



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS MATEMÁTICAS Y  
DE LA ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA

CAMINATAS ALEATORIAS CON REUBICACIONES PREFERENCIALES EN  
ESPACIOS HETEROGÉNEOS COMO MODELO PARA EL FORRAJE ANIMAL.

TESIS  
QUE PARA OPTAR POR EL GRADO DE:  
DOCTORA EN CIENCIAS

PRESENTA:  
JULIA ANDREA CATALINA FALCÓN CORTÉS

DIRECTOR DE LA TESIS  
DR. DENIS PIERRE BOYER  
INSTITUTO DE FÍSICA - UNAM

CIUDAD DE MÉXICO, 21 DE OCTUBRE DE 2019.



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



*A mi mamá, por mostrarme el camino correcto y por abrazarme con tanto amor a cada nuevo encuentro.*

*A mi papá, por enseñarme a soñar y por ser esa constante en mi vida que sé jamás desaparecerá.*

*A mi hermano, por quererme sin condiciones, por siempre estar y por ser mi maestro de bondad y nobleza.*

*Y a mi amado Andrés, por elegir acompañarme en este viaje y por elegir amar cada parte de mí.*



# Índice general

Agradecimientos . . . . .	4
<b>1. Introducción</b>	<b>9</b>
<b>2. Antecedentes</b>	<b>15</b>
2.1. Caminatas aleatorias reforzadas en medios homogéneos . . . . .	15
2.1.1. Caminatas aleatorias reforzadas . . . . .	16
2.1.2. Difusión anómala . . . . .	16
2.1.3. Caminatas aleatorias con reinicio estocástico . . . . .	17
2.1.4. Caminatas aleatorias preferenciales . . . . .	19
2.2. Movimiento Colectivo . . . . .	20
2.2.1. Estrategia de forrajeo grupal . . . . .	20
2.2.2. La comunicación mejora el forrajeo grupal . . . . .	22
2.2.3. Modelo basado en agentes para el forrajeo colectivo . . . . .	24
2.3. Análisis de trayectorias animales reales . . . . .	26
2.3.1. Identificando dos tipos de movimientos en ciervos . . . . .	26
2.3.2. Dos estados de movimiento durante la migración . . . . .	28
2.3.3. Modelo poblacional para herbívoros del Serengeti . . . . .	30
2.4. Objetivos Generales . . . . .	32
2.5. Objetivos Particulares . . . . .	33
<b>3. Modelo Mínimo</b>	<b>35</b>
3.1. Modelo . . . . .	35
3.2. Aproximación de decorrelación . . . . .	38
3.3. Dinámica a primeros vecinos . . . . .	41
3.4. Dinámica con vuelos de Lévy . . . . .	44
3.5. Comportamiento de la distribución $P_{\mathbf{n}}$ a $ \mathbf{n} $ grandes . . . . .	45
3.6. Exponentes críticos . . . . .	48
3.6.1. Parámetro de orden. . . . .	48
3.6.2. Longitud de localización. . . . .	50
3.7. Modelo mínimo en $2d$ y $3d$ . . . . .	51
3.8. Una simple modificación al modelo mínimo. . . . .	54
3.8.1. Expresión analítica en $1d$ . . . . .	55
3.8.2. Análisis para el caso $2d$ y $3d$ . . . . .	59

<b>4. Modelo Colectivo</b>	<b>63</b>
4.1. Modelo . . . . .	64
4.1.1. Ambiente . . . . .	64
4.1.2. Caminantes . . . . .	65
4.1.3. Otras redes de interacción . . . . .	65
4.2. Resultados . . . . .	66
4.2.1. Localización selectiva . . . . .	66
4.2.2. Cohesión del enjambre . . . . .	70
4.2.3. Efecto de las redes de interacción . . . . .	70
4.3. Vuelos de Lévy . . . . .	72
4.4. Decaimiento de Memoria . . . . .	76
<b>5. Modelos Jerárquicos</b>	<b>83</b>
5.1. Función de verosimilitud . . . . .	84
5.2. Modelos Bayesianos . . . . .	85
5.3. Modelos Jerárquicos . . . . .	89
5.4. Monte Carlo de Cadenas de Markov . . . . .	92
5.5. Algoritmo de dos pasos . . . . .	97
5.6. Comprobación Predictiva Posterior . . . . .	101
5.7. Criterio de Información de Akaike y de Watanabe-Akaike . . . . .	103
5.8. Los datos . . . . .	106
5.9. Modelos . . . . .	108
5.9.1. Modelo I . . . . .	108
5.9.2. Modelo II . . . . .	109
5.9.3. Modelo III . . . . .	109
5.9.4. Modelo IV . . . . .	110
5.10. Distribuciones Previas . . . . .	111
5.11. Parámetros determinados . . . . .	113
5.12. Comparación entre modelos . . . . .	124
<b>6. Conclusiones</b>	<b>129</b>
<b>7. Apéndices</b>	<b>137</b>
7.1. Nivel individual . . . . .	137
7.2. Nivel poblacional . . . . .	164
<b>Bibliografía</b>	<b>189</b>

# Agradecimientos

Como escribiera alguna vez aquél célebre filósofo romano Lucio Anneo Séneca: “Nada es más honorable que un corazón agradecido”. Pues bien, mi corazón está lleno de agradecimiento. Y aunque no por ello me siento más honorable, sí me siento profundamente necesitada de expresar mi gratitud. Hay tantas personas a las que me gustaría agradecer pues todas aquéllas que han pasado por mi vida me han llevado a donde estoy ahora, pero me gustaría sólo dedicar mis palabras a esas personas que ocupan los lugares más profundos de este agradecido corazón.

Obviamente el inicio de esta lista la ocupa mi amada y pequeña familia. Mamá, papá, no me alcanzará la vida para agradecerles todo lo que me han dado, todo lo que sé se han sacrificado para poner en mi camino oportunidades mucho mejores que las que ustedes tuvieron. Los días que me restan no serán suficientes para demostrarles lo mucho que los amo, lo mucho que los admiro y lo mucho, pero en serio mucho, que les agradezco cada una de las decisiones que tomaron para mejorar mi vida. Sin ustedes, sin su ejemplo, sin sus manos tomando las mías y sin su incondicional apoyo, no sé que sería de mí pero sí sé que no sería quien soy ahora.

Mamá, tú me has hecho la mujer que ahora soy. Gracias por el hermoso ejemplo de responsabilidad, rectitud, congruencia y moral inquebrantable que me has dado durante toda mi vida, ese ejemplo me enseñó que seguir tus pasos sería la mejor decisión que podría tomar. Gracias porque inculcaste en mí la disciplina y la tenacidad suficientes para siempre perseguir mis sueños. Gracias por permitirme tener en mi vida mujeres tan maravillosas como mis tías y mi abuelita, gracias por que por ti ahora puedo tener el gozo de decir que la vida me regaló cuatro hermosas madres que me marcaron para siempre. A ellas, que ya no están a nuestro lado, les agradezco a través de ti porque soy tan hija de ellas como tuya. ¡Gracias por elegir ser la maravillosa madre que tu también tuviste!

Papá, mi pilar inquebrantable, ni siquiera encuentro en mi pobre vocabulario las palabras para decirte lo que siento al escribirte estas líneas. Papá, eres el hombre más extraordinario que conozco, el más perseverante, el más terco, el más tenaz, el más caballeroso, el más amoroso... eres el papá que cualquier niña quisiera tener. Desde siempre me has alentado para perseguir todos mis sueños, hasta los más alocados de ellos. Gracias por convencerme que soy capaz de lograr lo que me proponga. Gracias por hacerme entender que ser mujer jamás debe de ser un impedimento y, que no importa lo que la sociedad diga, puedo jugar al fútbol si me lo propongo, puedo ser la mejor basquetbolista si así lo quiero, y puedo ser una científica exitosa si ese es mi sueño. No sabes cuánto te agradezco todo lo que has hecho por mí, por

mi mamá, por mi hermano. Sé lo mucho que has sacrificado pero quiero que sepas que, con ese sacrificio, me diste una de las mejores lecciones de mi vida: nada es más importante que la familia, de nada sirve tener dinero, éxitos y la aceptación de los otros, si no tenemos a nuestra familia con nosotros. Papi, quiero que seas eterno y me acompañes por el resto de mi vida. ¡Gracias, papi por elegir ser el papá que te hubiera gustado tener!, ese mismo padre que encontraste en mi abuelito Domi y que yo también hallé.

Chuy, m'hijo, mi hermanito... gracias por tu alegre presencia en mi vida. Ese 6 de julio en el que naciste ha sido uno de los días más felices de mi corta existencia, ¡y no era para menos!, ese día nació mi compañero de vida, ese ser humano con el que sé que siempre podré contar, uno de los hombres incondicionales de mi vida, mi gran maestro de nobleza y amor. Mi pequeño hermano (y es que siempre serás pequeño para mí)... creo que jamás te he dicho lo mucho que me costó irme de tu lado, el precio tan alto que pagué y sigo pagando (y que pagaré por el resto de mi vida) por la elección de perseguir mis sueños. Dejarte atrás ha sido una de las cosas más difíciles que he hecho, sabía cuánto me necesitabas pero también sabía cuanto anhelaba volar. Te confieso que sigo sin perdonarme el haberme ido. En eso has sido también un gran maestro de perdón. Gracias por tus sonrisas, por tus bromas (posees uno de los mejores sentidos del humor que he conocido), gracias por entenderme, por apoyarme, y por elegir ser ese hermano que te hubiera gustado tener.

Andrés, mi amado Andrés. Gracias por todo; por tu apoyo incondicional, por creer en mi siempre (incluso en aquellos momentos en los que yo no podía creer en mi misma), gracias por elegirme cada día. Gracias por acompañarme en esta aventura llamada doctorado, gracias por ayudarme a atravesar las aguas más peligrosas, por ayudarme a escalar las montañas más engañosas y por ayudarme a abrir mis alas y usarlas para lo que son... para volar. Gracias por aprender junto conmigo el verdadero significado del amor, ese amor que libera, que enaltece, que enorgullece, que suma alegría a nuestras vidas. Bien sabes que no te necesito en mi vida, así como tú no me necesitas en la tuya, pero ¡cuánto me encanta compartir mis días contigo! Mucho de lo que escribí en esta tesis se logró por compartir contigo las ideas, los avances y los problemas que surgieron en estos cuatro años de doctorado.

A todas esas personas que me han acompañado en estos años ¡mil gracias!. Gracias a mi suegra, a mis cuñadas, a mi cuñado y a mi sobrino por arroparme cuando me sentía extraña en una ciudad que no era la mía, lejos de mi familia y de todo lo que me era conocido. Sin ese apoyo yo no hubiera podido atravesar estos seis años de posgrado. Gracias a mis amigos Dani Cervantes y Sabri Andrade. Al primero por siempre estar dispuesto a ayudarme, por acompañarme en esa semana antes de irme a Bariloche, por su apoyo incondicional y sus abrazos siempre francos. A la segunda por ser ese rayito de luz en Argentina cuando todo parecía oscuro y sin sentido. Gracias a los dos por que, quizá sin saberlo, inyectaron ánimo a mi vida en los momentos complicados.

Gracias a la UNAM, en ella (la más grande casa de estudios de mi México) he conocido a grandes seres humanos y excelentes científicos que me han guiado en esta travesía. Gracias al Dr. Pedro Miramontes por su amistad, por su interés en mi y mi vida académica, por siempre estar dispuesto a recibirme en su cubículo para platicar. Gracias al Dr. Denis Boyer por su

excelente tutela en estos cuatro años de doctorado, por sus invaluable consejos académicos y por siempre tener tiempo para discutir mi proyecto de investigación, por abrirme las puertas al mundo permitiéndome exponer mi trabajo en diversos congresos y escuelas internacionales. Gracias por creer en mi y mostrarme el significado de la investigación científica. Gracias al Dr. Juan Manuel Morales por su cotutela en Bariloche, Argentina. Gracias porque con su visión me ayudó a darle un nuevo aire a mi doctorado y juntar las fuerzas y el ánimo que me hacían falta para terminar este proyecto. Gracias a los Drs. Francois Leyvraz, Gerónimo Uribe, Hernán Larralde y Gabriel Ramos por sus certeras aportaciones y sus constantes evaluaciones a mi trabajo. Abonaron de forma invaluable a que este proyecto fuera tan fructífero. Al Dr. Faustino Sánchez por enseñarme ese hermoso arte que significa impartir clases y transmitir conocimiento.

Gracias a CONACyT por el apoyo económico que me permitió realizar estos posgrados (maestría y doctorado) y que me permitió realizar actividades académicas fuera de mi país. Aquí se hace oportuno agradecer también a la UNAM, pues con su financiamiento (por medio de los apoyos de DGAPA y PAPIIT) pude viajar al extranjero a diversos congresos, cursos y escuelas, tanto nacionales como internacionales.

Gracias a los editores y árbitros de las revistas internacionales que evaluaron y aceptaron los artículos de los que se extiende esta tesis. En total hemos logrado publicar (hasta el momento) tres de ellos [1–3].

A todos, ¡mi infinita gratitud!



