedad de  $r$ -rodamiento libre se preserva en el límite. Es decir, sea  $\{r_n\}_{n \in \mathbb{N}}$  una sucesión de términos positivos convergente a  $r$  ( $r > 0$ ) y dado un conjunto  $A$  que satisface la propiedad de  $r_n$ -rodamiento libre, entonces  $A$  también satisface la propiedad de  $r$ -rodamiento libre. Estos autores proponen una propiedad  $(R)$  íntimamente ligada con las anteriores para garantizar que la frontera del conjunto sea suave (sin picos):

$(R)$  El conjunto  $A$  satisface la propiedad de rodamiento libre exterior con constante  $\tau$  y

la interior con constante  $\lambda$ .

La primera imagen de la *Figura 1.13* muestra un conjunto que satisface  $(R)$  y con frontera suave. En *Cuevas et al. (2012)* se demuestra que si  $S$  es compacto y  $r$ -convexo, entonces

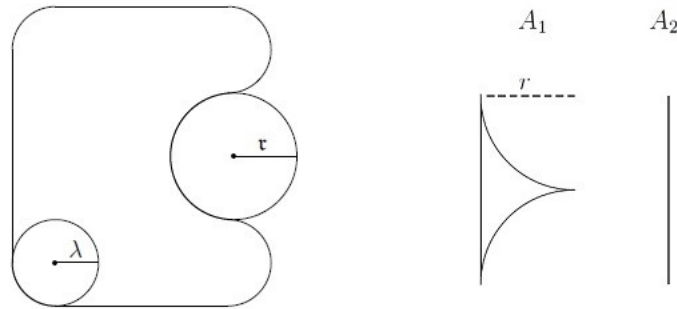


Figura 1.13: Conjunto verificando la propiedad de rodamiento libre exterior con radio  $\tau$  y la interior con radio  $\lambda$  (derecha).  $A_1 \cup A_2$  cumplen la condición de  $r$ -rodamiento, sin embargo,  $A_1 \cup A_2$  no es  $r$ -convexo (derecha).

cumple la propiedad de  $r$ -rodamiento libre exterior. El recíproco no es siempre cierto, como se muestra en la segunda imagen de la *Figura 1.13*. *Rodríguez-Casal y Saavedra Nieves (2016)* probaron que  $(R)$  es una condición suficiente para asegurar que la propiedad de  $\tau$ -rodamiento libre exterior implica la  $\tau$ -convexidad. Esta condición fue analizada por *Walther (1997)* y *Walther (1999)*, pero solamente para el caso en el que  $\tau = \lambda$ , en el cual vio que  $A$  es  $\tau$ -convexo.



## Capítulo 2

# Métodos de estimación del soporte para reconstruir el *home range*

La estimación no paramétrica del *home range* puede tratarse como un problema clásico de estimación del soporte donde se asume que las observaciones son independientes. En este trabajo, nos centraremos en los métodos de estimación globales, en los cuales se aplica un determinado algoritmo a todo el conjunto de datos a la vez, a diferencia de los métodos locales, que son aquellos en los que aplicamos un método global pero esta vez a cada uno de los subconjuntos de una partición de datos escogida adecuadamente para a continuación tomar la unión de todos los estimadores así obtenidos.

En los casos más generales de estimación del soporte, no se imponen condiciones geométricas sobre el conjunto teórico que se va a estimar. Sin embargo, se podrían considerar estimadores más sofisticados si se proporciona información adicional al respecto. Recordemos que el principal objetivo de la estimación no paramétrica del soporte es reconstruir  $S$  a partir de una muestra  $\aleph_n = \{X_1, \dots, X_n\}$  dentro de  $S$ . Este capítulo se organiza como sigue. En la *Sección 2.1*, revisaremos algunos estimadores para el soporte sin imponer sobre este restricciones de forma, en la *Sección 2.2*, impondremos restricciones de forma sobre el soporte y finalmente, a pesar de que no es el objetivo de este trabajo, citaremos brevemente algunas de las alternativas en la literatura que permiten incorporar la dependencia temporal al estimador del soporte. Revisaremos y representaremos los estimadores obtenidos para cada uno de los conjuntos de datos presentados en la *Sección 1.1*.

### 2.1. Sin restricciones de forma

A continuación, presentaremos los estimadores del soporte clásicos más generales, aquellos que no asumen ninguna restricción geométrica sobre el soporte  $S$  a estimar.

## Estimador muestral

La propia muestra  $\aleph_n$  es un estimador consistente del soporte en distancia de Hausdorff. Es decir, con probabilidad uno,  $d_H(\aleph_n, S) \rightarrow 0$  cuando  $n \rightarrow \infty$ . En la *Figura 2.1* representamos las muestras de datos de ubicaciones para las cabras y el lobo, respectivamente.

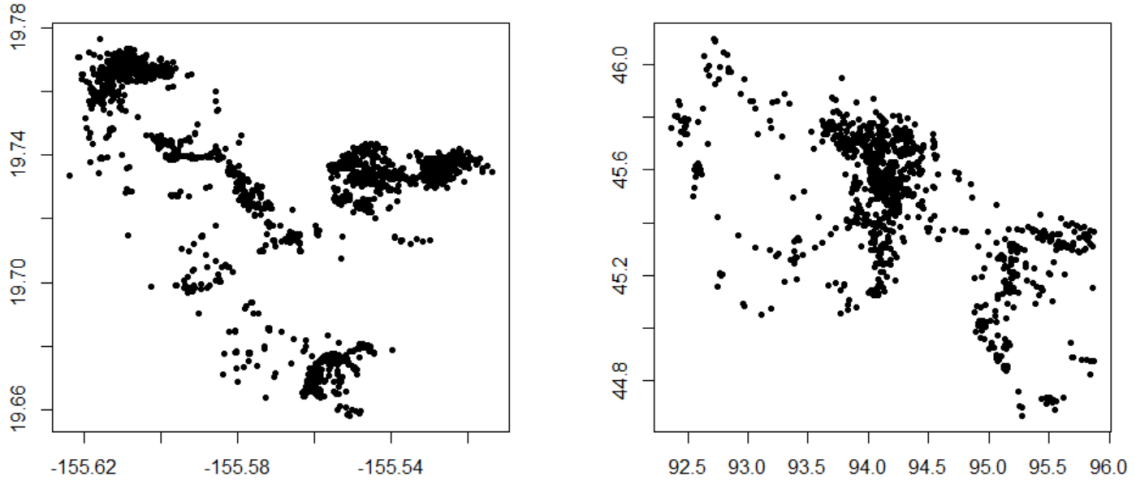


Figura 2.1: Muestras de ubicaciones de las cabras (izquierda) y del lobo (derecha).

## Estimador de Devroye y Wise

Si no se imponen restricciones de forma sobre el soporte  $S$ , existen varios estimadores que lo reconstruyen bajo condiciones muy generales en  $S$ . *Devroye y Wise (1980)* propusieron reconstruir el soporte como una versión suavizada de la muestra. Concretamente, presentaron el estimador para  $S$  dado por:

$$S_n = \bigcup_{i=1}^n B(X_i, \epsilon)$$

siendo  $B(X_i, \epsilon)$  la bola cerrada de centro  $X_i$  y radio  $\epsilon > 0$  para cada  $i \in \{1, \dots, n\}$ . Por tanto, este estimador es una versión “dilatada” de la muestra original.

*Devroye y Wise (1980)* demostraron que  $S_n$  es  $d_\mu$ -consistente en probabilidad y casi seguramente. Si  $\epsilon \rightarrow 0$  y  $n\epsilon^d \rightarrow \infty$ , entonces  $d_\mu(S, S_n) \rightarrow 0$ , en probabilidad.

El principal problema de este estimador, es seleccionar el radio  $\epsilon$  óptimo de las bolas. En la *Figura 2.2*, hemos representado el estimador de *Devroye-Wise* para distintos valores de  $\epsilon$  sobre las ubicaciones del lobo.

Mostramos en la primera imagen de la *Figura 2.2*, el resultado tomando como radio  $\epsilon = 0,03$ . Observamos que obtenemos un estimador con muchas componentes conexas y

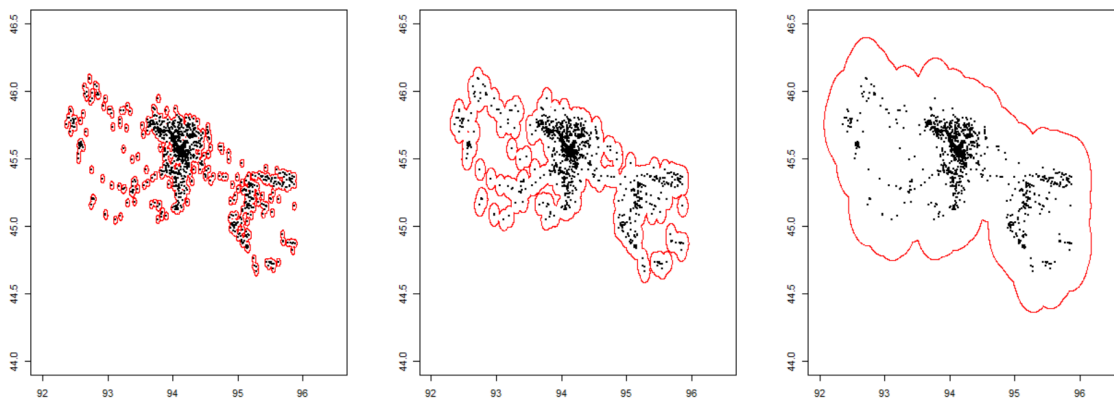


Figura 2.2: Estimadores Devroye-Wise para los datos del lobo con radios  $\epsilon = 0,03$ ,  $\epsilon = 0,08$  y  $\epsilon = 0,3$ , respectivamente.

muy fragmentado. De hecho para valores de  $\epsilon$  suficientemente pequeños tendríamos tantas componentes conexas como puntos en la muestra. Esto puede llegar a ser un problema grave, ya que con este método nos arriesgamos a dejar fuera del estimador áreas grandes que podrían pertenecer al *home range*. Es decir, al tomar un radio  $\epsilon$  demasiado pequeño, el estimador *Devroye-Wise* podría no aportarnos ninguna información nueva a la que ya sabemos conociendo los puntos muestrales. Por tanto, el *home range* se infraestimaría considerablemente.

Podemos pensar en escoger un radio mayor para solucionar este problema. Tomamos a continuación como radio  $\epsilon = 0,08$ , obteniendo el estimador presentado en la segunda imagen de la *Figura 2.2*. En este segundo caso, tenemos también varias componentes conexas, aunque un número mucho menor que en el primer caso tomando  $\epsilon = 0,03$ .

Por último, nos fijamos en la tercera imagen de la *Figura 2.2*, donde tomamos  $\epsilon = 0,3$ . Obviamente, cuanto mayor sea el radio  $\epsilon$ , menos componentes conexas presentará este estimador, como en este caso, que obtenemos solamente una componente conexas. Así, con radios más grandes, podríamos sobreestimar considerablemente el *home range* del lobo.

Para la muestra de cabras salvajes en Hawaii, hemos representado en la *Figura 2.3* el estimador de *Devroye-Wise* tomando como radios  $\epsilon = 0,004$ ,  $\epsilon = 0,01$  y  $\epsilon = 0,03$ , respectivamente.

En la primera imagen de la *Figura 2.3*, somos capaces de distinguir varias componentes conexas en el estimador, pero nuevamente, podemos pensar que estamos infraestimando el

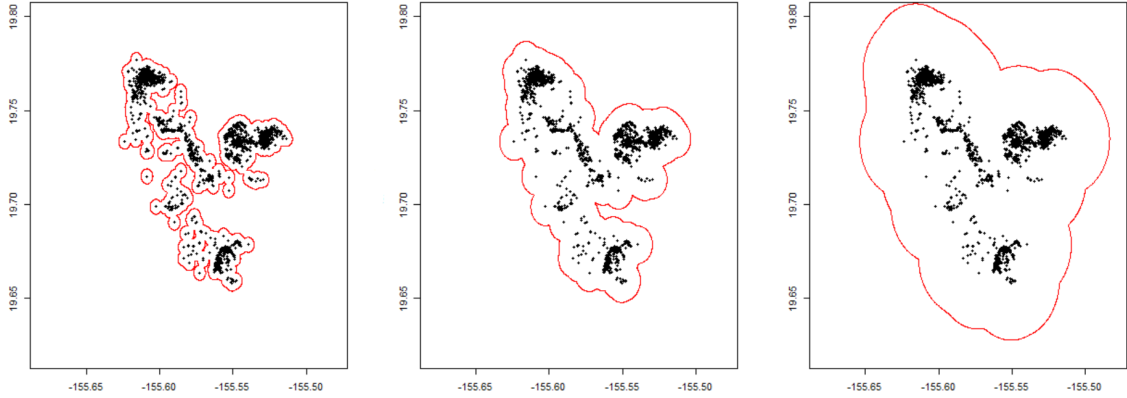


Figura 2.3: Estimadores Devroye-Wise para los datos de las cabras con radios  $\epsilon = 0,004$ ,  $\epsilon = 0,01$  y  $\epsilon = 0,03$ , respectivamente.

*home range*. Sin embargo, si se considera un radio mayor, observamos en la segunda imagen que pasamos a tener solamente una componente conexa, y tomando un radio todavía más grande, en la última imagen, obtenemos de nuevo una sola componente conexa, pero un estimador que presenta bastante diferencia con respecto al anterior, viendo en este caso que podemos estar cometiendo un error de sobreestimación considerable.

En definitiva, radios demasiado grandes sobreestimarán el *home range* y radios muy pequeños, lo infraestimarán. La clave sería seleccionar un radio  $\epsilon$  óptimo que pueda estimar el *home range* de forma que haya un balance equilibrado, sin caer en la sobreestimación del conjunto ni en la infraestimación, por lo que puede compensarnos imponer alguna restricción de forma a  $S$ . Existen algunos trabajos en la literatura que han abordado el problema de seleccionar  $\epsilon$ . En particular, [Baíllo y Cuevas \(2001\)](#) asumen que  $S$  es estrellado e incorporan esta información adicional al estimador de *Devroye-Wise* seleccionando el valor de  $\epsilon$  más pequeño de modo que  $S_n$  tenga también forma de estrella.

### Estimación de la densidad con estimador tipo núcleo (KDE)

[Cuevas y Fraiman \(1997\)](#) consideran una reconstrucción más sofisticada para el soporte. Se basan en el uso de un estimador no paramétrico de la densidad  $f_n$ . Si  $f$  es la densidad subyacente, el soporte esencialmente coincide con el conjunto  $\{f > 0\}$ . [Dobrow \(1992\)](#) sugiere una conexión explícita entre el problema del soporte y la teoría de estimación de la densidad, quien propuso una idea *plug-in* (estimadores de la forma  $\{f_n > c\}$  con  $c > 0$ ). La propuesta de [Dobrow \(1992\)](#) fue estimar  $S$  por

$$\tilde{S}_n = \{\hat{f}_n > 0\}$$

donde  $\hat{f}_n$  es un estimador de densidad tipo núcleo,

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n K_H(x - X_i)$$

con  $K_H(t) = |H|^{-1/2} K(H^{-1/2}t)$  siendo  $|\cdot|$  la notación de determinante,  $K : \mathbb{R}^d \rightarrow \mathbb{R}$  una función tipo núcleo y  $H$  una matriz de dimensión  $d \times d$  simétrica y definida positiva. El estimador  $\hat{f}_n$  dependerá de la matriz  $H$ , ver [Wand y Jones \(1995\)](#). De hecho, su elección es un problema clásico en la estadística no paramétrica.

Un estimador del tipo  $\tilde{S}_n$  es una elección simple y natural, pero presenta dos limitaciones: primero, estamos restringidos a usar  $K$  con soporte compacto, para evitar el estimador inútil  $\tilde{S}_n = \mathbb{R}^d$ . Segundo, si  $S(K) = \{K > 0\}$  es acotado, el estimador  $\tilde{S}_n = \{\hat{f}_n > 0\}$  es de nuevo una unión finita de tipo *Devroye-Wise*, donde las bolas  $B(X_i, \epsilon)$  son reemplazadas por  $X_i + hS(K)$ . [Cuevas y Fraiman \(1997\)](#) consideran una versión modificada introduciendo un nuevo parámetro de ajuste, además del parámetro de suavizado de  $\hat{f}_n$ . Su propuesta es estimar  $S$  como

$$S_n = \{\hat{f}_n > c_n\}$$

donde  $\hat{f}_n$  es un estimador de tipo núcleo y  $c_n$  es una sucesión numérica adecuada convergente a cero. Este parámetro proporciona más flexibilidad en la forma de  $S_n$ , que tendrá normalmente una frontera diferenciable (a diferencia del estimador de *Devroye-Wise*). Así, este estimador puede considerarse como una versión suavizada del propuesto por *Devroye-Wise*.

En este contexto, podemos considerar una definición del *home range* alternativa. De acuerdo con [Baíllo y Chacón \(2018\)](#) puede definirse como el conjunto de nivel  $\{f > c_{0,05}\}$  tal que su contenido en probabilidad es del 95 %, esto es, el conjunto de nivel asociado al nivel  $c_{0,05}$  tal que

$$0,95 = \int_{\{f > c_{0,05}\}} f(x) dx.$$

Este conjunto de nivel representa el soporte efectivo o substancial de la distribución, pues elimina o prescinde de las zonas de baja densidad. Así, un estimador *plug-in*, y por tanto, del soporte efectivo, será el conjunto de nivel análogo  $\{\hat{f}_n > \hat{c}_{0,05}\}$ , donde  $\hat{c}_{0,05}$  puede calcularse como el cuantil 0,95 de  $\{\hat{f}_n(X_1), \dots, \hat{f}_n(X_n)\}$ . Por tanto,

$$0,95 \geq \int_{\{\hat{f}_n > \hat{c}_{0,05}\}} \hat{f}_n(x) dx. \quad (2.1)$$

Además, se define la muestra efectiva como los puntos  $X_i$  de la muestra  $X_n$  que pertenecen al estimador del soporte efectivo definido en (2.1).

Por supuesto, valores diferentes a 0,95 podrían ser considerados dependiendo de las zonas de baja densidad que el usuario desee descartar. Por ejemplo, 0,99 o 0,90 son valores comunmente utilizados.

En la *Figura 2.4*, hemos representado las reconstrucciones del *home range* obtenidas para los datos del lobo con contenidos en probabilidad 95 % y 80 %, respectivamente. Para ello, hemos hecho uso del paquete `adehabitatHR`<sup>1</sup> de R. Observamos que para el primer

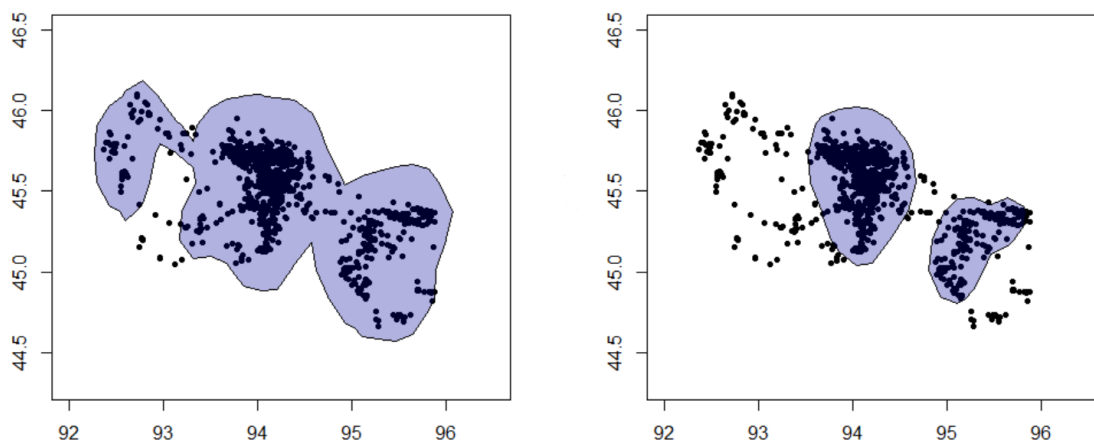


Figura 2.4: Estimadores KDE para los datos del lobo con contenido en probabilidad del 95 % y 80 %, respectivamente.

caso, obtenemos solamente una componente conexa, mientras que para el segundo, tenemos dos, correspondientes a las dos zonas geográficas donde la densidad toma valores más altos. En otras palabras, estimamos las dos modas más grandes.

En la *Figura 2.5* se muestra el estimador KDE aplicado a la submuestra seleccionada de ubicaciones de las cabras tomando como contenido en probabilidad 95 %, 90 % y 75 %, respectivamente.

Con contenido en probabilidad del 95 % este estimador nos proporciona solamente una componente conexa, mientras que, a medida que bajamos esta probabilidad, obtenemos dos y hasta tres componentes conexas bien diferenciadas, las cuales se corresponden con las modas de la densidad, es decir, las zonas donde los datos están más concentrados. Puede verse que cada una de las tres cabras de la muestra, se corresponde con una de las tres modas y, por tanto, ocupa una zona geográfica diferente.