# Capítulo 1

## Introducción

El movimiento es un elemento esencial de la vida y permite a los organismos sobrevivir y adaptarse a su ambiente. Nuestra ignorancia de cómo los organismos se mueven ha motivado un gran número de investigaciones científicas pero hay también otras razones para el creciente interés en este problema. De hecho, en un mundo en el cual tenemos una mayor conciencia de cómo los organismos, ecosistemas y sociedades están intrínsecamente interconectados, la pregunta de cómo y por qué los organismos se mueven se vuelve importante incluso por razones políticas, económicas, ambientales y de salubridad [4]. Por ejemplo, la complejidad del forrajeo y los patrones de movimiento, y la rica dinámica detrás de las relaciones de depredador-presa hacen posible, si no es que inevitable, entender de manera más profunda y sustancial las consecuencias de la actividad humana en la alteración del medio ambiente [4]. O bien, ya que el forrajeo puede verse como una búsqueda aleatoria de alimento, entender este fenómeno podría tener aplicaciones a otros problemas relacionados; como búsqueda de personas desaparecidas o búsqueda de criminales internacionales. En este contexto, Shlesinger [5] ha notado la relevancia de las búsquedas aleatorias en la caza de submarinos en la Segunda Guerra Mundial.

Así, el estudio del desplazamiento de los seres vivos, desde las bacterias hasta las ballenas, es de gran interés en muchas áreas de la ciencia, desde las ciencias exactas hasta las sociales pues el entendimiento de los mecanismos básicos subyacentes de este fenómeno pueden extrapolarse a otro tipo de dinámicas que compartan características básicas con el movimiento animal [4], como se ha mencionado brevemente en el párrafo anterior. En el sentido biológico y ecológico, entender los mecanismos de desplazamiento de los organismos y las técnicas que utilizan para navegar en su medio son el principio básico para entender el destino de los individuos, la estructura y la dinámica de las poblaciones, las comunidades y los ecosistemas, y la evolución y la diversidad de la vida [6, 7]. Este tema ha sido tratado desde diversas áreas de la ciencia como: la biología, la ecología, la física y la matemática. Se ha abordado tanto desde el punto de vista experimental, computacional y teórico [4, 7].

El desplazamiento de organismos vivos presenta características importantes tanto a escalas temporales como espaciales. Debido a su carácter esencialmente errático (aleatorio) el movimiento animal puede describirse adecuadamente empleando métodos estadísticos. La mayoría de estos modelos son caracterizados por Caminatas Aleatorias Simples (CAS), prin-

principalmente procesos Markovianos [4, 6, 8]. La bacteria *E. Coli* es uno de los mejores ejemplos de este tipo de movimiento. Esta bacteria toma pasos rectos que tienen una duración promedio de un segundo, interrumpidos periódicamente por vueltas (cambios de dirección) que duran 0.1 segundos [9]. Durante estos cambios de dirección, el movimiento es muy errático y la dirección del siguiente paso es independiente de la dirección anterior. El desplazamiento de otros organismos también se ha modelado con este tipo de movimiento. Los ejemplos incluyen la mosca *Phormia regina* [10], la mariposa *Euphydryas editha bayensis* [11] y peces que, al ser liberados en una misma área confinada del mar, tienden a difundirse como caminantes aleatorios una vez que la advección por corrientes del agua se ha tomado en cuenta [12]. Por otro lado, en el área de las búsquedas aleatorias, donde un caminante aleatorio tiene como fin encontrar objetos distribuidos en el espacio cuya posición es desconocida *a priori*, una de las preguntas más importantes es: ¿cuál es la mejor estrategia para que dicho buscador tenga éxito en su tarea? El enfoque principal para responder esta pregunta es estudiar la estadística de los tiempos de primer arribo y minimizar el tiempo medio de búsqueda. Particularmente, en las aplicaciones de las búsquedas aleatorias a la ecología, se ha mostrado que algunos modelos Markovianos pueden optimizar el éxito de forrajeo en muchas especies animales. Por ejemplo, para las aves albatros [13] se encontró que una estrategia de búsqueda Markoviana cuya distribución de saltos sigue una ley de potencia optimiza el forrajeo de estas aves, maximizando el cociente del número de sitios de alimento visitados entre la distancia recorrida (eficiencia de forrajeo). Estudios similares se han realizado para el microozoplankton [14] y los monos araña [15].

Paralelamente, estudios recientes han mostrado que muchas especies animales poseen habilidades cognitivas avanzadas como la memoria espacial, la cual les permite regresar a lugares familiares que no necesariamente se encuentran en su rango de visión inmediato [16–18]. Por ejemplo, en [16] se mostró que los bisontes americanos *Bison bison* escogen prados para visitar basándose en la información derivada de sus experiencias pasadas. Janson y Byrne proponen en [17] dos aproximaciones básicas para inferir los procesos cognitivos utilizados por primates en la toma de decisiones de movimiento, a partir de datos experimentales. En ambas aproximaciones (experimental y teórica) se encuentra que la memoria es una herramienta esencial. Además, el uso de memoria tiene efectos importantes sobre la dinámica del movimiento del animal (difusión súper lenta, uso heterogéneo del espacio, tendencia a visitar lugares específicos), por lo cual los procesos Markovianos no resultan ser una buena aproximación para modelar el desplazamiento animal [19–21]. Es por eso que en tiempos recientes se han utilizado nuevos tipos de caminatas aleatorias que toman en cuenta este nuevo ingrediente del uso de memoria para modelar el movimiento animal.

Una clase de caminatas aleatorias no Markovianas son aquéllas donde el movimiento del caminante depende de toda la trayectoria que éste ha recorrido. Estos procesos podrían resultar prometedores para modelar el uso de la memoria en animales pero la relativa ausencia de resultados analíticos limita el entendimiento de los efectos que causa la memoria en los patrones de desplazamiento [22–24]. Además, queda poco claro cómo la memoria mejoraría la eficiencia de forrajeo comparado con una caminata Markoviana [25]. Como ya se ha mencionado, estudios experimentales han probado que algunas especies animales tienen la capacidad de volver a sitios antes visitados por uso de la memoria [17, 18] por lo que

las caminatas aleatorias (CA) reforzadas podrían ser una buena aproximación al fenómeno del movimiento animal. Las CA reforzadas o auto atrayentes son ejemplos importantes de la dinámica no Markoviana; en estos procesos el caminante se mueve, por ejemplo, a sus vecinos más cercanos con probabilidad dependiente del número de veces que estos sitios han sido visitados en el pasado [26, 27]. Este tipo de caminatas pueden ser descritas por una jerarquía de funciones de probabilidad conjunta, lo cual hace imposible escribir las ecuaciones de la dinámica de manera directa y sencilla [26–29]. Esto complica poder encontrar la influencia que el uso de memoria tiene sobre la dinámica de desplazamiento, por ejemplo, responder si, de forma asintótica, el comportamiento es difusivo o localizado [30–32].

Con el fin de intentar resolver muchas de las problemáticas y retos que implica usar CA reforzadas, y que hemos mencionado en el párrafo anterior, en los últimos años se ha desarrollado una herramienta probabilística que ha ido ganando interés como estrategia eficiente de búsqueda: las CA con reinicios aleatorios. Estas caminatas consideran que el individuo, a tasa constante dada, interrumpe la caminata aleatoria y se reubica a su punto de partida [33]. El mayor efecto del reinicio aleatorio se da en la distribución de las posiciones de la partícula difusiva. En ausencia del reinicio, se tiene una distribución Gaussiana usual. Mientras que si el reinicio está presente, el proceso alcanza un estado estacionario fuera de equilibrio [33]. Esta dinámica de reinicio también tiene un impacto profundo en las propiedades de primer arribo cuando el caminante está en búsqueda de un objetivo cuya posición es desconocida. Como es bien sabido en la teoría de las búsquedas aleatorias, el tiempo de primer arribo a un sitio específico del espacio diverge para una partícula difusiva (sin reinicios) [33], en cambio, en presencia del reinicio, el individuo es capaz de llegar por primera vez al objetivo en un tiempo finito, más aún, existe una tasa óptima de reinicio a la cual el tiempo de primer arribo se minimiza.

Los hábitats donde los animales buscan y explotan recursos son en general espacialmente heterogéneos. Así pues, considerar en todos los modelos teóricos un espacio con esta característica resulta de gran importancia ecológica pues los animales que buscan comida, agua, lugares para dormir o pareja encuentran energéticamente conveniente visitar sitios con búsquedas exitosas. La explotación de regiones ricas en recursos (cualesquiera que estos sean) se basa en dos formas de memoria o en la combinación de ambas: una que usa la memoria del ambiente [34] y otra que usa las capacidades cognitivas del animal [18]. La primera, que está bien estudiada, se logra, por ejemplo, mediante el depósito de sustancias químicas como feromonas en las hormigas o marcas físicas para indicar la ruta directa a ubicaciones rentables específicas [35]. En este caso, el individuo no requeriría memoria, simplemente podría seguir el rastro de olor una vez encontrado. La última forma de explotación de recursos, que inspira el trabajo que se presenta aquí, utiliza la capacidad real de un individuo para recordar posiciones previamente visitadas y volver a ellas con preferencia [36]. Durante el forrajeo, los animales pueden tomar decisiones basadas en las experiencias exitosas pasadas, lo cual resulta en una mejora en la explotación de recursos [37, 38] y en un proceso de aprendizaje que les permite adaptarse a su medio ambiente [39, 40]. Por lo tanto, el conocimiento ecológico puede ser el resultado de un aprendizaje continuo a través de la vida de un individuo. Por ejemplo, en ciertas aves marinas, como los alcatraces, los adultos tienen un mejor conocimiento de las zonas provechosas y las explotan de manera más eficiente que los individuos

jóvenes [41]. Hasta ahora, y a nuestro saber, se entiende muy poco sobre los mecanismos de aprendizaje espacial y, prácticamente, no existen modelos matemáticamente solubles de CA con propiedades de aprendizaje. Es por eso que aquí presentamos un modelo que combina el movimiento aleatorio, un esquema de reforzamiento lineal y reinicios estocásticos que nos permitirá entender cómo el aprendizaje puede emerger durante una búsqueda en un espacio heterogéneo.

Dado que muchas especies animales viven en grupos [42, 43], uno podría preguntarse cómo el aprendizaje individual puede verse afectado, o mejorado, por la presencia de otros individuos de la misma especie. La sociabilidad trae consigo muchos beneficios, como un menor riesgo de depredación [43–46] y una mejora en las capacidades de detección y toma de decisiones [47–49] mientras se busca alimento en ambientes desconocidos o inciertos [50–53]. Las decisiones colectivas son sumamente importantes tanto a nivel individual como colectivo ya que proporciona a los individuos una mayor probabilidad de supervivencia y reproducción, y da al grupo una herramienta crucial para mantener la coordinación y la cohesión durante el desplazamiento [54, 55].

Cuando los individuos forrajean en grupo, el éxito de la búsqueda de alimento puede alcanzarse por la transferencia de información entre los miembros de dicho grupo [56–59]. Los insectos sociales como las abejas [60] y las hormigas [61] son ejemplos clásicos de forrajeo colectivo donde la información se intercambia entre individuos por medios externos (rastros químicos o néctar compartido). Estos mismos principios pueden aplicarse a los vertebrados, donde la información no necesariamente se almacena en el ambiente sino dentro de cada uno de los individuos, ya que cada uno de ellos puede utilizar su propia memoria para visitar lugares que se encuentran fuera de su rango de visión. Especies como los macacos [62], las golondrinas [63], las gacelas [64], los elefantes [65] y las hienas [66] se comunican a través de largas distancias (a veces de muchos kilómetros) por medio de vocalizaciones. La información que se transfiere por este medio puede tener un impacto positivo en el éxito de forrajeo de los grupos sociales [64]. En las orcas (ballenas asesinas), las hembras viejas tienden a liderar el grupo durante el movimiento colectivo cuando la abundancia de presas es baja, es decir, son ellas las que transfieren conocimiento ecológico valioso al resto del grupo [67]. En los monos araña, los individuos centrales en la red social tienden a ser seguidos con mayor frecuencia cuando poseen información sobre fuentes de alimento disponibles que los individuos no centrales [68]. Podemos decir, pues, que los mecanismos de coordinación se dividen en dos grandes categorías tomando en cuenta la forma en la que la información puede esparcirse en el grupo: mecanismo local o global. La primera, también llamada auto-organización, es aquella en donde los individuos no son capaces de tener un panorama completo de su entorno y sus reglas de movimiento se guían por la información local que colectan a través de las interacciones con sus vecinos [69]. La segunda se da principalmente en grupos pequeños donde uno o todos los individuos poseen un conocimiento global del entorno y las reglas de movimiento se basan en el conocimiento que dichos individuos poseen (*liderazgo*) [70]. En este trabajo abordaremos el primer tipo de mecanismo por medio de un modelo basado en agentes.

El estudio de animales salvajes tiende a ser una tarea complicada. Los animales generalmente son esquivos y residen en lugares remotos o difícilmente alcanzables para el observador

[71]. Afortunadamente, el desarrollo de dispositivos de telemetría ha facilitado y revolucionado los estudios de campo relacionados con la movilidad [72]. Los datos de telemetría contemporáneos se adquieren a través de dispositivos de comunicación satelital fijados a animales individuales. Estos dispositivos recolectan muchos tipos de datos, pero la mayoría de los estudios se centran en los datos de las posiciones espaciales de los individuos, principalmente para aprender sobre el efecto de la distribución de los recursos en el movimiento individual. Los datos de telemetría han abierto ventanas que nos permiten abordar algunas de las preguntas más fundamentales sobre el movimiento animal, como por ejemplo: ¿Dónde está el animal? (uso del espacio), ¿cómo llegó ahí? (movimiento), ¿dónde prefiere estar? (selección de recursos) y ¿qué está haciendo el animal? (comportamiento) [71].