<sup>1</sup><https://cran.r-project.org/web/packages/adehabitatHR/adehabitatHR.pdf>

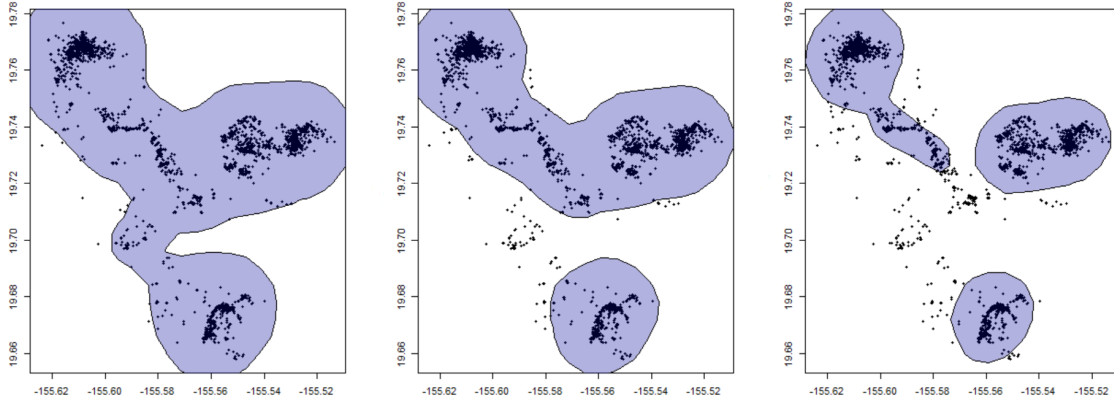


Figura 2.5: Estimadores KDE para los datos de las cabras con contenido en probabilidad del 95 % y 90 % y 75 %, respectivamente.

## 2.2. Con restricciones de forma

A continuación, revisaremos los estimadores para el soporte cuando se asume que éste satisface alguna propiedad geométrica. Comenzaremos imponiendo una condición que puede ser muy restrictiva, la convexidad, y a continuación, suavizaremos esta condición imponiendo que el soporte sea  $r$ -convexo o verifique la condición  $(R)$  introducida en la *Sección 1.2*.

### Envoltura convexa

La pregunta de cómo construimos un estimador adecuado tiene una respuesta bastante natural si asumimos  $S$  convexo. Así, bastaría considerar la envoltura convexa de la muestra,

$$S_n = \text{conv}(\mathfrak{N}_n)$$

la cual proporciona un estimador simple. Recordemos que la envoltura convexa de la muestra coincide con la intersección de todos los conjuntos convexos que incluyen a nuestra muestra de ubicaciones. Es decir, el menor convexo que contiene a la muestra.

*Korostel'ev y Tsybakov (1993)* se refieren a *Geffroy (1964)*, *Rényi y Sulanke (1963)*, y *Rényi y Sulanke (1964)* como los primeros trabajos en estimación del soporte bajo restricciones de convexidad. *Rényi y Sulanke (1963)* y *Rényi y Sulanke (1964)* estudiaron el caso en el que  $S \subset \mathbb{R}^2$  es un soporte convexo. *Dümbgen y Walther (1996)* estudiaron cómo de cerca puede estar  $S$  de la envoltura convexa de los puntos de la muestra. Esta proximidad se estudió en términos de la distancia de Hausdorff en una dimensión arbitraria  $d$ . Concretamente, se probó que  $d_H(S, S_n) = \mathcal{O}((\log n/n)^{1/d})$ , casi seguro.

Para los dos conjuntos de datos reales considerados en este trabajo, representaremos las envolturas convexas para nuestros datos empleando el paquete `tripack`<sup>2</sup> de R. En la *Figura 2.6*, se representa la envoltura convexa de los datos de ubicaciones del lobo. Claramente,

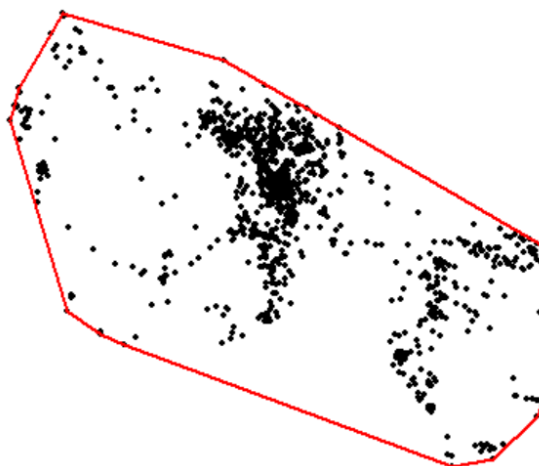


Figura 2.6: En rojo, envoltura convexa de los datos del lobo.

la suposición de convexidad es muy restrictiva en este caso particular, ya que solo estima una componente conexas y esta envoltura podría recoger lugares inaccesibles para los lobos, como ríos (lo cual puede provocar que tengamos más de una componente conexas), lagos, o puede presentar el problema de que el estimador no esté dentro de nuestras fronteras elegidas. Es decir, si quisiéramos centrarnos solamente en Mongolia, este estimador puede tomar dentro del *home range* zonas que no pertenezcan a este país. Además, a simple vista se detectan regiones geográficas sin puntos muestrales que claramente no pertenecen al soporte de la distribución.

En el caso de las cabras, podemos visualizar la envoltura convexa de la muestra con los datos de ubicaciones de los 13 individuos frente a la de cada uno de ellos, viendo así cuánto puede condicionar cada uno a este estimador. En la *Figura 2.7*, se muestra la envoltura convexa de la muestra completa (rojo) y las envolturas convexas calculadas a partir de las localizaciones de cada una de las 13 cabras. La unión de estas 13 envolturas no coincide exactamente con la envoltura representada en rojo. Además, podemos observar que la mayoría de individuos se encuentran en la zona norte del mapa, siendo el individuo correspondiente a color azul e incluso el correspondiente al color violeta muy influyentes para la envoltura, ya que si los eliminamos, el estimador se reduciría prácticamente a la

---

<sup>2</sup><https://cran.r-project.org/web/packages/tripack/tripack.pdf>

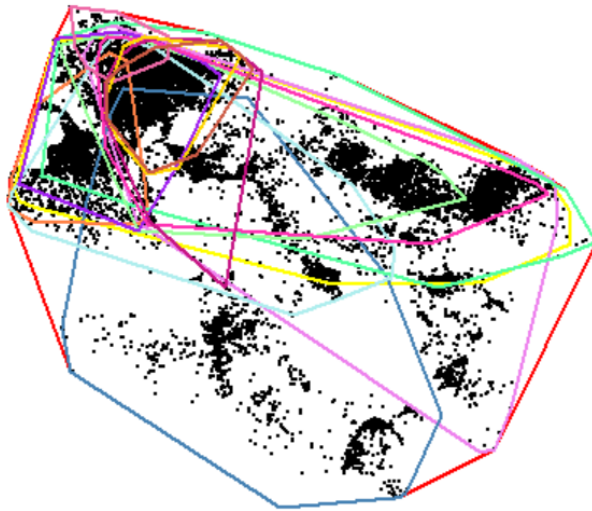


Figura 2.7: Envoltura convexa de la muestra total, en rojo, frente a las envolturas de los demás individuos.

mitad. En la *Figura 2.8*, comparamos el estimador empleando la muestra total (primera imagen), sin el individuo correspondiente al color azul (segunda imagen), el cual contiene 1282 ubicaciones, y sin el individuo correspondiente al color violeta (tercera imagen), con 1254 ubicaciones, quedándonos entonces con 28 964 ubicaciones al eliminar un individuo y con 27 710 ubicaciones al eliminar los dos. Vemos que eliminando un individuo la envoltura

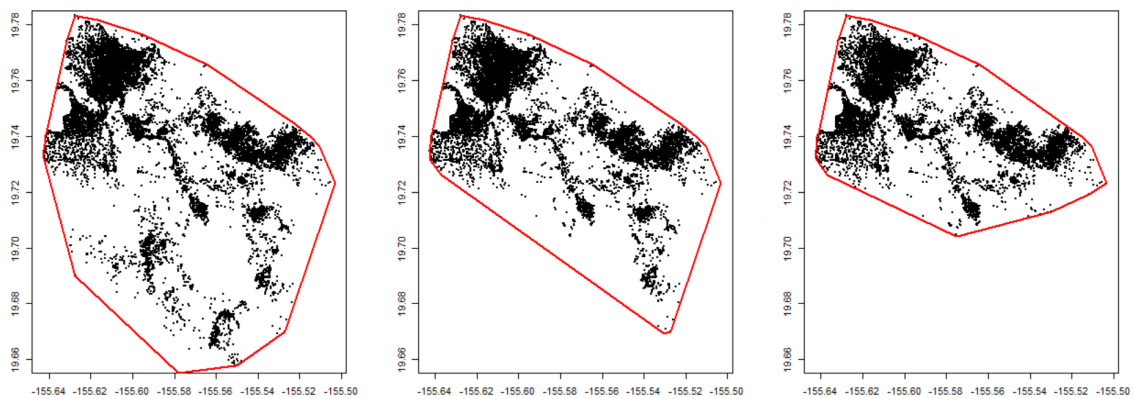


Figura 2.8: Envoltura convexa de la muestra total, eliminando el individuo azul y el violeta.

se reduce casi un tercio con respecto a la original, y eliminando ambos, esta se reduce prácticamente a la mitad, es decir, solamente estos dos individuos tienen un gran poder para condicionar la forma del estimador con la muestra total.

En la *Figura 2.9*, representamos la envoltura convexa con los 13 individuos frente a cada una de las envolturas convexas de los tres individuos que seleccionamos para construir una submuestra representativa (izquierda) y la envoltura de estos nuevos datos (derecha). Se

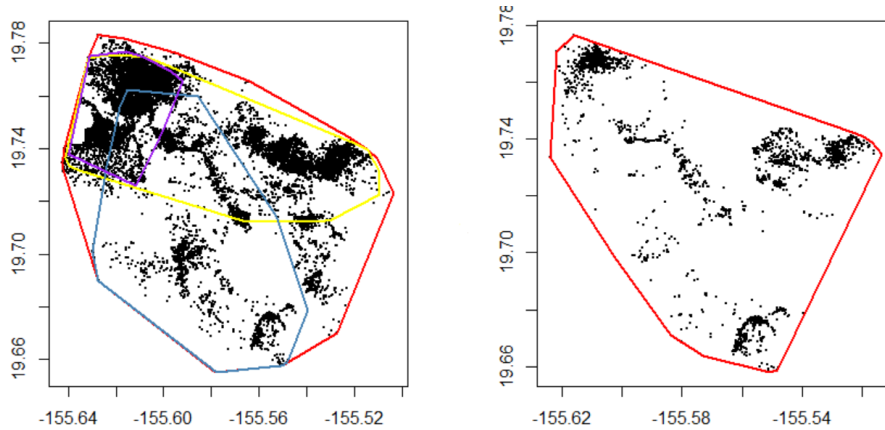


Figura 2.9: Envoltura convexa de la muestra total y de los tres individuos seleccionados (izquierda) y envoltura convexa de estos tres individuos (derecha).

observa que la envoltura calculada a partir de la muestra completa y la submuestra formada por los tres individuos son bastante parecidas. De nuevo, se observan zonas sin datos en ambas envolturas por los que éste podría no ser el estimador del *home range* más adecuado.

### Envoltura $r$ -convexa

En la geometría computacional, la noción de  $r$ -convexidad ( $r > 0$ ) se introdujo como una generalización del concepto de convexidad, para relajarlo y flexibilizarlo. En la *Sección 1.2.2*, vimos que la envoltura  $r$ -convexa de la muestra es la intersección de los complementarios de todas las bolas abiertas de radio  $r$  que no contienen ningún punto de la muestra. Por tanto, la frontera de este estimador está formada por arcos de circunferencia de radio  $r$ . A diferencia de la envoltura convexa que solo estima una componente conexa, la envoltura  $r$ -convexa permite reconstruir un soporte con varias componentes o grupos.

En general, si se asume que el soporte  $S$  es  $r$ -convexo, un estimador  $S_n$  natural sería la envoltura  $r$ -convexa de los puntos de la muestra,  $C_r(X_n)$ . El papel del parámetro de forma  $r$  (que es en principio desconocido) es similar al del radio  $\epsilon$  para el estimador de *Devroye-Wise*. Valores grandes de  $r$  nos proporcionan estimadores con pocas componentes conexas, y valores pequeños, estimadores mucho más fragmentados. En las *Figuras 2.10* y *2.11*, se muestran las envolturas  $r$ -convexas para los datos del lobo y las cabras, tomando

distintos valores del radio de las bolas  $r$ . Si  $r$  es suficientemente cercano a cero, el número de

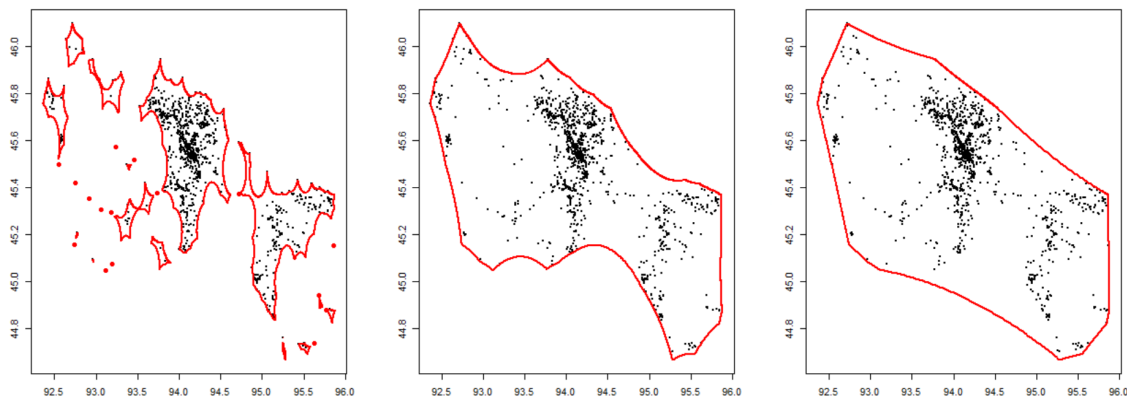


Figura 2.10: Envoltura  $r$ -convexa de los datos del lobo con  $r = 0,1$ ,  $r = 1$  y  $r = 10$ , respectivamente.

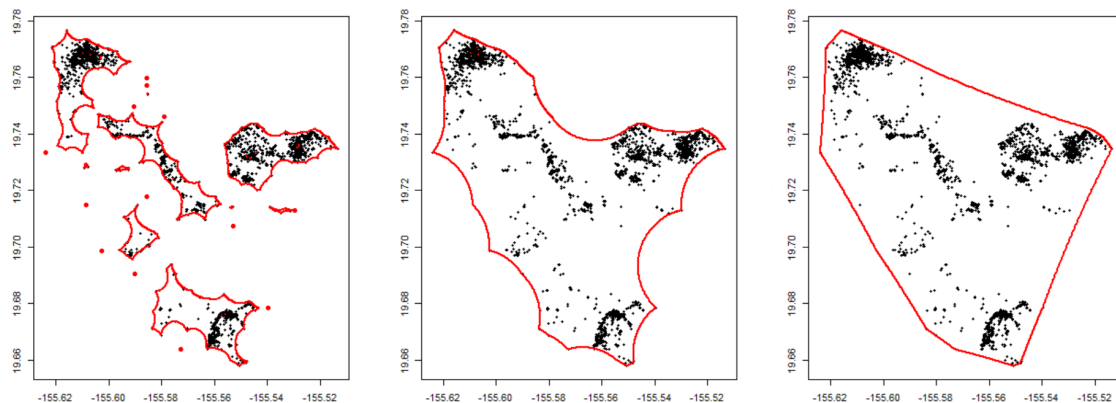


Figura 2.11: Envoltura  $r$ -convexa de los datos de las cabras con  $r = 0,005$ ,  $r = 0,02$  y  $r = 0,5$ , respectivamente.

componentes conexas de la envoltura coincidirá con el tamaño de muestra. Por el contrario, si  $r$  tiende a infinito, la envoltura  $r$ -convexa coincidirá con la envoltura convexa de la muestra. Por tanto,  $r$  juega el papel de un parámetro de forma y/o suavizado.

Uno de los problemas principales para reconstruir el *home range* si se asume que su soporte es  $r$ -convexo es elegir el parámetro  $r$  para que nuestra estimación sea óptima. En la *Figura 2.10*, aplicando distintos radios a los datos de ubicaciones del lobo, efectivamente obtenemos estimadores más fragmentados para radios pequeños, similares a los del tipo *Devroye-Wise*, como muestra la primera imagen de la *Figura 2.10* con  $r = 0,1$ , y estima-

dores cada vez más similares a la envoltura convexa a medida que el radio aumenta, como se muestra en la segunda imagen con  $r = 1$ , donde todavía podemos distinguir que la frontera está formada por arcos de circunferencia, y en la tercera con  $r = 10$ , muy similar a la envoltura convexa.

En la *Figura 2.11*, se muestran envolturas  $r$ -convexas para el conjunto de ubicaciones de cabras salvajes tomando como radios  $r = 0,005$ ,  $r = 0,02$  y  $r = 0,5$ . Obtenemos conclusiones similares que en caso de las ubicaciones del lobo.

Entonces, queda claro que el desconocimiento del valor real del parámetro  $r$  puede provocar que los estimadores obtenidos sobreestimen o infraestimen el *home range* considerablemente. Por eso, resolver el problema de seleccionarlo a partir de los datos de forma automática es clave para obtener estimadores del soporte realistas. *Rodríguez-Casal y Saavedra-Nieves (2019)* proponen un método automático para elegir el valor óptimo del

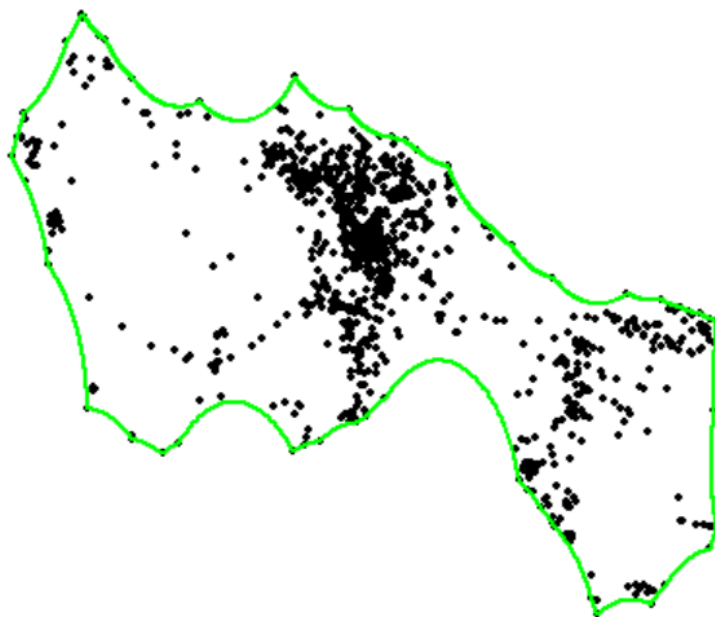


Figura 2.12: Envoltura  $r$ -convexa para los datos de ubicaciones del lobo con  $r$  estimado usando el método propuesto en *Rodríguez-Casal y Saavedra-Nieves (2019)*.

parámetro  $r$  asumiendo, entre otras condiciones de regularidad, que el soporte  $S$  cumple la condición  $(R)$  introducida en la *Sección 1.2.2*. En esta memoria no describiremos el algoritmo propuesto, que puede ser consultado en *Rodríguez-Casal y Saavedra-Nieves (2019)*. Sin embargo, ilustraremos su funcionamiento usando los dos conjuntos de datos reales presentados previamente.

Empleando este algoritmo, el valor estimado del parámetro  $r$  para los datos de ubica-

ciones del lobo es 0,43. En la *Figura 2.12*, se representa la envoltura 0,43-convexa de los puntos muestrales. Observamos que este estimador del soporte  $S$  nos proporciona una sola componente conexa.

Si en lugar de estimar el soporte completo  $S$ , estimamos el soporte efectivo con contenidos en probabilidad 0,99 y 0,85, el número de componentes conexas aumenta. Una vez seleccionada la muestra efectiva, aplicamos el algoritmo propuesto por *Rodríguez-Casal y Saavedra-Nieves (2019)* para estimar el parámetro  $r$ .

En la *Figura 2.13*, se muestra la envoltura  $r$ -convexa aplicada al 99 % y 85 % de los datos, obteniendo un radio estimado de  $\hat{r} = 0,2$  y  $\hat{r} = 0,28$ , respectivamente. Observamos

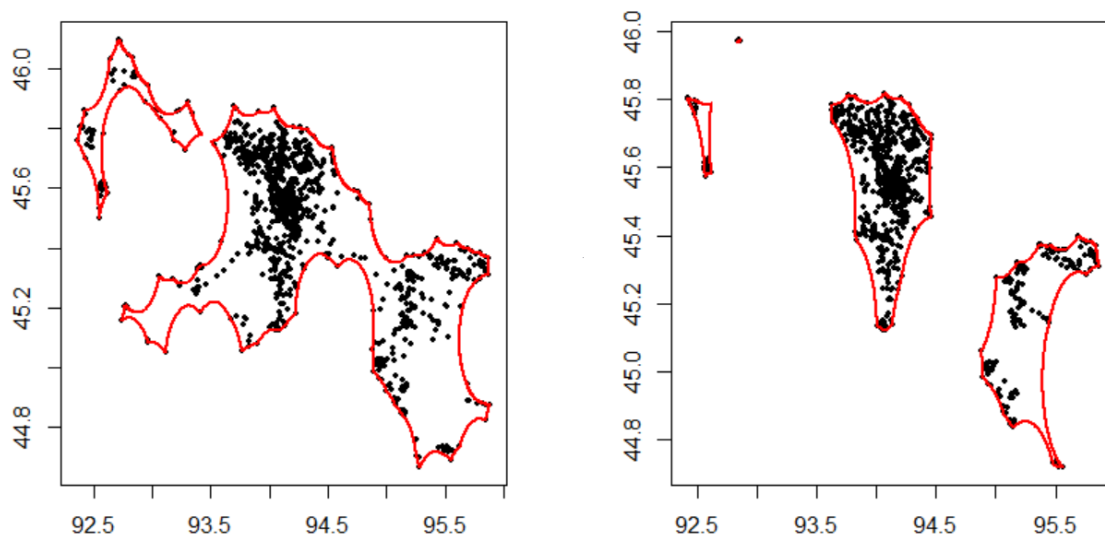


Figura 2.13: Envoltura  $r$ -convexa para los datos de ubicaciones del lobo con el 99 % y 85 % de los datos respectivamente.

que ya con el 99 % de la muestra obtenemos dos componentes conexas y con el 85 % llegamos a tener tres y un punto aislado. Vemos que la componente conexa más grande con la muestra al 99 % se divide en dos componentes al pasar al 85 % de la muestra, y la restante, vuelve a dividirse. Es razonable pensar por tanto que el *home range* del lobo pueda tener más de una componente conexa si las ubicaciones en zonas de baja probabilidad son eliminadas.