La manera más común de estudiar y analizar todo estos nuevos datos disponibles es por medio de modelos estadísticos, muchos de estos modelos han sido propuestos en años recientes y varían dependiendo de la motivación del proyecto y del análisis que desea hacerse sobre los datos [73]. En este trabajo abordaremos los modelos estadísticos bayesianos discretos en tiempo y espacio. Entendiendo por un modelo bayesiano aquél que se basa en la inferencia bayesiana, que es el tipo de inferencia estadística en la que las evidencias u observaciones se emplean para actualizar o inferir la probabilidad de que una hipótesis dada pueda ser cierta. El nombre “bayesiana” proviene del uso frecuente que se hace del teorema de Bayes durante el proceso de inferencia [74]. Dicho de manera más precisa: los modelos bayesianos incorporan, primordialmente, el conocimiento previo obtenido de los datos disponibles para estimar un conjunto de parámetros que define un modelo que se ha planteado desde la propia experiencia del observador o desde una teoría probabilística bien establecida, como es el caso de las CA, para describir características relevantes del conjunto de datos observados o predecir el comportamiento de datos aún no recolectados. La estadística bayesiana provee cantidades tanto conocidas como desconocidas, lo que permite incorporar los datos observados dentro de la estimación de los parámetros dados inicialmente, logrando así un proceso de ajuste más rico en información al hacer inferencias sobre las cantidades desconocidas. Todos los modelos bayesianos tienen en común la asignación de la probabilidad como medida de creencia de una hipótesis [74]. Más adelante en este mismo escrito definiremos de manera formal los modelos bayesianos y toda la terminología usada en esta rama de la estadística.

Los ecólogos utilizan desde hace tiempo el teorema de Bayes, por ejemplo, las probabilidades condicionales calculadas usando el teorema de Bayes se utilizaron de manera extensa en modelos dinámicos para el comportamiento durante el forrajeo [75, 76] y para la evasión de depredadores [77]. Estos estudios consideran de manera explícita que los animales forrajeadores utilizan su experiencia previa para modificar actividades de forrajeo futuras y toman una total ventaja de la naturaleza iterativa del teorema de Bayes. La aplicación de la inferencia bayesiana a cuestiones ecológicas ha florecido desde la publicación en 1996 de una serie de artículos sobre inferencia bayesiana para la investigación ecológica y la toma de decisiones ambientales [78]. Los métodos bayesianos se han utilizado más ampliamente en ecología de poblaciones y comunidades, en la que existen muchos modelos que compiten para explicar un mismo fenómeno ecológico [79]. Aunque también ha sido frecuentemente aplicada a nivel individual para estudiar y desarrollar, por ejemplo, modelos de movimiento y uso del espacio [80]. En este escrito expondremos varios modelos bayesianos para analizar

las características del uso del espacio y aprendizaje en ciervos canadienses, estos modelos toman las características esenciales del modelo mínimo que propondremos en los primeros capítulos de esta tesis.

En el capítulo 2 de esta tesis se expone un resumen de los trabajos ya existentes en la literatura relativos a los diversos temas que trataremos a lo largo del desarrollo de este escrito y que ayudarán al lector a familiarizarse con ciertos conceptos y uso de lenguaje que utilizaremos en capítulos posteriores. Específicamente, en la sección 2.1 nos enfocaremos en las caminatas aleatorias con reinicio y caminatas aleatorias auto-atrayentes, exponiendo cuatro trabajos que marcaron el hito en la investigación dentro de esta área. En la sección 2.2 se presenta una recopilación de los principales trabajos existentes y de interés sobre el aprendizaje y movimiento colectivos. En la sección 2.3 se muestran tres de los trabajos más sobresalientes (a nuestro criterio) que ejemplifican de manera clara el uso de inferencia bayesiana en el análisis y ajuste de datos de animales salvajes, tanto a nivel individual como poblacional (jerárquico). En las secciones 2.4-2.5 se presentan los principales objetivos, generales y particulares (respectivamente), que incluyen cómo entender el efecto de la memoria en la difusión y determinar las condiciones bajo las cuales dicho uso de memoria resulta benéfico para la búsqueda y explotación de recursos. En el capítulo 3 se muestra nuestra propuesta de modelo, el análisis teórico y los principales resultados. El capítulo 4 aborda una extensión a colectivos de agentes del modelo mínimo presentando en el capítulo 3, en la que se considera una tasa de transferencia de información entre individuos relativa a sitios visitados y recursos disponibles en los mismos. En el capítulo 5 mostramos las principales definiciones y herramientas dentro del área del análisis bayesiano y exponemos los resultados obtenidos al aplicar cuatro modelos bayesianos jerárquicos distintos para aproximar y predecir las trayectorias reales de 21 Ciervos Canadienses obtenidas a través de datos recolectados en campo. El capítulo 6 presenta las conclusiones de este trabajo de investigación que abarca los modelos individual y colectivo de los capítulos 3 y 4, respectivamente, y el estudio de inferencia bayesiana expuesto en el capítulo 5. Finalmente, en el capítulo 7 se muestra un compendio de material suplementario (tablas y figuras) para la correcta lectura e interpretación de los resultados mostrados en el capítulo 5.

# Capítulo 2

## Antecedentes

### 2.1. Caminatas aleatorias reforzadas en medios homogéneos

Las caminatas aleatorias auto-atrayentes (reforzadas) son una clase importante de procesos no Markovianos con memoria de largo alcance y han sido estudiadas desde el punto de vista matemático desde 1987 por Coopersmith y Diaconis [81], Pemantle [82] y Davis [26]. Las caminatas aleatorias reforzadas son procesos auto-atrayentes donde, típicamente, un caminante aleatorio, que se desplaza sobre una red unidimensional regular, tiende a visitar los sitios vecinos más cercanos de manera preferencial, i.e., el caminante decide brincar a un vecino u a otro con probabilidades proporcionales al tiempo total que haya permanecido en cada uno de ellos. La dinámica anterior es conocida en la literatura como “reforzamiento de sitios”, pero existe también una variante de caminata auto-atrayente en donde el caminante refuerza los vínculos entre sitios (“reforzamiento de vínculos”). Aquí, la probabilidad de ir de un sitio de la rejilla a otro dependerá del número de veces que ese vínculo o camino haya sido utilizado [26]. Podemos imaginar pues que los sitios visitados por el caminante forman una red cuyos pesos de las aristas, sin sentido, nos indican la probabilidad de que el caminante decida usar ese mismo camino o crear uno nuevo. Estas caminatas han sido utilizadas para describir el desplazamiento de las hormigas o las bacterias en donde los caminantes aleatorios depositan una sustancia no difusiva que modifica el ambiente local [27]. En ecología, pueden representar modelos simples para describir la formación de *ámbitos hogareños* [16, 21, 83], entendiendo por ámbito hogareño el área en el cual un organismo cumple sus funciones, incluyendo los límites máximos hasta donde se desplaza para efectos de alimentación, descanso, reproducción, exploración o refugio [84]. Muchos de estos modelos de caminatas reforzadas son definidos a través de las probabilidades de transición que dependen del número de visitas recibidas a cada sitio, o cada vínculo, dando como resultado una dinámica fuertemente dependiente del camino.

### 2.1.1. Caminatas aleatorias reforzadas

Uno de los primeros estudios de este tipo de caminatas aleatorias fue hecho por Davis en [26]. En este trabajo se define un movimiento aleatorio a primeros vecinos en una rejilla unidimensional ( $\vec{X} = X_0, X_1, \dots$ ) en los enteros como sigue: Inicialmente el peso de cada intervalo  $(i, i + 1)$ ,  $i$  un entero, es igual a 1. Si al tiempo  $n$  un intervalo  $(i, i + 1)$  ha sido cruzado exactamente  $k$  veces por  $\vec{X}$ , su peso será  $1 + \sum_{j=1}^k a_j$ , con  $a_j$  una sucesión de números no negativos y  $j \geq 1$ . Dado  $(X_0, X_1, \dots, X_n) = (i_0, i_1, \dots, i_n)$ , la probabilidad de que  $X_{n+1}$  sea  $i_n - 1$  o  $i_n + 1$  es proporcional a los pesos al tiempo  $n$  de los intervalos  $(i_n - 1, i_n)$  y  $(i_n, i_n + 1)$ . Davis prueba que  $\vec{X}$  puede tener dos tipos de comportamiento; recurrente o localizado. Estos distintos comportamientos dependerán de la sucesión de enteros  $a_j$  de la siguiente manera: Si  $\sum_{n=1}^{\infty} (1 + \sum_{i=1}^n a_i)^{-1} = \infty$ , entonces el proceso será recurrente, i.e.,  $\vec{X}$  visitará un número infinito de veces todos los enteros c.s.. Por lo contrario, si  $\sum_{n=1}^{\infty} (1 + \sum_{i=1}^n a_i)^{-1} < \infty$ , entonces la caminata será localizada o de rango finito, i.e.,  $\vec{X}$  visitará un número finito de enteros, oscilando eventualmente en un par de ellos, c.s. Aquí cabe mencionar el caso particular en que la sucesión de enteros  $a_j$  es una sucesión constante, i.e.,  $a_j = a \forall j$  y  $a \in \mathbb{N}$ , en cuyo caso  $\vec{X}$  será recurrente.

Este estudio fue un parteaguas para el análisis de otro tipo de reglas de reforzamiento, mostrando que diferentes dinámicas (difusivas o localizadas) pueden emerger dependiendo de la fuerza de la memoria y de la dimensión espacial [26–28, 85, 86]. Simulaciones numéricas han mostrado una variedad de modelos que parecen exhibir una transición de fase entre ambos estados, donde además se ha presentado evidencia de que la difusión es anómala (*subdifusión*) en el caso no localizado [29, 32]. Este tipo de difusión se ha estudiado de manera extensa en otros contextos [87] y es caracterizada por el hecho de que el desplazamiento cuadrático medio (MSD por sus siglas en inglés) de la partícula no sigue la ley de Smoluchowski-Einstein del movimiento browniano que establece que el cuadrado del desplazamiento medio observado de una partícula que se difunde en un medio viscoso es proporcional a la duración de dicha observación, i.e.,  $\langle X_t^2 \rangle \propto t$ . Así pues, la difusión anómala se caracterizará por  $\langle X_t^2 \rangle \propto t^\mu$ , si  $\mu < 1$  entonces se hablará de subdifusión y si  $\mu > 1$  de súper-difusión. Estos dos anteriores son los casos más estándar de la difusión anómala, pero existe también un caso de subdifusión en el que  $\langle X_t^2 \rangle \propto \ln(t)^\beta$ ,  $\beta > 1$ , el cual mencionaremos más a detalle a continuación.

### 2.1.2. Difusión anómala

Vendruscolo y Marsili establecen en [88] un puente entre las caminatas reforzadas y las caminatas Markovianas en medios desordenados, en particular, utilizan el modelo expuesto por Sinai en [89]. Gracias a esta conexión demuestran que las caminatas reforzadas exhiben difusión anómala. El modelo de Sinai define un tipo de caminata aleatoria llamada CA Aleatoria (CAA) que se caracteriza, en una línea infinita, por la asignación a cada uno de

los sitios  $i = 0, \pm 1, \pm 2, \dots$  de una variable aleatoria  $p_i \in [0, 1]$  tomada de una distribución  $P\{p \leq p_i < p + dp\} = \phi(p)dp$ . La evolución de la posición  $x_t$  de la CAA será:  $x_{t+1} = x_t + 1$  con probabilidad  $p_{x_t}$  y  $x_{t+1} = x_t - 1$  de otra forma. Una de las características más sorprendentes de esta dinámica es que la difusión es extremadamente lenta:  $\langle X_t^2 \rangle \sim \ln(t)^4$  [89]. Este modelo es la base para el modelo expuesto en [88] que es uno de los primeros en incluir la memoria. Para ello, Marsili y Vendruscolo consideran una distribución uniforme entre cero y uno para la medida de desorden  $p_i$ . Se supone que el caminante se mueve en el ambiente sin conocer nada sobre la realización  $\{p_i\}$  del desorden. La única información disponible es el número  $n_{i,t}$  de visitas del caminante al sitio  $i$  y el número  $k_{i,t}$  que define la cantidad de veces que el caminante se ha movido del sitio  $i$  al sitio  $i + 1$ . Esto es, como hemos mencionado ya en la subsección anterior, una caminata aleatoria con reforzamiento de vínculo. Para este caso en particular, el reforzamiento es lineal y proporcional a  $k_{i,t}$ , siendo éste pues un proceso recurrente.<sup>1</sup>

Usando las cantidades  $n_{i,t}$  y  $k_{i,t}$  es posible encontrar una probabilidad de transición efectiva dada por:

$$p_{n,k}^a = \int dP(p|n,k)p = \frac{k+1}{n+2} \quad (2.1)$$

Es decir, la probabilidad de que el caminante vaya del sitio  $i$  al sitio  $i + 1$ , dado que ha hecho el mismo salto  $k$  veces después de  $n$  visitas previas, es  $p_{n,k}^a$ . Esta probabilidad de transición es consistente con la historia pasada del proceso. A este modelo definido por la Ec. (2.1) le llamaron una CA con Memoria (CAM). Esta dinámica también presenta una difusión súper lenta que va como  $\langle X_t^2 \rangle \sim \ln(t)^4$ . En este contexto de CAM este tipo de difusión puede entenderse de manera sencilla con el siguiente argumento<sup>2</sup>: Primero notemos que los valores de  $k_i$  y  $n_i$  en sitios diferentes no son independientes. Por ejemplo, es fácil checar que  $t = \sum_i n_i$  y  $x_t = \sum_i (2k_i - n_i)$ . En general  $n_i = k_{i-1} + n_{i+1} - k_{i+1}$ . En esta relación las  $k_i$ 's están distribuidas uniformemente entre 0 y  $n_i$ . Entonces, aproximadamente, esta relación tiene la forma  $n_{i+1} \simeq C_i n_i$  con  $C_i$  una variable aleatoria. En otras palabras, la variable  $\ln(n_i)$  se comportará como una caminata aleatoria sobre  $i$ , lo que se deriva en que el valor máximo típico de  $n_i$  para  $i \in [0, L(t)]$  (con  $L$  el tamaño de la rejilla con fronteras periódicas donde se mueve el caminante) será  $n_{max} \sim \exp\sqrt{L(t)}$ . Ya que este valor también domina la suma  $t = \sum_i n_i$ , se puede concluir que  $L(t) \sim \ln(t)^2$ .

### 2.1.3. Caminatas aleatorias con reinicio estocástico

Otro tipo de procesos que han ganado interés en los últimos años y que, como lo veremos más adelante, se prestan fácilmente a generalizaciones para incorporar efectos de reforzamiento, son las caminatas aleatorias con reinicio. El reinicio estocástico, según Evans y Majumdar

<sup>1</sup>La sucesión de enteros que define el peso del vínculo de  $i$  a  $i + 1$  es una sucesión de constantes con  $a_j = 1 \forall j$

<sup>2</sup>En estos renglones finales de esta subsección se hace una simplificación de notación: El tiempo  $t$  será fijo y se omitirá el subíndice  $t$  de las cantidades  $n$  y  $k$ .

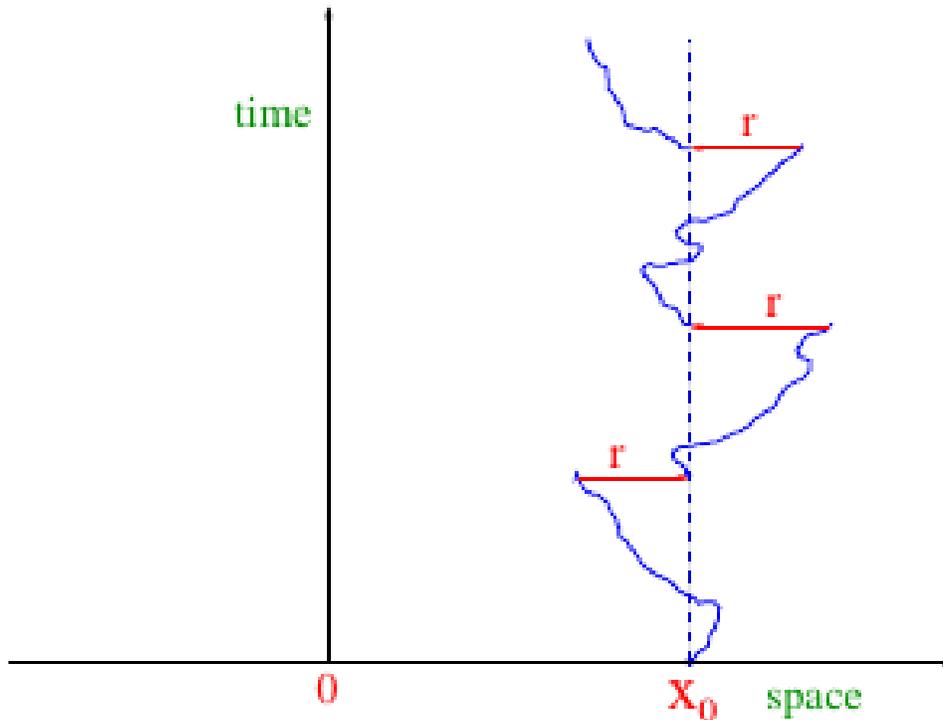


Figura 2.1: Esquema en el espacio-tiempo de una trayectoria de un movimiento browniano que inicia en  $x_0$  y regresa estocásticamente a su posición inicial  $x_0$  a tasa  $r$ . Figura tomada de [33].

[33], es un proceso muy común en la vida cotidiana. Si una búsqueda aleatoria no tiene éxito en su primer intento, el buscador volverá a su sitio de partida para reiniciar una nueva búsqueda hasta alcanzar el objetivo. En [33] se expone un modelo simple donde una partícula difusiva, de forma estocástica, regresa a su posición inicial a tasa constante  $r$  (ver Fig. 2.1).

La ecuación maestra para la probabilidad de que la partícula esté en el sitio  $x$  al tiempo  $t$  ( $p(x, t|x_0)$  en  $1d$ ), dado que inició en  $x_0$  al tiempo 0 va como:

$$\frac{\partial p(x, t|x_0)}{\partial t} = D \frac{\partial^2 p(x, t|x_0)}{\partial x^2} - rp(x, t|x_0) + r\delta(x - x_0) \quad (2.2)$$

con  $D$  la constante de difusión,  $\delta$  la función delta de Dirac y con condición inicial  $p(x, 0) = \delta(x - x_0)$ . Los últimos dos términos del lado derecho de la Ec.(2.2) representan una probabilidad de flujo negativo fuera de cada sitio  $x$  ( $rp(x, t|x_0)$ ) y la correspondiente probabilidad de flujo positivo hacia  $x_0$  ( $r\delta(x - x_0)$ ). Es fácil mostrar que en el límite a tiempos grandes, la partícula alcanza una distribución estacionaria en una dimensión espacial dada por:

$$p_{st}(x|x_0) = \frac{\alpha_0}{2} \exp(-\alpha_0|x - x_0|) \quad (2.3)$$

donde  $\alpha_0 = \sqrt{r/D}$  es el inverso de la longitud de escala correspondiente a la distancia típica difundida por la partícula entre reinicios. Así, una tasa de reinicio finita crea una fuente de probabilidad hacia  $x_0$  mientras que dicha probabilidad se pierde cuando  $x \neq x_0$ , podemos decir pues que hay una circulación de probabilidad incluso a tiempos grandes lo que lleva a un estado estacionario fuera de equilibrio (NESS por sus siglas en inglés) con fluctuaciones no gaussianas para la posición de la partícula en  $1d$ . Este NESS está definido por (2.3). Además, el tiempo promedio en el que la partícula encuentra por primera vez un objetivo fijo en el espacio es finito y tiene un valor mínimo en una tasa de reinicio óptima  $r^*$  [33].

#### 2.1.4. Caminatas aleatorias preferenciales

Inspirados en el modelo expuesto por Evans y Majumdar [33] y sus extensiones estudiadas por Kusmierz en [90, 91], Boyer y Solis-Salas proponen en [85] un modelo de reforzamiento donde el caminante puede realizar movimientos aleatorios (exploración local) o reubicarse en sitios visitados en el pasado (por medio de pasos largos). Contrario a otros modelos usuales de reforzamiento, este modelo es analíticamente tratable, pudiéndose obtener resultados exactos sobre la dinámica y el efecto del reforzamiento sobre ésta. Dicho modelo depende de un parámetro constante  $q$  que describe la tasa relativa de los dos tipos de movimiento arriba mencionados como sigue: Se considera un caminante con posición  $\mathbf{X}_t$  al tiempo  $t$  en una rejilla regular, uniespaciada  $d$ -dimensional y cuya posición inicial está dada por  $\mathbf{X}_0 = \mathbf{0}$ . Sea  $q$  un parámetro constante.  $0 < q < 1$ . A cada paso discreto de tiempo,  $t \rightarrow t + 1$ , el caminante decide con probabilidad  $1 - q$  visitar alguno de sus vecinos más cercanos, como en la CAS. Con probabilidad complementaria  $q$ , se reubica directamente en un sitio visitado en el pasado. En este caso, la probabilidad de escoger un sitio específico ya visitado de la rejilla es proporcional al número de visitas que dicho sitio haya recibido en el intervalo  $[0, t]$ . Esta regla de revisita preferencial es equivalente a escoger un entero aleatorio  $t'$  de manera uniforme en el intervalo  $[0, t]$  y regresar al sitio visitado en  $t'$  con lo que el proceso es altamente no Markoviano.

La ecuación que define la dinámica expuesta arriba, en  $1d$ , está dada por:

$$P(n, t + 1) = \frac{1 - q}{2} P(n - 1, t) + \frac{1 - q}{2} P(n + 1, t) + \frac{q}{t + 1} \sum_{t'=0}^t P(n, t') \quad (2.4)$$

El último término de la Ec.(2.4) indica que el sitio  $n$  puede ser visitado (desde cualquier otro sitio de la rejilla) de acuerdo a la regla de memoria dado que el caminante estuvo en el sitio  $n$  al tiempo  $t'$ . Abusando un poco del lenguaje, llamaremos a la Ec.(2.4) la *ecuación maestra* que describe la dinámica expuesta en [85].<sup>3</sup>

En este tipo de dinámica la memoria modifica profundamente la difusión normal del proceso y genera patrones complejos en la ocupación del espacio. Por ejemplo, el MSD crece

---

<sup>3</sup>Como bien sabemos, una ecuación maestra es tal sólo cuando describe procesos Markovianos, este proceso recién descrito no es para nada Markoviano y la justificación de la Ec.(2.4) no es trivial como puede verse en [85].

asintóticamente muy lentamente con el tiempo y está definido por  $\langle X_t^2 \rangle \simeq \frac{1-q}{q} [\ln(qt) + \gamma]$ , con  $\gamma$  la constante de Euler. Por otro lado, la distribución  $P(n, t)$  tiende a una Gaussiana, como en la difusión normal, pero con una varianza que sigue la ley anómala logarítmica mencionada arriba en lugar de la ley normal en la que  $\langle X_t^2 \rangle \propto t$ . Así, la inclusión del uso de memoria dentro del movimiento puramente aleatorio lleva, pues, a una generalización del Teorema del Límite Central [92].

Este modelo se utilizó para aproximar datos tomados en campo de monos capuchinos (*Cebus capucinus*) dando un ajuste satisfactorio y revelando un uso de memoria relativamente pequeño para estos mamíferos ( $q \approx 0.1$ ) [85].

Así pues, debido a su naturaleza, las caminatas aleatorias reforzadas y las caminatas aleatorias con reinicio ofrecen una alternativa prometedora a las caminatas aleatorias Markovianas como modelo del movimiento ecológico y la búsqueda de recursos [4, 6, 8, 16]. En este trabajo introduciremos en el capítulo 3 un modelo con reglas de reforzamiento y reinicio estocásticos en un espacio heterogéneo, terreno que ha sido, a nuestro saber, poco estudiado y que podría resultar prometedor para explicar de forma más precisa y realista cómo el uso de la memoria puede inducir una adaptación emergente al ambiente, es decir, bajo qué condiciones ocurre una transición entre una dinámica localizada y difusiva y cómo una localización específica puede llevar a la aparición de fenómenos tales como el aprendizaje espacial y la adaptación por medio de una localización del caminante alrededor de sitios de recursos o de atrapamiento.

## 2.2. Movimiento Colectivo

Como hemos mencionado ya en la introducción de esta tesis, muchas de las aplicaciones a ecología de las caminatas aleatorias como modelo de movimiento animal se enfocan en individuos aislados, y dejan de lado el hecho de que muchas especies animales son sociables y viven en colectivo. Con el fin de entender el efecto que esta característica de sociabilidad tiene sobre la toma de decisiones, el desarrollo de mecanismos de transferencia de información y el éxito de forrajeo grupal se han propuesto diversos modelos teóricos que van desde el uso de ecuaciones diferenciales hasta modelos de agentes, agentes que no toman en cuenta el conocimiento adquirido al interactuar con su ambiente. A continuación expondremos cuatro estudios que consideramos esenciales en el estudio del movimiento colectivo y el efecto que trae consigo la colectividad sobre el éxito de forrajeo y cómo éste se mejora considerablemente cuando los individuos sociables son capaces de intercambiar información valiosa (como la ubicación de un sitio rico en alimento).

### 2.2.1. Estrategia de forrajeo grupal

En [42] Clark y Mangel exponen un modelo simple de movimiento colectivo en parvadas (*flocking*) forrajeadoras. Se considera que esta estrategia de forrajeo grupal es “una estrategia

evolutivamente estable” y que proporciona a la parvada un aumento en la información útil grupal debido a la información útil recolectada por cada individuo aún y cuando ello no se traduzca en un aumento global de la tasa de alimento consumido. El modelo, que no es espacialmente explícito, se centra en considerar un ave que busca su alimento en un área dada del ambiente que está formado de parches distribuidos aleatoriamente y de tamaño uniforme. Si la probabilidad de encontrar un parche en el intervalo de tiempo  $(t, t + dt)$  es independiente del patrón de forrajeo, ésta está dada por:

$$p = \lambda dt \quad (2.5)$$

donde  $\lambda$  es una constante igual a la tasa promedio de éxito (el forrajeador encuentra un parche). Si se considera, además, el tiempo requerido para que el ave consuma un parche completamente y reinicie la búsqueda ( $\tau$ ), entonces la tasa promedio de alimentación, dado que el tiempo esperado para encontrar un parche es  $1/\lambda$  viene dada por:

$$f_\lambda = \frac{B}{\tau + 1/\lambda} = \frac{B\lambda}{\tau\lambda + 1} \quad (2.6)$$

donde  $B$  es la cantidad de alimento que el parche contiene (peso del parche).