Pasamos ahora a representar la envoltura  $r$ -convexa con el  $r$  estimado para los datos de las cabras, empleando de nuevo el algoritmo de *Rodríguez-Casal y Saavedra-Nieves (2019)*, obteniendo un radio de  $\hat{r} = 0,0092$ . Presentamos este estimador en la *Figura 2.14*.

Vemos que obtenemos una gran componente conexa junto con tres puntos aislados.

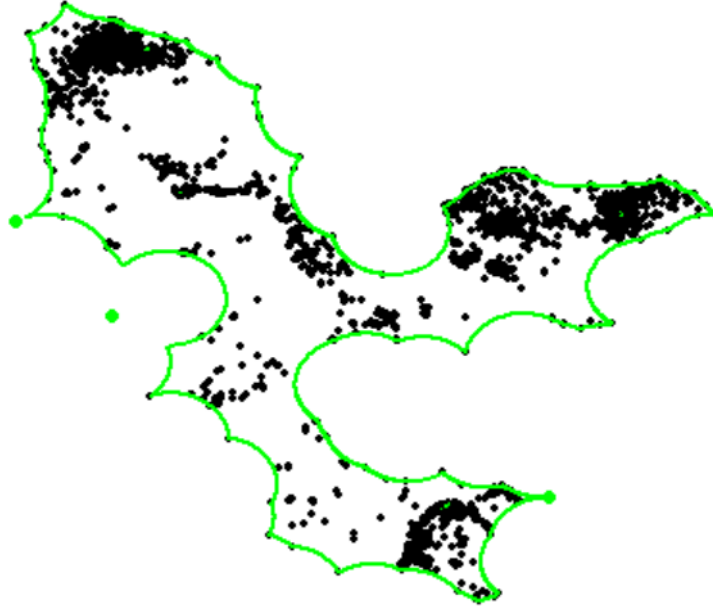


Figura 2.14: Envoltura  $r$ -convexa para los datos de ubicaciones de las cabras con radio óptimo  $\hat{r} = 0,0092$ .

En la práctica, podríamos pensar entonces que el *home range* de las localizaciones de las cabras tiene por tanto una sola componente conexa.

Para obtener todos los estimadores vistos hasta ahora, se ha asumido que los datos de ubicaciones son independientes, pero existen alternativas en la literatura que tienen en cuenta la dependencia temporal. Esta dependencia puede incorporarse en el estimador del *home range*, ya que también puede ser una fuente de información del comportamiento del animal. Algunas propuestas existentes en la literatura para incorporar la dependencia del tiempo son la estimación de densidad geográfica (TGDE), la estimación de densidad tipo núcleo incorporando la variable tiempo (MKDE), o el T-LoCoH. Dado que la revisión de estos métodos no son el objetivo de este trabajo, no serán descritos. Sin embargo, información detallada puede consultarse en [Baíllo y Chacón \(2018\)](#), que revisan en profundidad este problema.

## Capítulo 3

# Ranking de *home ranges*

Una de las preguntas clave después de calcular varios estimadores del *home range* basados en los mismos datos de ubicaciones es cómo seleccionar el más realista.

En la mayoría de los estudios, los biólogos eligen el resultado más sólido de acuerdo con la información previa sobre el animal o su especie. La posibilidad de hacer una elección automatizada, basada en criterios estadísticos objetivos, entre una colección de estimaciones del *home range* es, en cierto sentido, todavía una cuestión abierta.

Ha habido algunos intentos de comparar los *home range* de acuerdo con diferentes criterios. Una posibilidad, especialmente si las ubicaciones se tratan como datos independientes, es separar la muestra en dos submuestras de ubicaciones: una muestra para construir el estimador de *home range* y otra de prueba para verificar su precisión predictiva. [Fleming et al. \(2015\)](#) compara las áreas de las estimaciones del *home range* obtenidas con la muestra completa con las obtenidas con la primera mitad de los datos (cronológicamente hablando). [Long y Nelson \(2015\)](#) calculan las áreas de diferentes estimadores del *home range* obtenidos con métodos distintos. [Kenward et al. \(2001\)](#) estudia la relación entre el logaritmo del área del *home range* y, por ejemplo, los factores ambientales que influyen en el comportamiento animal (como la disponibilidad de alimentos o la densidad de población). Además del área, [Steiniger y Hunter \(2013\)](#) usan la complejidad de la forma dada teniendo en cuenta agujeros y parches.

En [Baíllo y Chacón \(2018\)](#), se propone una nueva forma automática de elegir el *home range* “más adecuado” entre una colección, o al menos un criterio para clasificarlos. La idea es minimizar una medida del exceso de extensión cometida por el *home range* en comparación con las ubicaciones originales. El procedimiento es válido para cualquier tipo de *home range*, independientemente de la forma en que haya sido construido.

En este trabajo, consideraremos la propuesta de [Fleming et al. \(2015\)](#). Es decir, compararemos las áreas obtenidas al estimar el *home range* con la muestra completa de ubica-

ciones y la primera mitad cronológicamente. Realizamos estas comparaciones en la *Sección 3.1* para las localizaciones del lobo y para las ubicaciones de las cabras en la *Sección 3.2*.

### 3.1. Ranking de *home ranges* para las ubicaciones del lobo en Mongolia.

En la *Figura 3.1*, se representan la muestra completa de los datos del lobo frente a la primera mitad cronológica, en la cual tenemos 728 datos de ubicaciones. A simple vista,

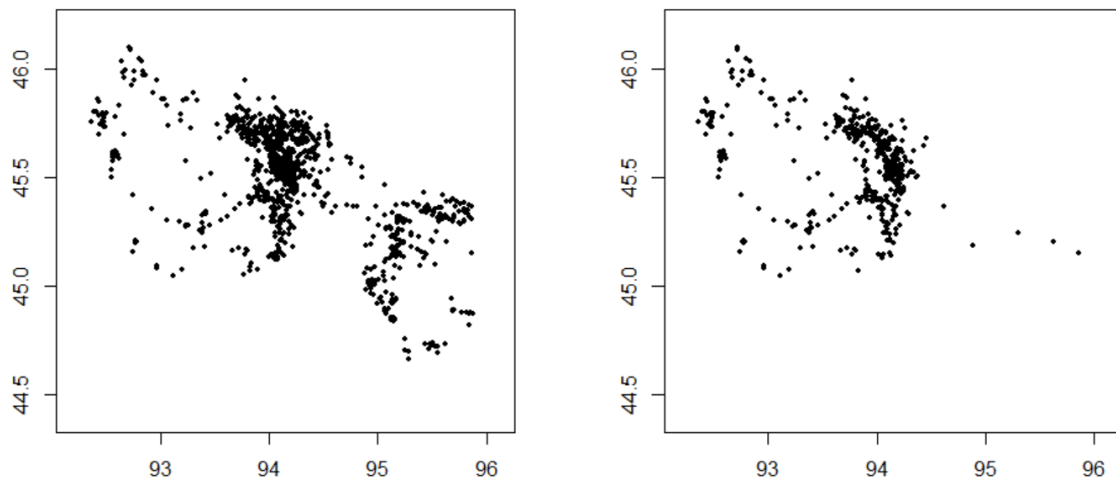


Figura 3.1: Muestra completa de datos de ubicaciones del lobo frente a la primera mitad cronológica de la muestra.

al escoger solamente la primera mitad de los datos, tenemos mucha diferencia entre esta muestra y la original, ya que observamos en la segunda imagen que prácticamente desaparecen todas las ubicaciones situadas en el sureste del mapa, lo que nos puede hacer pensar que por ejemplo, con el paso del tiempo, los recursos naturales de esa zona comenzaron a escasear por lo que el lobo pudo haber tenido que trasladarse. Puede verse a simple vista que los estimadores del *home range* de estos conjuntos de datos diferirán en área mucho el uno del otro. De todas formas, comprobamos esto más detenidamente aplicando los métodos de estimación del soporte vistos hasta ahora sin imponer e imponiendo restricciones de forma.

### Devroye-Wise:

Aplicamos este estimador sin restricciones de forma a ambos conjuntos de datos tomando dos radios distintos. En la *Figura 3.2*, presentamos en las dos primeras imágenes este estimador con radio  $\epsilon = 0,03$  para la muestra total y la primera mitad cronológica, respectivamente, y con  $\epsilon = 0,08$  para las dos últimas, considerando la muestra total en la primera de ellas y la primera mitad cronológica en la última.

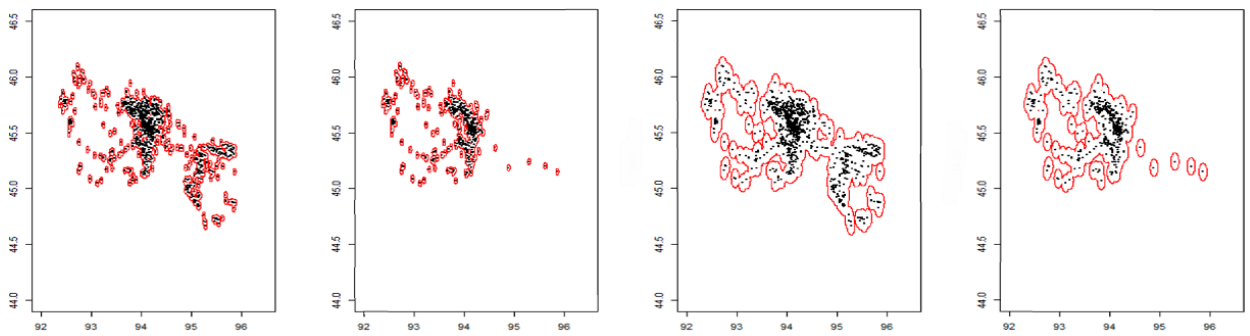


Figura 3.2: Estimador de *Devroye-Wise* con radios  $\epsilon = 0,03$  en las dos primeras imágenes para la muestra total y la primera mitad, respectivamente, y con  $\epsilon = 0,08$  para la muestra total y la mitad cronológica en las dos últimas.

Como era de esperar, puede apreciarse a simple vista que el área de este estimador aplicado a la muestra total es considerablemente mayor con respecto a la primera mitad.

### KDE:

De nuevo sin tomar restricciones de forma en el soporte, aplicamos el estimador KDE a ambos conjuntos de datos. En la *Figura 3.3*, representamos estos estimadores tomando contenidos de probabilidad del 95 %, en las dos primeras imágenes, para la muestra total y la primera mitad cronológica, respectivamente, y con contenido en probabilidad del 80 % en las dos últimas.

Nuevamente, vemos que con contenido en probabilidad del 95 % el área se reduce prácticamente a la mitad, sin embargo, con contenido de 80 %, aún habiendo diferencia, no es tan notable.

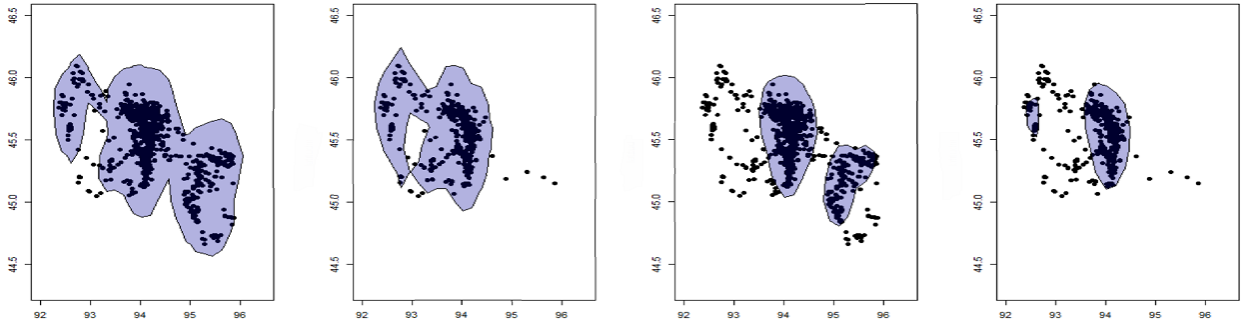


Figura 3.3: Estimador KDE con contenidos en probabilidad de 95% en las dos primeras imágenes para la muestra total y la primera mitad cronológica, respectivamente, y 80% en las dos últimas.

### Envoltura convexa:

Suponiendo que el soporte es convexo, presentamos en la *Figura 3.4* las envolturas convexas para la muestra total (izquierda) y la primera mitad cronológica (derecha).

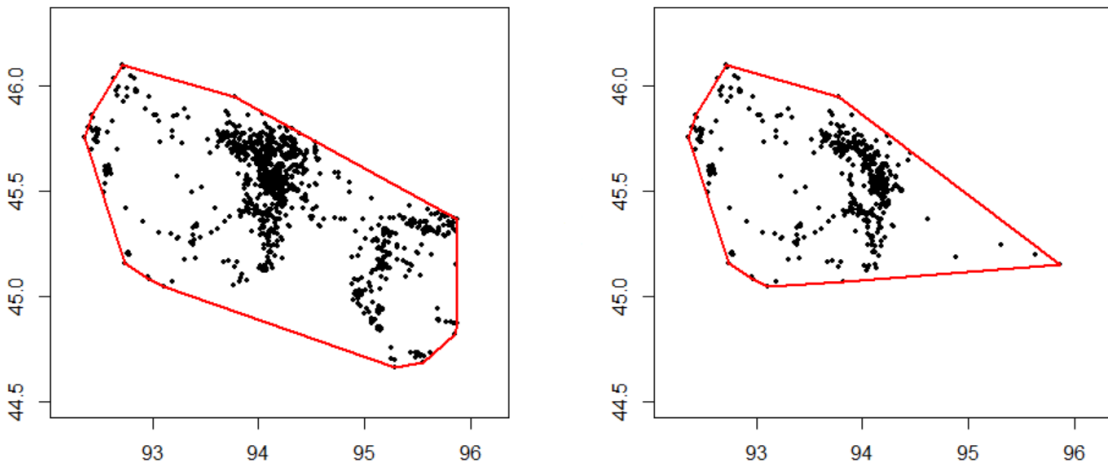


Figura 3.4: Envoltura convexa para la muestra total (izquierda) y para la primera mitad cronológica (derecha).

Ahora, el paquete `geometry`<sup>1</sup> de R nos permite calcular el área de estas envolturas. Obtenemos para la muestra total un área de  $8,18 u^2$  y para la primera mitad,  $7,65 u^2$ , casi una unidad menos.

<sup>1</sup><https://cran.r-project.org/web/packages/geometry/geometry.pdf>

### Envoltura $r$ -convexa:

Suponiendo que el soporte del *home range* es  $r$ -convexo, aplicamos el método propuesto en [Rodríguez-Casal y Saavedra-Nieves \(2019\)](#) para seleccionar el  $r$  óptimo de forma automática.

Para la muestra total, tenemos  $\hat{r} = 0,43$  y para la primera mitad cronológica,  $\hat{r} = 0,64$ . En la [Figura 3.5](#) se muestra la envoltura  $r$ -convexa con radio estimado para la muestra total (izquierda) y para la mitad de la muestra (derecha).

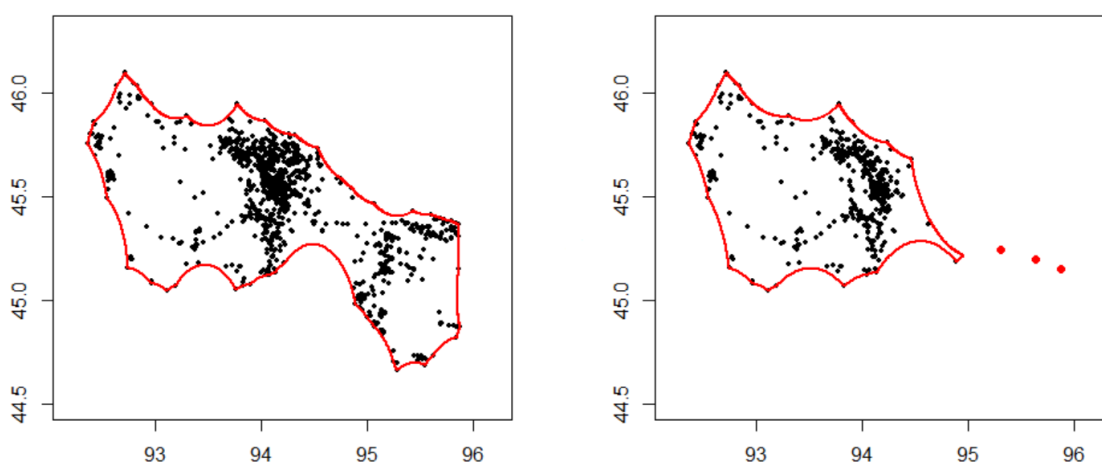


Figura 3.5: Envoltura  $r$ -convexa para la muestra total con radio óptimo  $\hat{r} = 0,43$  (izquierda) y para la primera mitad con radio óptimo  $\hat{r} = 0,64$  (derecha).

Con el paquete `alphahull`<sup>2</sup> de R calculamos el área de estas envolturas, obteniendo para la muestra total un área de  $2,25 u^2$  y para la primera mitad cronológica,  $1,48 u^2$ . El área se reduce a casi la mitad al pasar de una muestra a otra.

## 3.2. Ranking de *home ranges* para las ubicaciones de las cabras salvajes en Hawaii.

Realizamos este test de comparación de áreas a los datos de las cabras en Hawaii pensando en obtener mejores resultados, ya que esta muestra tiene un número mayor de ubicaciones, representa a más de un individuo, y además, los individuos fueron seleccionados en grupos distintos para obtener una muestra más representativa.

<sup>2</sup><https://cran.r-project.org/web/packages/alphahull/alphahull.pdf>

En la *Figura 3.6* presentamos, en la imagen de la izquierda, las ubicaciones de la muestra total frente a las ubicaciones de la primera mitad cronológica, a la derecha, la cual consta de 1966 datos. Observamos que en este caso, al contrario que con el lobo, las dos

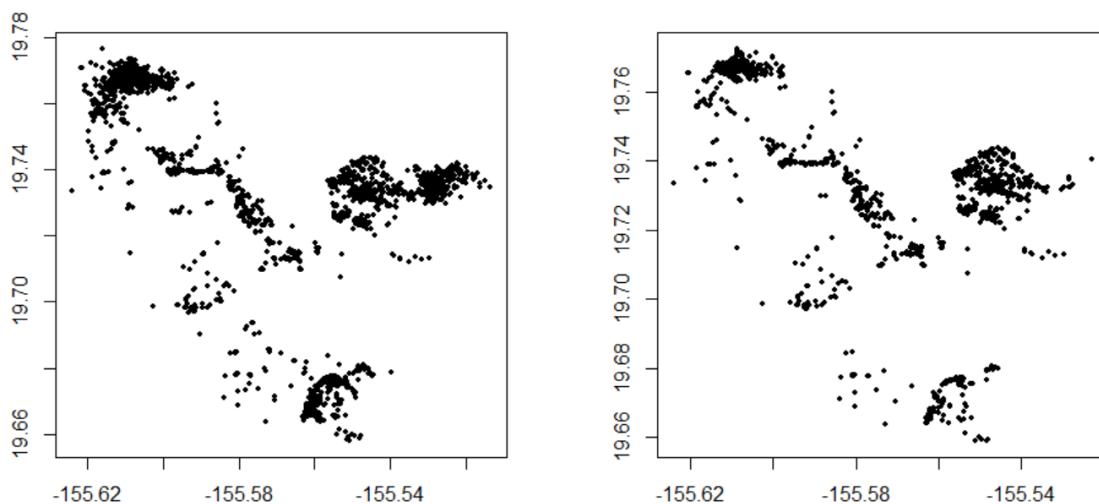


Figura 3.6: Muestra con tres individuos de datos de ubicaciones de las cabras frente a la primera mitad cronológica de la muestra.

muestras presentan muchas más similitudes, por lo que puede ser interesante comparar los estimadores del *home range* de ambas. Estimaremos el *home range* asumiendo y sin asumir restricciones de forma.

### Devroye-Wise:

Sin imponer restricciones de forma, comparamos los estimadores tomando dos radios distintos. En la *Figura 3.7*, presentamos estas dos comparaciones, siendo las dos primeras imágenes el estimador tomando como radio  $\epsilon = 0,004$  para la muestra total y la primera mitad cronológica, respectivamente; y, de forma análoga, las dos últimas tomando como radio  $\epsilon = 0,01$ .