Tomando en cuenta este comportamiento individual, se considera luego que hay dos aves forrajeando sobre la misma área. Cada una de ellas puede escoger entre dos estrategias diferentes: La estrategia  $N$ , que define una búsqueda y alimentación independiente entre aves o bien, la estrategia  $F$ ; durante la búsqueda una de las aves también observa a su compañera y cada que su par encuentra un parche ésta va y consume del mismo. En esta última estrategia, el ave multiplica por dos la cantidad de información que posee sobre la ubicación de los parches. Considerando lo anterior, se calcula la tasa promedio de alimentación para cada ave, dependiendo de las estrategias usadas por ambas, asumiendo que las dos aves dividen en partes iguales el alimento disponible y que no se gasta tiempo peleando sobre el parche.

Si las dos aves usan la estrategia  $N$ , entonces para cada una la tasa promedio de alimentación viene dada por (2.6). Si ambos individuos utilizan la estrategia  $F$ , y el tiempo que tarda cada ave en alcanzar el parche encontrado por su par es pequeño comparado con  $\tau$  y  $1/\lambda$ , entonces el tiempo esperado entre encuentros (encontrar parches) es aproximadamente  $1/2\lambda$ , el tiempo de consumo es  $\tau/2$  y cada ave come  $B/2$  por parche. Por lo tanto, la tasa promedio de alimentación sigue estando dada por (2.6). Finalmente, si solo una de las aves usa la estrategia  $F$ , entonces las tasas promedio de alimentación son  $2f_\lambda - f_{\lambda/2}$  y  $f_{\lambda/2}$  para la ave que usó estrategia  $F$  y para la que no, respectivamente, esto ya que la segunda ave consume sólo la mitad de cada uno de los parches que encuentra por lo que su tasa promedio de alimentación es  $(B/2)/(\tau/2 + 1/\lambda) = f_{\lambda/2}$ , y se asume que la primer ave va al encuentro de la segunda cada vez que ésta encuentra un parche aún y si ella se encuentra consumiendo su propio parche, además, la primer ave es capaz de volver al parche que dejó sin consumir y terminar de hacerlo. Con esto, la tasa promedio de alimentación para la primer ave será lo que ella consume por sí misma ( $f_\lambda$ ) más lo que consume al compartir los parches con la segunda ave  $((B/2)/(\tau/2 + 1/2\lambda) = f_{\lambda/2})$  menos la tasa promedio de alimentación de la segunda ave ( $f_{\lambda/2}$ ). Así, sin importar qué estrategia escoja el ave 2, la primer ave siempre

obtendrá un mayor provecho al usar la estrategia  $F$ . Por lo tanto, podría decirse que una estrategia de búsqueda grupal resulta mucho más eficiente (comida consumida / tiempo de traslado) que el forrajeo individual, por lo menos para aquellos individuos que eligen una estrategia de búsqueda colectiva. Este es uno de los primeros modelos en los que se maneja la transferencia de información entre forrajeadores, mismos que carecen de memoria.

### 2.2.2. La comunicación mejora el forrajeo grupal

Uno de los primeros modelos estocásticos que investiga la relación entre la comunicación y la eficiencia en el forrajeo es el propuesto en [57]. Martínez-García *et al.* consideran  $N$  partículas (individuos) que se mueven en un ambiente  $2d$  y que se rigen por caminatas aleatorias correlacionadas las cuales son más apropiadas para modelar la persistencia en dirección del desplazamiento animal. El movimiento está sesgado por los gradientes de la calidad del ambiente (información local) y por la interacción entre individuos a través de un mecanismo de comunicación que es activado cuando un buen recurso es encontrado, lo que proporciona información de la calidad del ambiente en otras áreas (información no local). La dinámica para cada una de las partículas  $i = 1, \dots, N$  es:

$$\dot{\mathbf{r}}_i(t) = B_g \nabla g(\mathbf{r}_i) + B_C \nabla S(\mathbf{r}_i) + \eta_i(t) \quad (2.7)$$

donde  $\eta_i(t)$  es el término del ruido blanco Gaussiano caracterizado por  $\langle \eta_i(t) \rangle = 0$  y  $\langle \eta_i(t) \eta_j(t') \rangle = 2D \delta_{ij} \delta(t - t')$ , con  $D$  una intensidad de ruido (o coeficiente de difusión del animal en ausencia de los otros términos). El término  $B_g \nabla g(\mathbf{r}_i)$  se refiere a la búsqueda local, donde  $g(\mathbf{r})$  es la función que define la calidad del ambiente (cantidad de césped, presa, agua, etc.) del área de estudio y  $B_g$  es el parámetro de sesgo de la búsqueda. En [57] se define  $g(\mathbf{r})$  a partir de tres funciones Gaussianas no normalizadas, para asegurar que  $g(\mathbf{r}) \in [0, 1]$ , centradas en distintos puntos de un espacio unitario con fronteras periódicas. La configuración del ambiente permanece fija en el tiempo. Luego,  $B_C \nabla S(\mathbf{r}_i)$  es el término de la búsqueda no local, con  $B_C$  el parámetro de sesgo no local y  $S(\mathbf{r}_i)$  es la función de información disponible del individuo  $i$ . ésta representa la información que llega a la posición del individuo  $i$  como resultado de la comunicación con el resto de la población. Este término hace que los individuos se muevan a lo largo de gradientes de información recibida y es una función de la superposición de los pares de interacción entre el individuo  $i$  y cada uno del resto de sus pares.

$$S(\mathbf{r}_i) = F \left( \sum_{j=1, j \neq i}^N A[g(\mathbf{r}_j)] V(\mathbf{r}_i, \mathbf{r}_j) \right) \quad (2.8)$$

donde  $F$  es una *función de percepción* arbitraria, la cual debe de fijarse en respuesta a las necesidades específicas de la aplicación del modelo, i.e., tomando en cuenta la especie animal que ha decidido estudiarse. El término  $V(\mathbf{r}_i, \mathbf{r}_j)$  es la interacción entre la partícula receptora  $i$  dependiendo de su posición  $\mathbf{r}_i$  y la partícula emisora localizada en  $\mathbf{r}_j$  y  $A[g(\mathbf{r}_j)]$  es la función de activación (típicamente una función de Heaviside que activa la interacción cuando la calidad del recurso sobrepasa cierto umbral  $\kappa$ , i.e.,  $A[g(\mathbf{r})] = \Theta(g(\mathbf{r}) - \kappa)$ ) que nos indica si el individuo en  $\mathbf{r}_j$  llama a los otros dependiendo si se encuentra o no en un buen sitio de recurso.

Utilizando simulaciones de Monte Carlo para las dinámicas individuales definidas por la Ec.(2.7), fijando las ecuaciones de densidad para definir la distribución de recursos en el espacio, y definiendo la función de información disponible como:

$$S(\mathbf{r}_i) = \sum_{j=1, j \neq i}^N A[g(\mathbf{r}_j)] \frac{\exp(-\frac{(\mathbf{r}_i - \mathbf{r}_j)^2}{2\sigma^2})}{2\pi\sigma^2} \quad (2.9)$$

con un umbral  $\kappa = 0.85$ , se muestra que la búsqueda se optimiza (el tiempo promedio de primera llegada al mejor sitio entre los buscadores es mínimo) cuando las escalas de comunicación son intermedias (ver Fig.2.2). Es decir, el número de individuos del cuál un animal recibe información valiosa crece con la escala de interacción ( $\sigma$ ). Cuando esta escala es muy pequeña, los individuos reciben muy poca información (cero información cuando  $\sigma = 0$ ) y por lo tanto exhiben una eficiencia de búsqueda muy pequeña. La eficiencia se mide directamente a partir de la cantidad  $\tau$  que se define como el tiempo en el que la mitad de la población encuentra, al menos, un buen sitio de recurso. Así, un  $\tau$  alto dará una eficiencia pobre y un  $\tau$  pequeño se traducirá en una gran eficiencia. Algo similar ocurre cuando las escalas de interacción son muy grandes, esto lleva a los individuos a verse saturados con la cantidad de información que reciben de todas direcciones y por lo tanto no pueden decidir de manera eficiente cuál de ellas elegir. Sólo las escalas de comunicación intermedias proporcionan a los individuos la cantidad de información necesaria con la cuál pueden localizarse de manera eficiente en aquellos sitios con recursos valiosos.

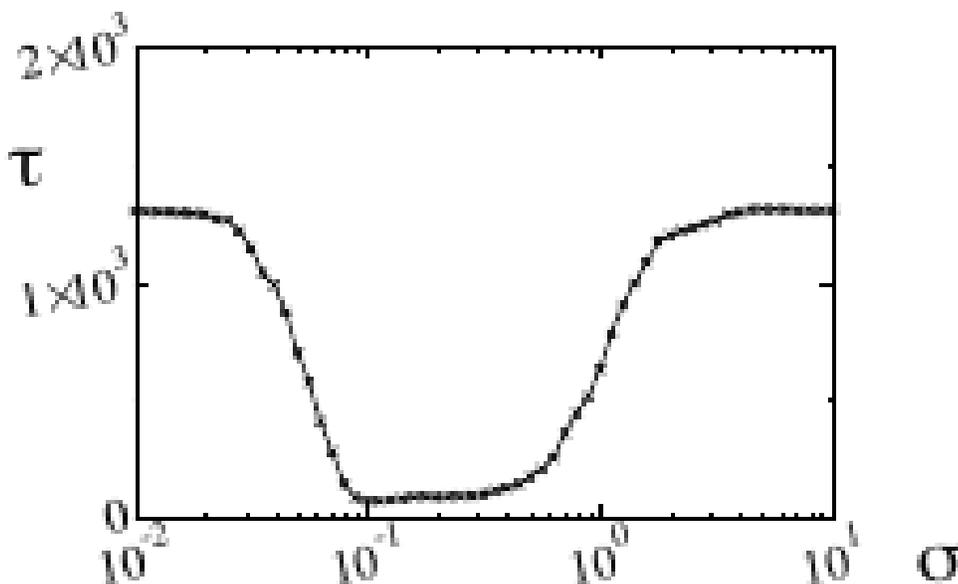


Figura 2.2: Tiempo de búsqueda  $\tau$  como función de  $\sigma$ , usando la Ec. (2.7) con función de interacción definida en (2.8). Figura tomada de [57].

Tanto el exceso como la ausencia de información hacen que la búsqueda falle. En [57]

también se muestra una aplicación al caso particular de la comunicación acústica entre las gacelas de Mongolia (*Procapra gutturosa*) obteniendo buenos resultados entre las predicciones del modelo y los datos de campo disponibles. Los agentes en este modelo carecen de memoria, dándose la adaptación sólo por medio de las interacciones locales (en tiempo y espacio) entre agentes.

### 2.2.3. Modelo basado en agentes para el forrajeo colectivo

Uno de los trabajos más recientes que abordan el problema del forrajeo colectivo es el de Bhattacharya y Vicsek [93], en él proponen un modelo basado en agentes para el forrajeo colectivo en el cual el movimiento individual de cada forrajeador puede ser aleatorio en la ausencia de interacciones. Este modelo resulta de interés pues, además de considerar agentes que pueden comunicarse entre sí, es de fácil implementación y posee una cantidad de parámetros relativamente pequeña, lo que resulta valioso al momento de la interpretación de los resultados. En este trabajo se expone, además, una técnica de modelaje que nos será de inspiración para el capítulo 4 de este escrito.

Los forrajeadores son caminantes en una rejilla dos dimensional, donde los parches de recurso se regeneran aleatoriamente en el espacio luego de ser consumidos. Los caminantes tienen información local sobre la distribución de estos sitios de recurso y cada forrajeador tiene información global sobre el estado actual de los demás caminantes en el espacio [93]. El modelo se define de manera formal como sigue:

La rejilla, que hace las veces de ambiente, es de tamaño  $L$  con fronteras periódicas. Estas fronteras son válidas tanto para el movimiento de los individuos como para la regeneración de los parches, las cuales explicaremos a detalle más adelante. En dicha rejilla hay  $N_F$  forrajeadores cuyas posiciones iniciales se distribuyen al azar. Los parches (alimento) se regeneran de tal suerte que a cada paso de tiempo hay una cantidad constante  $N_T$  de sitios de recurso en el espacio. Se define la vecindad de un individuo como la región circular de radio  $R$  centrada en él. Se asume que a cada instante de tiempo cada individuo tiene la capacidad de censar cuáles de sus vecinos han tenido éxito en identificar y consumir alimento, estos vecinos se denominarán *forrajeadores exitosos*. Un forrajeador sólo es capaz de detectar sitios de recurso cuando ha llegado a ellos. Sea  $(x_i^t, y_i^t)$  la posición del individuo  $i$  al tiempo  $t$  y sea  $j$  el vecino más cercano a  $i$  que ha consumido alimento al tiempo  $t$ . Las reglas que dictan el movimiento de  $i$  son las que siguen:

- (1) Si hay alimento en el sitio  $(x_i^t, y_i^t)$ , entonces el individuo se queda en esa posición y consume uno, y sólo uno, de los recursos disponibles. Sea  $h^t(x, y)$  la cantidad de alimento en el sitio  $(x, y)$  al tiempo  $t$ . Entonces  $(x_i^{t+1}, y_i^{t+1}) = (x_i^t, y_i^t)$  y  $h^{t+1}(x, y) = h^t(x, y) - 1$ . Si a cualquier instante de tiempo el número de forrajeadores excede el número de parches en cualquier sitio de la rejilla, entonces los parches se distribuyen al azar y de manera uniforme entre los individuos.
- (2) Si el sitio  $(x_i^t, y_i^t)$  está vacío, y el individuo llegó a éste por un *paso aleatorio* (se explica abajo), o el animal estuvo en el tiempo anterior en él por presencia de alimento, entonces

el siguiente movimiento del individuo se decide de la siguiente manera:

- (a) Si no hay ningún forrajeador exitoso dentro de la vecindad del individuo  $i$ , entonces el animal toma un paso aleatorio, es decir,  $(x_i^{t+1}, y_i^{t+1})$  será alguno de los cuatro vecinos más cercanos  $(x_i^t \pm 1, y_i^t)$  o  $(x_i^t, y_i^t \pm 1)$  con igual probabilidad cada uno.
  - (b) Si existe un vecino exitoso  $j$ , entonces con probabilidad  $p_i$  el individuo “se mueve” en la dirección a ese vecino y con probabilidad  $1 - p_i$  ejecuta un paso aleatorio. Aquí  $p_i = \exp(-\alpha d_{ij})$  donde  $d_{ij}$  es la distancia euclideana entre  $(x_i^t, y_i^t)$  y  $(x_j^t, y_j^t)$ . Así, la interacción entre individuos está caracterizada por el parámetro  $\alpha$ . Además, el movimiento en dirección a  $j$  implica escoger alguno de los cuatro posiciones vecinas más cercanas de tal suerte que la distancia entre  $(x_i^{t+1}, y_i^{t+1})$  y  $(x_j^t, y_j^t)$  sea menor que  $d_{ij}$ . A este tipo de movimiento se le llama un *paso con dirección* y se asume, además, que el individuo  $i$  recuerda la distancia  $d_i^t = d_{ij}$ .
- (3) Si el sitio  $(x_i^t, y_i^t)$  está vacío y el animal llegó a éste por medio de un paso con dirección, entonces su movimiento se decide como sigue:
- (a) Misma regla que en (2)(a).
  - (b) Si hay un forrajeador exitoso  $j$  en su vecindad y  $d_{ij} < d_i^{t-1}$ , entonces el individuo  $i$  hace un paso con dirección a  $j$  y  $d_i^t = d_{ij}$ .
  - (c) Si hay un forrajeador exitoso  $j$  en su vecindad y  $d_{i,j} > d_i^{t-1}$ , entonces sigue la misma regla ya explicada en (2)(b).
- (4) La regeneración de los sitios de recurso a cada instante se hará de la siguiente manera: Sea  $\Delta N(t)$  la cantidad de parches consumida a cada instante de tiempo  $t$ . Se selecciona al azar uno de los parches restantes  $N_T - \Delta N(t)$ . A una distancia  $d$  ( $1 \leq d \leq L/\sqrt{2}$ ) del parche escogido de entre los restantes, se coloca un nuevo parche en una dirección aleatoria. La distancia  $d$  se escoge de una distribución de cola larga  $P(d) \sim d^{-\gamma}$ , con  $\gamma$  un parámetro. Este proceso se repite  $\Delta N(t) - 1$  veces de tal suerte que el número de parches al inicio del tiempo  $t + 1$  sea  $N_T$ .

Con estas reglas de movimiento, resulta claro que en el límite  $\alpha \rightarrow \infty$ , las trayectorias de los forrajeadores son esencialmente caminatas aleatorias. La tendencia de los individuos de viajar a través de sitios lejanos, cuando un parche ha sido encontrado por alguno de sus vecinos, crece conforme decrece el valor de  $\alpha$ . Por lo tanto, las trayectorias se vuelven una combinación entre pasos aleatorios y pasos con dirección hacia lugares con recursos explotados por otros.

Uno de los principales resultados expuestos es que la presencia de interacciones dentro de un grupo de forrajeadores resulta esencial para maximizar la eficiencia en la búsqueda de alimento (cociente entre el total de alimento consumido entre la distancia total viajada). Por otro lado, la estrategia óptima en la cual todos los individuos se benefician, se da de la combinación de una búsqueda individual e independiente del grupo y una búsqueda colectiva dada por las interacciones (valores intermedios pero pequeños de  $\alpha \in (10^{-2}, 10^{-1})$ )<sup>4</sup>, sin

---

<sup>4</sup>Cabe mencionar que la eficiencia aumenta conforme  $\alpha$  crece, alcanza su máximo en los valores intermedios arriba mencionados, decrece al aumentar  $\alpha$  hasta alcanzar un valor de saturación cuando  $\alpha \rightarrow \infty$ .

importar la distribución de los parches, es decir, los parámetros  $\alpha$  y  $\gamma$  son independientes.

Los modelos presentados en esta sección estudian la importancia de las interacciones y la comunicación en un grupo de animales que forrajea un espacio heterogéneo, y todos llegan a conclusiones importantes y más o menos similares; el forrajeo colectivo es más eficiente (bajo ciertas condiciones específicas) que el forrajeo individual. Encuentran, también, mecanismos particulares bajo los cuales la eficiencia del forrajeo colectivo se maximiza. Pero como se ha mencionado ya, ninguno de estos modelos considera el papel del uso de la memoria (la experiencia acumulada por cada individuo a lo largo de su historia) en el éxito del forrajeo colectivo. Uno de nuestros objetivos es tomar en cuenta, justamente, el uso de memoria, el cual permite la posibilidad de describir fenómenos adaptativos y de aprendizaje, mismos que dan pie a otro mecanismo para la emergencia de un forrajeo colectivo exitoso.

## 2.3. Análisis de trayectorias animales reales

En la introducción de esta tesis discutimos el hecho de que en tiempos recientes se ha tenido acceso a una cantidad impresionante de datos de telemetría sobre la posición individual de animales, lo que ha llevado, a su vez, al desarrollo de modelos estadísticos para el estudio y análisis de dichos datos. En esta subsección final de los antecedentes, presentaremos tres trabajos que ejemplifican la aplicación de modelos matemáticos basados en caminatas aleatorias para ajustar datos observados de animales salvajes. Nos hemos enfocado en aquellos modelos que consideran distintos tipos de dinámica (modos de comportamiento) pues en el capítulo 5 presentaremos un compendio de modelos, también basados en caminatas aleatorias, que consideran el uso de memoria para el forrajeo animal. Así nuestros modelos, además de considerar el uso de memoria para el movimiento de los animales, consideran de forma intrínseca dos tipos de movimiento: un movimiento “azaroso” en el cual el individuo no hace uso de su experiencia para escoger su próxima posición (este tipo de movimiento es análogo a la exploración), y otro tipo de movimiento en el cual el animal hace uso de la información obtenida por la experiencia para escoger su próximo sitio de alimento (este tipo de movimiento es análogo a la explotación).

### 2.3.1. Identificando dos tipos de movimientos en ciervos

Una forma ampliamente estudiada para analizar trayectorias animales, es caracterizar las mismas no sólo por las coordenadas espaciales de cada una de las posiciones, sino también por las longitudes y ángulos de giro entre posiciones. La longitud será simplemente la distancia euclídeana entre dos posiciones consecutivas y se entiende por ángulo de giro, el ángulo formado por las líneas rectas que conectan tres posiciones también consecutivas. Ya que es de esperar que un animal se mueva rápido a través de un hábitat con poca disponibilidad de recursos y que se mueva de manera lenta en ambientes que mejoran de manera significativa la cantidad de recursos disponibles para el individuo, se puede decir que cuando el animal está en un estado de búsqueda intensiva (como después de encontrar parche con abundante alimento) la longitud de los pasos tenderá a ser corta y los giros con ángulos grandes. En

contraste, cuando está en un estado de búsqueda extensivo, la longitud de pasos será larga y giros con ángulos pequeños [80]. Considerando estos dos estados de comportamiento, Morales *et al.* asumen en [80] que la trayectoria de un animal se compone de dos tipos de caminatas aleatorias, i.e, cada paso y cada giro son asignados a una caminata aleatoria que caracteriza un estado de comportamiento distinto. Así pues, cada caminata aleatoria estará definida por las distribuciones de la longitud de pasos y los ángulos de giro. En [80] buscan clasificar cada observación en una de esas CA y obtener los parámetros que definen las distribuciones que las caracterizan.

De forma general, en [80] la asignación de los estados de comportamiento se infiere de los datos observados de movimiento y puede incluir el tipo de hábitat en el cual el animal está localizado. El cambio entre distintos estados de comportamiento puede modelarse de manera explícita utilizando una matriz de transición entre estados que puede ser estimada directamente de los datos de la trayectoria, o bien, de las características del ambiente. El ajuste del modelo se hace bajo el marco de la inferencia bayesiana, metodología que expondremos de manera mucho más detallada en las primeras secciones del capítulo 5 de esta tesis. Así pues, se hace necesario estimar los parámetros que definen el modelo de movimiento en cada uno de los estados de comportamiento y los parámetros que definen las transiciones entre estados.

De manera formal, en [80] se considera una sucesión de  $T$  variables observadas  $y_t$ , de tal suerte que a cada medición  $y_t(t = 1, \dots, T)$  se le asocia con un modo de comportamiento no observable latente  $I_t = i, i \in \{1, \dots, M\}$  donde  $M$  es el número de los distintos estados de movimiento a considerar. Cada observación se asigna a uno y sólo uno de los estados  $M$ . Las observaciones  $y_t = [r_t, \phi_t]$ , son los pares de la longitud del  $t$ -ésimo paso ( $r_t$ ) y su respectivo ángulo de giro ( $\phi_t$ ). En [80] se asume que las observaciones son independientes entre sí y que se obtienen de la multiplicación de una distribución de Weibull para longitud de paso y una distribución de Cauchy para los ángulos de giro. Cada una está definida como sigue:

$$W(x) = abx^{b-1}\exp(-ax^b) \quad (2.10)$$

para la distribución Weibull de parámetros  $a, b$ , y la función de densidad para la Cauchy sigue la forma,

$$C(\phi) = \frac{1}{2\pi} \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos \phi - \mu} \quad 0 \leq \phi \leq 2\pi, 0 \leq \rho \leq 1 \quad (2.11)$$

con parámetros  $\mu$  y  $\rho$  que definen la dirección media y el coseno medio de la distribución angular, respectivamente.

Así pues, la distribución Weibull estará definida por los parámetros  $a_i$  y  $b_i$  ( $i \in \{1, \dots, M\}$ ) y la distribución Cauchy por los parámetros  $\mu_i$  y  $\rho_i$ . Así para un vector de estados dado  $I_t$ , la función de probabilidad (likelihood o función de verosimilitud) que define la trayectoria está dado por:

$$P(y|a, b, \mu, \rho) = \prod_{t=1}^T W(r_t|a_{I_t}, b_{I_t})C(\phi_t|\mu_{I_t}, \rho_{I_t}) \quad (2.12)$$

donde  $W$  y  $C$  denotan a las distribuciones Weibull y Cauchy, respectivamente. De manera general, si suponemos valores fijos para cada uno de los parámetros e, igualmente, un número fijo de estados para cada paso, la función de probabilidad  $P$  representa la probabilidad de generar la trayectoria observada con el modelo propuesto. El principal objetivo es pues encontrar los valores de los parámetros y la cantidad de estados tal que  $P$  se maximice. Los métodos bayesianos son particularmente útiles para encontrar los valores de  $a, b, \mu$  y  $\rho$  que ajustan mejor para cada estado de movimiento, la manera en la que se realiza este análisis y procedimiento, será explicado a detalle en las primeras secciones del capítulo 5 de esta tesis.

En este trabajo presentado por Morales *et al.* [80] se exponen siete modelos de movimiento diferentes que consideran una, dos o hasta tres CA combinadas, algunos de estos modelos contemplan las tasas de cambio entre distintas CA, donde estas tasas dependen o no del ambiente. Si las tasas de cambio no dependen del ambiente, se considerarán tasas de transición fijas, si las tasas dependen del ambiente, se incluirán a la función likelihood definida en (2.12) y el número de parámetros a ajustar aumentará dependiendo de las variables ambientales a considerar (para mayor detalle ir a [80]).

El análisis se hizo sobre los siete modelos considerando como objeto de estudio a cuatro ciervos que fueron trasladados (junto con otros 116 ciervos) del Parque Nacional de la Isla del Ciervo, en Alberta Canadá, a la parte Este-Central de Ontario Canadá. Las observaciones se recolectaron en el Este-Central, el período promedio entre observaciones fue de dos horas durante 218 días, lo que da un promedio total de 2,616 muestras por individuo. Para estos cuatro ciervos particulares es uno de los siete modelos el que mejor ajusta las trayectorias de tres de los cuatro sujetos de estudio. El modelo es aquél que considera dos CA con tasas de cambio dadas y constantes. Las dos CA corresponden a los estados de “exploración” (pasos largos y ángulos pequeños) y “acampada” (pasos cortos y ángulos grandes). Así, a cada paso de tiempo el animal puede decidir cambiar de su estado actual al otro con probabilidad fija definida por las tasas de cambio entre modos de movimiento. En el estado de exploración, la distribución de la longitud de pasos es forzada a tener una moda mayor que cero.

Se puede decir entonces que el movimiento de los ciervos puede pensarse como un movimiento multifacético: los ciervos gastan la mayor parte de su tiempo en un estado de acampada (en el cual la longitud de los pasos es del orden de los cientos de metros, y los ángulos de giro tienden a ser muy grandes), o en un estado de exploración (con longitud de pasos en el orden de los kilómetros y ángulos de giro pequeños). La aplicación de este modelo revela, de manera consistente con las observaciones, que los ciervos acampan de manera más probable en hábitats abiertos (campos de agricultura o bosques abiertos) pero no encuentra una asociación con el ambiente para el estado exploratorio [80].