Los resultados obtenidos muestran que se conserva la forma esencialmente en los dos casos, y las áreas de los estimadores parecen similares para ambos radios, incluso, para el radio más pequeño, obtenemos en ambas muestras las mismas componentes conexas, cinco.

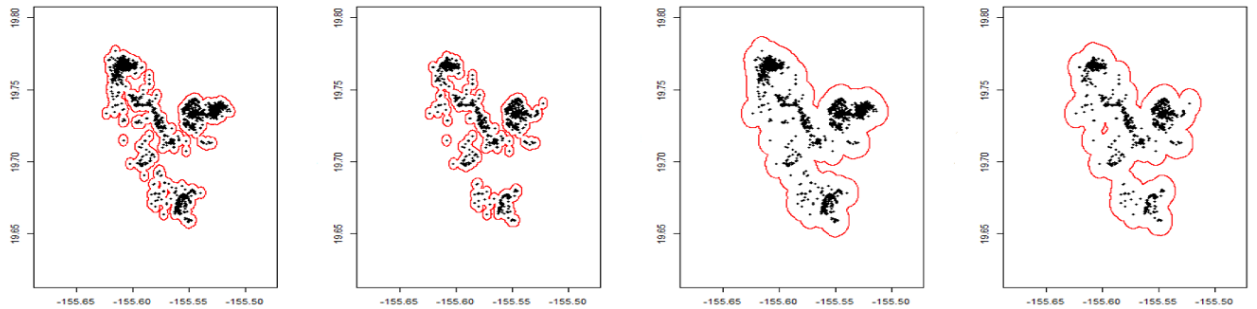


Figura 3.7: Estimadores de *Devroye-Wise* comparados con radios  $\epsilon = 0,004$  en las dos primeras imágenes para la muestra total y la mitad cronológica, respectivamente, y con  $\epsilon = 0,01$  en las dos últimas.

### KDE:

Sin imponer restricciones de forma al soporte del *home range*, vamos a comparar ambos estimadores KDE con un contenido en probabilidad del 95 %, 90 % y 75 % en las dos muestras. En la *Figura 3.8*, mostramos los estimadores con contenido en probabilidad del 95 % en las dos primeras imágenes para la muestra total y la primera mitad cronológica, respectivamente, y con contenido en probabilidad del 90 % en las dos últimas.

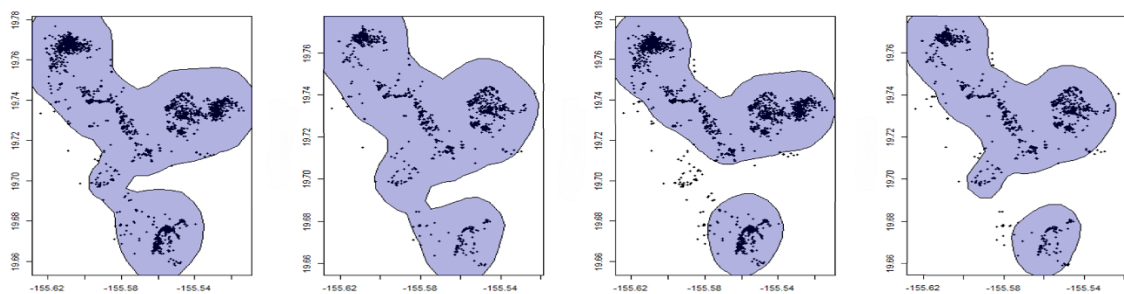


Figura 3.8: Estimadores KDE con contenido en probabilidad del 95 % en las dos primeras imágenes sobre la muestra total y la primera mitad cronológica y con contenido en probabilidad del 90 % en las dos últimas.

Observamos que ambos estimadores son relativamente parecidos, ya que mantienen en ambas probabilidades de contenido las mismas componentes conexas y podemos ver a simple vista que presentan áreas similares. En la *Figura 3.9*, mostramos los resultados de tomar un 75 % de contenido de probabilidad para la muestra total (izquierda) y la primera

mitad cronológica (derecha). En este caso, respecto a las áreas, vemos a simple vista que son similares, aunque el estimador para la muestra total presente más componentes con-

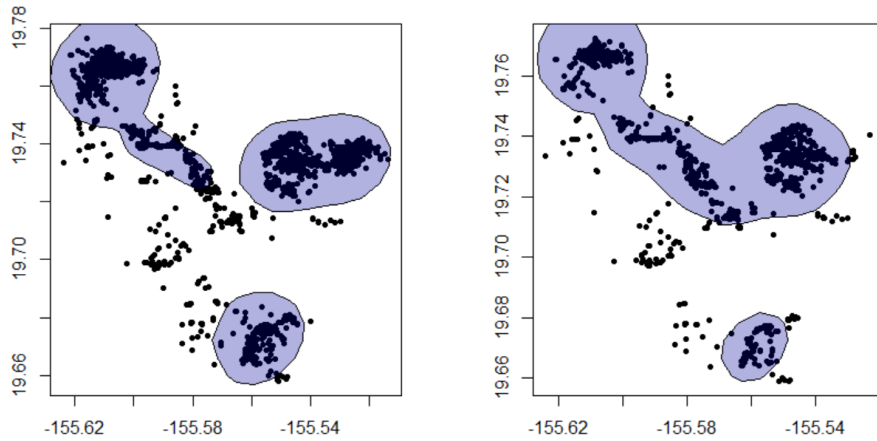


Figura 3.9: Estimadores KDE con la muestra completa (izquierda) y la primera mitad (derecha) con un contenido en probabilidad del 75 %.

exas que el obtenido para la mitad de los datos.

### Envoltura convexa:

En la *Figura 3.10*, mostramos, imponiendo la restricción de convexidad al soporte, los estimadores resultantes para ambas muestras, la total (izquierda), y la primera mitad cronológica (derecha).

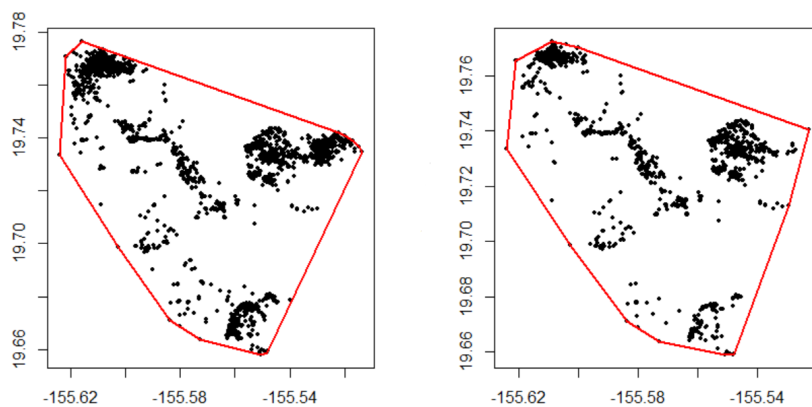


Figura 3.10: Envoltura convexa para la muestra con todos los datos (izquierda) y para la primera mitad cronológica (derecha).

nológica (derecha), que resultan prácticamente iguales en forma. Empleando de nuevo el paquete `geometry`<sup>3</sup> de R obtenemos un área de  $0,35 u^2$  para la envoltura convexa de la muestra total y de  $0,33 u^2$  con la primera mitad cronológica de esta. Por tanto, ambas envolturas son prácticamente iguales.

### Envoltura $r$ -convexa:

Suponiendo que el soporte del *home range* es  $r$ -convexo, aplicamos el método propuesto en [Rodríguez-Casal y Saavedra-Nieves \(2019\)](#) para calcular  $r$  de forma automática.

Presentamos en la [Figura 3.11](#) ambas envolturas  $r$ -convexas aplicadas a nuestros dos conjuntos de datos, la muestra total y la primera mitad cronológica, respectivamente. El  $r$  óptimo obtenido para la muestra total era  $\hat{r} = 0,0092$ . Calculando este radio para la primera mitad cronológica de ubicaciones, obteniendo  $\hat{r} = 0,041$ .

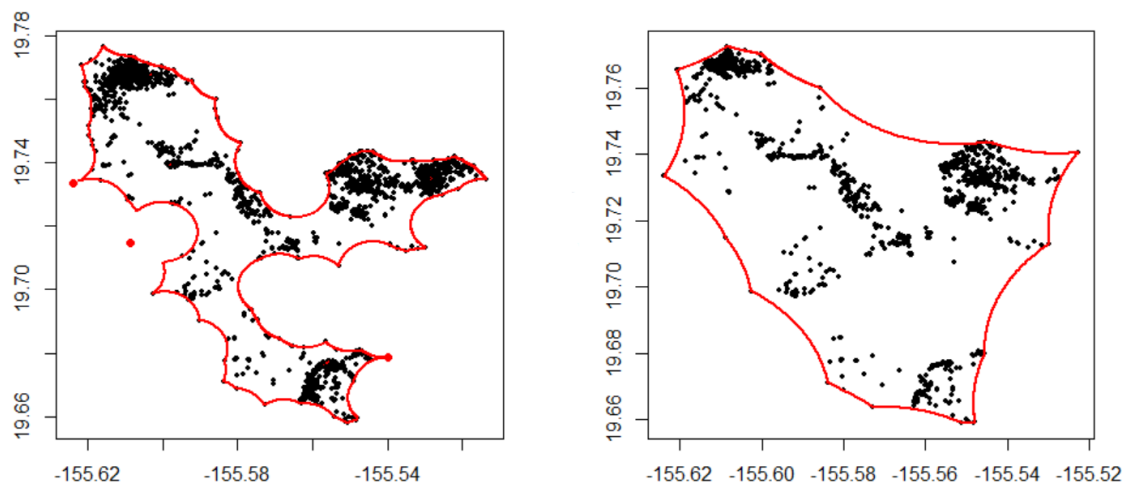


Figura 3.11: Envoltura  $r$ -convexa para la muestra con todos los datos (izquierda) y para la primera mitad cronológica (derecha) con  $r$  óptimo.

Con estos resultados podemos observar diferencias entre ambos estimadores, aunque los dos constan de una sola componente conexa diferenciada (obviamos los puntos aislados del primer caso), notamos que con la mitad de la muestra (derecha) el estimador presenta mucho más parecido a la envoltura convexa, mientras que en el primer caso obtenemos un estimador con forma más similar a los obtenidos con *Devroye-Wise* o KDE, es decir, sin imponer condiciones de forma al *home range*.

<sup>3</sup><https://cran.r-project.org/web/packages/geometry/geometry.pdf>

Viendo estas diferencias, calculamos nuevamente sus áreas empleando el paquete `alphahull`<sup>4</sup> de R, obteniendo para la muestra total un área de  $0,00402 u^2$  y para la primera mitad cronológica,  $0,0061 u^2$ . Notamos que en este caso, a diferencia de la envoltura convexa, ya no logramos áreas tan similares entre ambos estimadores, siendo esta vez el conseguido con la primera mitad de ubicaciones mayor. Una de las posibles causas de los resultados obtenidos es que el tamaño de la muestra más pequeña no sea suficiente.