### 2.3.2. Dos estados de movimiento durante la migración

Uno de los principales obstáculos al momento de analizar datos de trayectorias animales es que, en muchos casos, los datos sólo pueden recolectarse en períodos distantes, por ejemplo, en muchos de los animales marinos, la recolección sólo se hace posible cuando éstos se

encuentran en la superficie o relativamente cerca de ésta, lo cual provoca mediciones ruidosas y con alta variabilidad. En 2007, Jonsen *et al.* presentan en [94] un modelo novedoso para el movimiento animal que tiene la capacidad de trabajar con datos con grandes medidas de error y alta variabilidad en las dinámicas de movimiento, además, este modelo propuesto en [94] es capaz de identificar patrones ocultos en los datos. Esta última fue la aproximación que Morales *et al.* implementaron en [80] y que se expuso brevemente arriba. Así pues, el principal objetivo del modelo presentado en [94] es identificar comportamientos de forrajeo a partir de datos de rastreo tomados de manera remota (la observación directa no es posible). Para lograrlo se necesitan estimar tres cosas: las posiciones “verdaderas” no observables, los parámetros de movimiento de un modelo de proceso específico y los estados de comportamiento ocultos en los datos observables.

El modelo se basa en el uso de caminatas aleatorias correlacionadas a primera diferencia (CACD) como modelo de movimiento. Las CACD asumen una caminata aleatoria correlacionada en las diferencias de las posiciones sucesivas, no en las posiciones previas propiamente. Esto tiene cierto sentido intuitivo, pues se desea modelar cómo los animales cambian su velocidad y dirección, no cómo cambian su posición *per se*. La CACD tiene la siguiente forma [94]:

$$\mathbf{d}_t \sim N_2[\gamma \mathbf{T}(\theta) \mathbf{d}_{t-1}, \Sigma] \quad (2.13)$$

donde  $\mathbf{d}_{t-1}$  es la diferencia entre las posiciones  $\mathbf{x}_{t-1}$  y  $\mathbf{x}_{t-2}$ .  $\mathbf{T}(\theta)$  es la matriz de transición que da el ángulo de rotación necesario para pasar de  $\mathbf{d}_{t-1}$  a  $\mathbf{d}_t$ , donde  $\theta$  es el ángulo de giro promedio.  $N_2$  es una distribución normal bivariada con matriz de covarianza  $\Sigma$ . Se incluye  $\gamma$  para permitir la variabilidad en la autocorrelación de la dirección y la velocidad [94]. Entonces decimos que la diferencia  $\mathbf{d}_t$  entre las posiciones sucesivas  $\mathbf{x}_t$  y  $\mathbf{x}_{t-1}$  es una variable aleatoria que se distribuye conforme la normal  $N_2$  definida anteriormente.

Con el modelo arriba explicado se pueden estimar los estados no observados correspondientes a cada posición observada  $\mathbf{x}_t$  y los parámetros de la CACD  $\gamma$ ,  $\theta$  y  $\Sigma$ . Para identificar distintos modos de comportamiento en los datos se necesita especificar un proceso para cada uno ellos, se propondrán tantos modos como se cree que haya (esto generalmente se deduce de la experiencia del experto en la especie a estudiar). Se usa pues un subíndice  $b_t$ , que denota un modo de comportamiento, en los parámetros  $\theta$  y  $\gamma$  de la CACD. Por lo tanto:

$$\mathbf{d}_t \sim N_2[\gamma_{b_t} \mathbf{T}(\theta_{b_t} \mathbf{d}_{t-1}, \Sigma)] \quad (2.14)$$

donde  $b_t = k, k \in \{1, \dots, B\}$  y  $B$  el número total de modos de comportamiento. En [94] se consideran sólo 2 posibles modos; el modo *tránsito* y el modo *forrajeo*. Estos estados, o modos, son análogos a los modos definidos por Morales *et al.*, pero dando una formulación matemática diferente para la aproximación de la dinámica (trayectoria). Se tienen que estimar: dos conjuntos de variables no observadas ( $\mathbf{x}_t$  y  $b_t$ ) y los parámetros  $\theta$ ,  $\gamma$  y  $\Sigma$ . Para estimar  $b_t$  se asume que lo que el animal hace en el presente se ve influenciado de alguna forma del pasado, para esto se estiman las dos transiciones siguientes: El parámetro  $\alpha_1$  que es la probabilidad de que el animal esté en tránsito al tiempo  $t$  dado que estaba en tránsito al tiempo  $t-1$ . Y el parámetro  $\alpha_2$  que es la probabilidad de que el animal esté en tránsito al tiempo  $t$  dado

que estaba en forrajeo al tiempo  $t - 1$  [94].

Se utiliza la aproximación anterior para ajustar el modelo a los datos de posiciones de cinco tortugas laúd de las costas del este de Canadá y el noreste de Estado Unidos. Los mapas de estado asociados a los modos de comportamiento [94] indican que dos de las tortugas gastan una proporción considerable de tiempo forrajeando en las aguas de Canadá y el noreste de Estado Unidos. Los resultados muestran, además, que los dos modos considerados en el modelo representan, de hecho, clases distintas de movimiento. Como era de esperarse, los animales forrajeadores típicamente exhiben alguna forma de búsqueda restringida con ángulos de giro grandes y tasas de viaje lentas, las tortugas tienden a revertir su dirección de manera frecuente. Las tortugas muestran también una baja autocorrelación entre sus ángulos de giro y su velocidad, indicando una falta de persistencia en el ángulo de giro y la tasa de viaje entre un paso de tiempo y el siguiente. Estos resultados sirven para identificar de manera más precisa los hábitats de forrajeo de las tortugas laúd [94].

Hasta aquí hemos expuesto dos modelos, similares en el núcleo pero distintos en el procedimiento de aproximación, de movimiento animal que tienen como objetivo analizar datos de telemetría para extraer información de la trayectoria, tales como los modos de movimiento y cómo éstos afectan a diversos fenómenos como la migración y el forrajeo. Estas aproximaciones, y muchas de las presentes en la literatura [4, 95] se enfocan en un análisis a nivel individual, sin embargo, muchos estudios de interés sobre el movimiento de animales están relacionados a un análisis a nivel poblacional. Es decir, para varios individuos, ¿hay evidencia de respuestas de comportamiento consistentes a las variables ambientales? Para obtener una aproximación a nivel poblacional, el enfoque bien aceptado es utilizar un modelo jerárquico con efectos aleatorios para los individuos que se agrupan a nivel de la población [71]. En el marco teórico del capítulo 5 de esta tesis, abordaremos toda la formalidad de los modelos jerárquicos.

### 2.3.3. Modelo poblacional para herbívoros del Serengueti

En [96] Hopcraft *et al.* exponen un modelo que tiene como objetivo determinar cómo la calidad de alimento, su abundancia y el riesgo de depredación influyen en la distancia y dirección que los migrantes del Serengueti escogen para viajar día con día y cómo estas características de movimiento determinan las diferencias entre los ñu y las cebras de este ecosistema. Ya que cada animal rastreado puede considerarse distinto a los otros, pero no estadísticamente independiente, Hopcraft *et al.* consideran en [96] una aproximación jerárquica para modelar los datos de movimiento de los ñu y las cebras para determinar cuáles variables ambientales describen mejor su comportamiento. Basándose en las diferencias fisiológicas entre estas dos especies y los factores que regulan su abundancia en el ecosistema, Hopcraft *et al.* adelantan que: la calidad del alimento afecta al ñu individual más que la cebra, la exposición al riesgo afecta más el movimiento de la cebra que del ñu, y que la cantidad de comida afecta a ambas especies, mucho más en la temporada de sequía cuando los recursos son limitados.

La aproximación que utilizan en [96] es la de estimar los parámetros para una caminata

aleatoria correlacionada espacialmente informada, es decir, además de considerar la inercia de la dirección previa para elegir su nueva dirección, el animal también considerará características espaciales específicas que pueden, o no, modificar su trayectoria. Más que modelar los cambios entre comportamientos discretos (como acampada o migración), determinan los parámetros que gobiernan las distribuciones de los pasos diarios (los desplazamientos entre posiciones consecutivas diarias) y los ángulos de giro como funciones logarítmicas de las variables ambientales. Se adapta esta aproximación para capturar las variaciones individuales entre animales rastreados haciendo uso de modelos jerárquicos. Para modelar la distribución de los pasos diarios se utiliza una distribución Weibull de parámetros  $\alpha$  y  $\beta$  y para los ángulos de giro se utiliza una distribución Cauchy de parámetros  $\mu$  y  $\rho$ , misma aproximación ya utilizada en [80].

La longitud de pasos y los ángulos de giro para los animales individuales se relacionan con las características del ambiente por medio del modelaje del parámetro de escala  $\alpha$  de la distribución Weibull, y la variabilidad y la dirección media de los ángulos de giro ( $\rho$  y  $\mu$ ) como funciones continuas de varias características ambientales como la calidad del alimento, la cantidad de éste, el riesgo de depredación, etc. Aquellas características ambientales que influyen en la reducción de las distancias de desplazamiento (aumento en  $\alpha$ ) y que aumentan la variabilidad en los ángulos de giro (reducen  $\rho$ ), resultan en un patrón de movimiento de búsqueda restringida. En caso contrario, si la interacción con ciertas características ambientales lleva a un aumento en la longitud de paso (disminución en  $\alpha$ ) y una disminución en la variabilidad en los ángulos de giro (aumenta  $\rho$ ), entonces se tendrá el abandono de cierta área del ambiente. Así pues, los parámetros  $\alpha$ ,  $\rho$  y  $\mu$  dependerán de las características del ambiente como sigue:

$$\log(\alpha_t^i) = b_0^i + \sum_{k=1}^8 b_k^i x_{k,t}^i \quad (2.15)$$

$$\text{logit}(\rho_t^i) = c_0^i + \sum_{k=1}^8 c_k^i x_{k,t}^i \quad (2.16)$$

$$\mu_t^i = \text{logit} \left( d_0^i + \sum_{k=1}^8 d_k^i x_{k,t}^i \right) \quad (2.17)$$

con  $\text{logit}(x) = \log(x/1-x)$  y donde los subíndices indican la observación del día  $t$  del  $i$ -ésimo individuo (superíndice) y hay ocho variables ambientales predictivas ( $x_k$ ) que son medidas al mismo tiempo en el mismo día  $t$  para el animal  $i$ . Los interceptos de la regresión  $b_0$ ,  $c_0$  y  $d_0$  se muestrean de una distribución a nivel poblacional; por lo tanto, aunque cada animal cuenta con un valor específico para cada uno de sus respectivos interceptos, dichos interceptos retratan un comportamiento a nivel poblacional y no individual. El parámetro de forma de la distribución Weibull  $\beta$  también es estimado a nivel poblacional pero no depende de las condiciones ambientales. Además, los coeficientes de regresión  $b_k$ ,  $c_k$  y  $d_k$  para  $\alpha$ ,  $\rho$  y  $\mu$  (respectivamente) se estiman también a nivel poblacional, permitiendo así variaciones de las longitudes de paso y ángulos de giro entre animales en respuesta a cada una de las ocho

variables ambientales [96].

Este tipo de análisis es útil para determinar cómo la depredación y la abundancia de alimento afectan la dirección de los giros ( $\mu$ ), la tortuosidad del movimiento ( $\rho$ ), y la distancia recorrida diaria ( $\alpha$ ). En el caso particular de estudio de [96], los ñu tienden a responder a los factores relacionado a alimento, especialmente en las planicies, pero se muestran mucho más reactivos a los factores de depredación en los bosques. En contraste, la cebra equilibra el acceso a alimentos de alta calidad evitando las áreas de riesgo asociadas con los depredadores y los humanos cuando los recursos son abundantes en las planicies, pero cuando los recursos disminuyen y se ven obligados a ingresar a los bosques, la cebra se enfoca en acceder a suficiente comida en parches de hierba de alta biomasa. Los factores relacionados con los depredadores no afectan el movimiento de la cebra en los bosques tanto como lo hacen en las planicies. Las respuestas de la cebras a nivel individual tiende a variar más que las de los ñus, que se comportan de manera muy similar entre sí. Por ejemplo, la proximidad a las fuentes de alimento altas en nitrógeno es una covariable importante para predecir el desplazamiento diario de la cebra a nivel poblacional [96].

Hemos presentado hasta aquí tres formulaciones matemáticas distintas que tienen un objetivo común: identificar, dentro de un conjunto de datos observados, distintos modos, o estados, de movimiento que definen comportamientos animales diferentes y que permiten entender cómo dichos tipos de comportamiento influyen durante el desarrollo de fenómenos tales como el forrajeo y la migración. Este tipo de análisis nos proporciona también una herramienta para caracterizar distintas especies animales por sus modos de comportamiento y el tiempo que permanecen en cada uno de ellos, dicha caracterización nos permite cuantificar las diferencias de movimiento entre una especie y otra, como en el caso del trabajo presentado brevemente en esta última subsección. El enfoque presentado aquí nos servirá de inspiración para el modelo que presentaremos más adelante en el capítulo cinco y en el cual agregamos, además de algunas características ya mencionadas en esta sección, el uso de memoria y cómo ésta puede caracterizar, de manera intrínseca, dos tipos de movimiento distintos. Exponemos a continuación los principales objetivos generales y particulares que deseamos alcanzar con este trabajo.

## 2.4. Objetivos Generales

- i) Entender, matemáticamente, el efecto de la memoria sobre la difusión de modelos simples. Surgimiento de difusión no browniana, subdifusión o localización en medios heterogéneos.
- ii) Determinar en qué condiciones la memoria es ventajosa para un organismo que recorre un medio espacialmente heterogéneo, comparado con reglas Markovianas y si el uso de memoria puede llevar a la adaptación por medio del aprendizaje.
- iii) Determinar si el uso de memoria y transferencia de información pueden facilitar un aprendizaje colectivo en problemas de búsqueda en ambientes complejos.