---

<sup>4</sup><https://cran.r-project.org/web/packages/alphahull/alphahull.pdf>

## Capítulo 4

# Conclusiones y discusión

En este trabajo, tras introducir ciertas propiedades geométricas sencillas, nuestro objetivo fue revisar los estimadores clásicos del soporte y aplicarlos para reconstruir el *home range*. Para visualizar los distintos estimadores, se emplearon dos conjuntos de datos reales, de lobos y cabras salvajes, haciendo uso del software libre R para ello.

Entre los estimadores del soporte, se han distinguido dos casos:

1. El soporte no satisface ninguna condición de forma.
2. El soporte verifica alguna propiedad geométrica concreta.

En el primer caso, el estimador muestral claramente presenta la desventaja de infraestimar el *home range*, ya que solamente tiene en cuenta los propios datos de ubicaciones de los que disponemos, problema que surge también en el estimador de *Devroye-Wise* cuando escogemos un radio  $\epsilon$  demasiado pequeño, al contrario de lo que ocurre si tomamos este radio muy grande, entrando así en riesgo de sobreestimar el *home range*. Esto no ocurre en el estimador KDE, en el cual presentamos el *home range* como un conjunto de nivel y escogemos un contenido en probabilidad determinado. Si asumimos que el soporte es convexo, la envoltura convexa de los puntos de la muestra es un estimador natural, cuyo inconveniente es estimar solamente una componente conexas, por lo que puede contener lugares de imposible acceso para los animales. Suavizando esta imposición, hemos asumido que el soporte es  $r$ -convexo para un cierto  $r > 0$  desconocido que ha sido estimado a partir de la muestra de forma automática. Para los dos conjuntos de datos considerados, este estimador proporciona reconstrucciones del *home range* bastante realistas.

En el caso de las ubicaciones del lobo, además de representar los estimadores estudiados, pensando que el *home range* de este animal podría tener más de una componente conexas, realizamos una modificación a los datos quedándonos con la muestra efectiva que pertenece

al conjunto de nivel con contenido en probabilidad del 99 %, eliminando así las ubicaciones no efectivas o no sustanciales por si estas condicionan demasiado la forma de nuestro estimador. Con la consideración de la muestra efectiva y suponiendo con esta nueva muestra que el soporte es  $r$ -convexo, obtenemos un nuevo estimador con dos componentes conexas, lo cual nos indica, efectivamente, que es posible que el *home range* no sea conexo, sino que existan, por lo menos, dos componentes distintas.

Para las ubicaciones de las cabras salvajes, realizamos un estudio sobre la gran influencia que puede llegar a tener un solo individuo sobre el estimador del *home range* viendo la envoltura convexa de cada individuo dentro de la muestra y los cambios drásticos que se producen con la eliminación de uno de ellos.

Finalmente, para medir o evaluar la calidad de nuestros estimadores, aplicamos el criterio propuesto en *Fleming et al. (2015)* de comparación de áreas. En la *Tabla 4.1*, se muestran los resultados obtenidos para la envoltura convexa y  $r$ -convexa (con  $r$  estimado):

	Lobo		Cabras	
	Total (en $u^2$ )	Mitad (en $u^2$ )	Total (en $u^2$ )	Mitad (en $u^2$ )
Envoltura convexa	8.18	7.65	0.35	0.33
Envoltura $r$ -convexa	2.25	1.48	0.00402	0.0061

Tabla 4.1: Áreas de las envolturas convexas y  $r$ -convexas de la muestra total y la mitad de la muestra para los dos conjuntos de ubicaciones.

Para los datos del lobo, las áreas resultantes no presentan similitudes entre ellas, ya que la muestra total difiere mucho sobre la mitad de esta, por lo que puede apreciarse a simple vista que las áreas de los estimadores serán distintas. La diferencia entre la muestra total y la mitad puede deberse a que estos datos de ubicaciones contienen únicamente información de un solo individuo que, a lo largo del tiempo, pudo haber tenido la necesidad de trasladarse, por ejemplo, por escasez de recursos naturales a otras zonas. Al contrario, con los datos de localizaciones de las cabras, estos dos conjuntos de ubicaciones presentan más similitudes, ya que en este caso disponemos de ubicaciones de más de un individuo y la muestra fue tomada escogiendo cabras de distintos grupos para que fuese representativa. Por tanto, para el conjunto de datos de las cabras salvajes, las áreas de los estimadores presentados son prácticamente coincidentes para ambos conjuntos, lo que indica el buen comportamiento de los estimadores. Aunque los valores son bastante parecidos, el hecho de que la envoltura  $r$ -convexa presente un área mayor para la mitad de la muestra que para el total, puede deberse al bajo tamaño de muestra considerado.

Finalmente, es importante destacar que los estimadores del *home range* que hemos analizado asumen de forma poco realista que las observaciones son independientes. Aunque existen en la literatura alternativas que solventan esta restricción, no han sido consideradas en este trabajo donde el principal objetivo era abordar la reconstrucción del *home range* a partir de la teoría de la estimación no paramétrica del soporte.



# Bibliografía

- [1] Baíllo, A. y Chacón, J. E. (2018), *A survey and a new selection criterion for statistical home range estimation*. arXiv preprint arXiv:1804.05129.
- [2] Baíllo, A. y Cuevas, A. (2001). *On the estimation of a star-shaped set*. Advances in Applied Probability, 33, 717–726.
- [3] Burt, W.H. (1943). *Territoriality and home range concepts as applied to mammals*. Journal of Mammalogy, 24, 346-352.
- [4] Calenge, C. (2006). *The package “adehabitat” for the R software: a tool for the analysis of space and habitat use by animals*. Ecological modelling, 197(3-4), 516-519.
- [5] Calenge, C. Package “adehabitatHR”. *Home Range Estimation*.
- [6] Cuevas, A. (2010), *Set estimation: Another bridge between statistics and geometry*. Boletín de Estadística e Investigación Operativa, 25, 71-85.
- [7] Cuevas, A., Fraiman, R. y Pateiro-López, B. (2012). *On statistical properties of sets fulfilling rolling-type conditions*. Advances in Applied Probability, 44, 311–329.
- [8] Cuevas, A. y Fraiman, R. (1997). *A plug-in approach to support estimation*. Annals of statistics, 25, 2300-2312.
- [9] Devroye, L. y Wise, G. L. (1980). *Detection of abnormal behavior via nonparametric estimation of the support*. SIAM Journal on Applied Mathematics, 38, 480–488.
- [10] Dobrow, R. (1992). *Estimating level sets of densities*. *Unpublished manuscript*.
- [11] Dümbgen, L. y Walther, G. (1996). *Rates of convergence for random approximations of convex sets*. Advances in Applied Probability, 28, 384–393.
- [12] Edelsbrunner, H., Kirkpatrick, D. G. y Seidel, R. (1983). *On the shape of a set of points in the plane*. Information Theory, IEEE Transactions on, 29, 551–559.

- [13] Fleming, C.H., Fagan, W.F., Mueller, T., Olson, K.A., Leimgruber, P. y Calabrese, J.M. (2015). *Rigorous home range estimation with movement data: a new autocorrelated kernel density estimator*. *Ecology*, 96, 1182-1188.
- [14] Geffroy, J. (1964). *Sur un problème d'estimation géométrique*. Publications de l'Institut de statistique de l'Université de Paris, 13, 191–210.
- [15] Jean-Romain Roussel, Barber C. B., Kai Habel, Raoul Grasman, Robert B. Gramacy, Pavlo Mozharovskiy y David C. Sterratt (2190). Package “geometry”. *Mesh Generation and Surface Tessellation*.
- [16] Kenward, R.E., Clarke, R.T., Hodder, K.H. y Walls, S.S. (2001). *Density and linkage estimators of home range: nearest-neighbor clustering of multinuclear cores*. *Ecology*, 82, 1905-1920.
- [17] Korostelëv, A. y Tsybakov, A. (1993). *Minimax Theory of Image Reconstruction*. Lecture Notes in Statistics, 82. Springer-Verlag, New York.
- [18] Long, J. y Nelson, T. (2015). *Home range and habitat analysis using dynamic time geography*. *The Journal of Wildlife Management*, 79, 481-490.
- [19] Movebank for animal tracking data. Wolves Mongolia 2003.2005 [n=2] from [www.movebank.org](http://www.movebank.org)
- [20] Movebank for animal tracking data. Non-native feral goats, Chynoweth, Hawaii Island [n=13] from [www.movebank.org](http://www.movebank.org)
- [21] Pateiro-López, B. y Rodríguez-Casal, A. (2010). *Generalization of the convex hull of a sample: The R package alphahull*. *Journal of Statistical Software*, 34, 1-28.
- [22] Pateiro-López, B. y Rodríguez-Casal, A. (2019). The R package “alphahull”. *Generalization of the Convex Hull of a Sample of Points in the Plane*.
- [23] Renka, R. J., Gebhardt, A., Eglén, S., Zuyev, S. y White, D. (2019). Package “tripack”. *Triangulation of Irregularly Space Data*.
- [24] Rényi, A. y Sulanke, R. (1963). *Über die konvexe Hülle von n zufällig gewählten Punkten*. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 2, 75–84.
- [25] Rényi, A. y Sulanke, R. (1964). *Über die konvexe Hülle von n zufällig gewählten Punkten. II*. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 3, 138–147.

- [26] Rodríguez-Casal, A. y Saavedra-Nieves, P. (2016). *A fully data-driven method for estimating the shape of a point cloud* *ESAIM: Probability and Statistics*. Vol. 20, pp. 332-348.
- [27] Rodríguez-Casal, A. y Saavedra-Nieves, P. (2019). *Extent of occurrence reconstruction using a new data-driven support estimator*. arXiv preprint arXiv:1907.08627.
- [28] Saavedra-Nieves, P. (2014). *Nonparametric data-driven methods for set estimation*. Thesis.
- [29] Seton, E.T. (1909). *Life-histories of Northern Animals: an Account of the Mammals of Manitoba*. Vol.I-Grass-Eaters. Charles Scribner's Sons.
- [30] Steiniger, S. y Hunter, A.J.S. (2013). *A scaled line-based density estimator for the retrieval of utilization distributions and home ranges from GPS movement tracks*. *Ecological Informatics*, 13, 1-8.
- [31] Walther, G. (1997). *Granulometric smoothing*. *Annals of Statistics*. 25, 2273–2299.
- [32] Walther, G. (1999). *On a generalization of Blaschke's rolling theorem and the smoothing of surfaces*. *Mathematical Methods in the Applied Sciences*, 22, 301–316.
- [33] Wand, M. y Jones, M. (1995). *Kernel Smoothing*. Chapman and Hall.