- iv) Ajustar y predecir trayectorias reales de Ciervos Canadienses con distintos modelos bayesianos basados en memoria.

## 2.5. Objetivos Particulares

- i) Resolver analíticamente un modelo de caminata con memoria en un ambiente que contiene una heterogeneidad. Es de interés estudiar tanto el caso en el que dicha heterogeneidad representa un sitio de atrapamiento (el individuo permanece en ella cierto tiempo) como el caso en el que este sitio particular repele la trayectoria del individuo, ya sea por movimiento aleatorio o uso de memoria.
- ii) Considerar distribuciones de salto o dimensiones espaciales tales que el proceso Markoviano subyacente (en ausencia de memoria) sea transiente o recurrente y estudiar el efecto sobre la adaptación.
- iii) Extender el algoritmo a varios caminantes y varias inhomogeneidades espaciales con el fin de probar el éxito en la búsqueda y explotación de los recursos. Agregar un parámetro que describa la probabilidad de que dos caminantes se *comuniquen* con el fin de transferir información sobre posibles sitios de interés.
- iv) Utilizar modelos bayesianos para ajustar y predecir las trayectorias de 21 Ciervos Canadienses. Se utilizarán cuatro distintos modelos para este fin y se hará un estudio estadístico exhaustivo para identificar los mecanismos básicos de uso de memoria que llevan al patrón de uso de espacio observado para cada individuo y sus consecuencias en el aprendizaje.



## Capítulo 3

# Modelo mínimo: un caminante, una inhomogeneidad

Como se ha mencionado ya en anteriores capítulos de esta tesis, muchas especies animales tienen la capacidad de utilizar la información que han obtenido al interactuar con su ambiente para tomar decisiones en su presente. Es decir, la interacción con el ambiente y la capacidad de memoria que poseen (habilidad para volver a sitios previamente visitados y que se encuentran, posiblemente, fuera de su rango de visión) les permite desarrollar nuevos comportamientos adaptativos como resultado de su experiencia.

Dentro del ámbito matemático existen diversos tipos de modelos que intentan encontrar los mecanismos básicos que definen este tipo de procesos de memoria y aprendizaje, tal es el caso de las CA con visitas preferenciales; las cuales resultan matemáticamente retardoras, y las preguntas básicas sobre bajo qué condiciones los comportamientos asintóticos de la caminata son difusivos o localizados, se vuelven muy complicadas de responder [30–32]. Menos entendido es cómo un espacio heterogéneo (compuesto de recursos) puede afectar a los procesos de difusión y localización. Existen varios modelos motivados en la biología que han sido estudiados numéricamente y que han intentado dar luz a estas preguntas [21, 25, 83, 97].

En este capítulo presentamos un modelo que se puede resolver analíticamente en una aproximación bien definida que puede resultar ser exacta en el límite de tiempos largos. Este modelo que combina el movimiento aleatorio con un esquema de reforzamiento lineal, permitiéndonos entender cómo el aprendizaje espacial puede emerger durante una búsqueda. El esquema de reforzamiento lineal y la dinámica de reubicación en las cuales se basa nuestro modelo son los mismos ya expuestos en el capítulo 2.1 de este mismo escrito. El modelo propuesto es el siguiente:

### 3.1. Modelo

Consideremos un caminante con posición  $\mathbf{X}_t$  en una rejilla cúbica  $d$ -dimensional infinita uniespaciada, donde la variable temporal  $t$  es discreta y la posición inicial del caminante

viene dada por  $\mathbf{X}_0 = \mathbf{x}_0$ . La rejilla contiene una inhomogeneidad (o sitio de atrapamiento), representando un pozo de agua o un parche de comida, ubicada en el origen. Dependiendo de su posición en el espacio, el caminante aleatorio puede seguir dos tipos de dinámica. (i) Puede difundirse con una dinámica reforzada que consiste en reubicarse a sitios visitados en el pasado por medio del reinicio [21, 85, 86], o (ii) puede permanecer *atrapado* en el origen durante algún tiempo (Fig.3.1). De forma más precisa, a cada paso de tiempo  $t \rightarrow t + 1$ :

- (a) Si el caminante no se encuentra en la inhomogeneidad, con probabilidad  $1 - q$  selecciona un desplazamiento aleatorio  $\ell_t$  dado por una distribución simétrica  $p(\ell)$  y  $\mathbf{X}_{t+1} = \mathbf{X}_t + \ell_t$  (*movimiento aleatorio*). Con probabilidad complementaria  $q$  el caminante regresa a un sitio visitado en el pasado, esto es  $\mathbf{X}_{t+1} = \mathbf{X}_{t'}$  donde  $t'$  es un entero aleatorio escogido de forma uniforme del intervalo  $[0, t]$ . Por lo tanto, la probabilidad de escoger un sitio particular para reubicarse es proporcional al tiempo total acumulado gastado en ese sitio (*reforzamiento lineal*). Este esquema de reforzamiento es el mismo expuesto en [85] ya explicado con anterioridad.
- (b) Si el caminante se encuentra en la inhomogeneidad ( $\mathbf{X}_t = \mathbf{0}$ ) se queda en ella al tiempo  $t+1$  con probabilidad  $\gamma$ , o se mueve acorde a las reglas descritas en (a) con probabilidad  $1 - \gamma$ .

Definiendo  $\text{Prob}[\mathbf{X}_{t'} = \mathbf{n} \text{ and } \mathbf{X}_t = \mathbf{0}]$  como la probabilidad conjunta de estar en el sitio  $\mathbf{n}$  al tiempo  $t'$  y en el origen al tiempo  $t$ , podemos escribir la dinámica anterior como sigue:

$$\begin{aligned}
P_{\mathbf{n}}(t+1) &= (1-q) \sum_{\ell} p(\ell) (1 - \gamma_{\mathbf{n}-\ell}) P_{\mathbf{n}-\ell}(t) + \gamma_{\mathbf{n}} P_{\mathbf{n}}(t) \\
&+ \frac{q(1-\gamma)}{t+1} \sum_{t'=0}^t \text{Prob}[\mathbf{X}_{t'} = \mathbf{n} \text{ and } \mathbf{X}_t = \mathbf{0}] \\
&+ \frac{q}{t+1} \sum_{t'=0}^t \text{Prob}[\mathbf{X}_{t'} = \mathbf{n} \text{ and } \mathbf{X}_t \neq \mathbf{0}]
\end{aligned} \tag{3.1}$$

donde  $P_{\mathbf{n}}(t) = \text{Prob}[\mathbf{X}_t = \mathbf{n}]$  y  $\gamma_{\mathbf{n}} = \gamma \delta_{\mathbf{n}, \mathbf{0}}$ .

Los primeros dos términos del lado derecho de la Ec.(3.1) describen el movimiento aleatorio y de *permanencia* del caminante, respectivamente. Los últimos dos términos de la Ec.(3.1) se refieren a la probabilidad de reubicación al sitio  $\mathbf{n}$  (si éste ha sido visitado en un tiempo previo  $t'$ ) desde un sitio que puede ser la inhomogeneidad  $\mathbf{0}$  o cualquier otro. El término  $1/(t+1)$  es la distribución de probabilidad uniforme de la variable  $t'$ . La ecuación (3.1) describe un proceso no Markoviano con memoria infinita que evoluciona en un medio heterogéneo, siendo claramente una ecuación no cerrada. Esta ecuación se deriva de la misma manera que la ecuación maestra desarrollada por Boyer y Solis-Salas en [85].

En ausencia de la heterogeneidad espacial ( $\gamma = 0$ ) el modelo presenta sub-difusión para cualquier probabilidad de reubicación o uso de memoria ( $0 < q < 1$ ). La distribución  $P_{\mathbf{n}}$  tiende a una Gaussiana, como en la difusión normal, pero con una varianza que sigue la ley

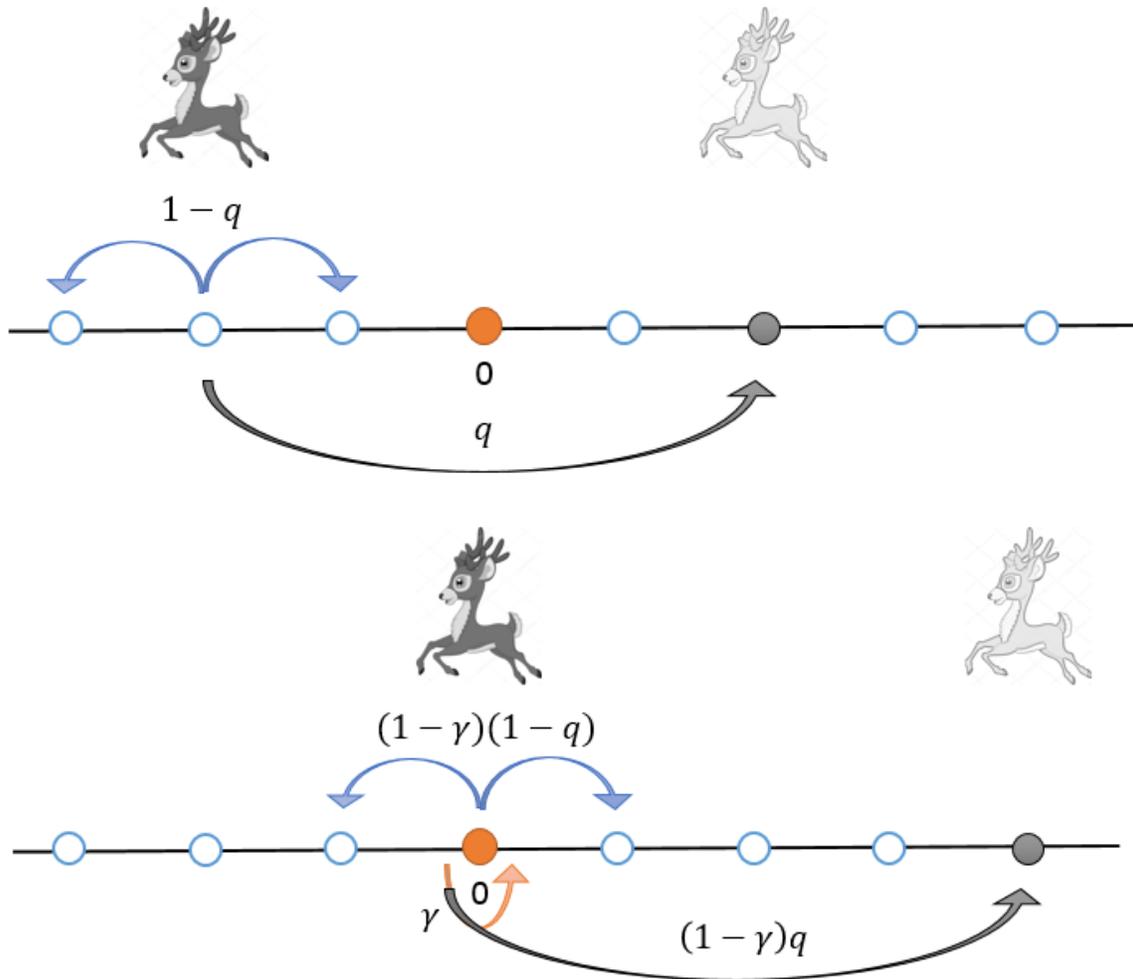


Figura 3.1: Diagrama del modelo mínimo que describe las reglas de movimiento (a)-(b) de un caminante aleatorio en una rejilla que contiene una heterogeneidad en el origen. **Arriba:** Cuando  $\mathbf{X}_t \neq \mathbf{0}$ , el caminante toma un paso aleatorio con probabilidad  $1 - q$  (flechas azules) y usa su memoria de revisita preferencial con probabilidad  $q$  (flecha gris). **Abajo:** Cuando  $\mathbf{X}_t = \mathbf{0}$ , el caminante elige quedarse en ese sitio un paso más de tiempo con probabilidad  $\gamma$  y se mueve siguiendo las reglas arriba descritas con probabilidad  $1 - \gamma$ .

anómala logarítmica en lugar de la ley normal en la que  $\langle X_t^2 \rangle \propto t$ . Así  $\lim_{t \rightarrow \infty} P_{\mathbf{n}} = 0 \forall \mathbf{n}$  [85]. Cuando  $\gamma \neq 0$ , cada visita al origen tiende a durar más que en cualquier otro sitio. Aquí surge la pregunta de si el reforzamiento puede suprimir la difusión completamente y llevar la dinámica hacia un NESS, uno similar al que presentan Evans y Majumdar en [33]. A nuestro NESS llamémosle  $P_{\mathbf{n}} \equiv \lim_{t \rightarrow \infty} P_{\mathbf{n}}(t) \neq 0$ , centrado alrededor del  $\mathbf{0}$  e independiente de la posición inicial del caminante  $\mathbf{x}_0$ . Cuando el estado NESS es alcanzado decimos que el caminante se ha localizado como resultado de la adaptación por aprendizaje.

La Fig.3.2 muestra de manera esquemática lo anterior. En la Fig.3.2 - Abajo, dos capturas de pantalla a tiempos iniciales y finales de la dinámica en  $2d$  (resp.), en ellas se muestra en naranja la posición inicial del caminante y en rojo la heterogeneidad con peso  $\gamma = 0.9$ , se fija el parámetro de uso de memoria en  $q = 0.2$ . La Fig.3.2 - Arriba, el número de visitas a la posición inicial (triángulos verdes) y a la inhomogeneidad (rombos azules) como función de  $t$ . Resulta claro que una vez que el caminante encuentra el sitio de interés, el número de visitas a su posición inicial alcanza un valor asintótico y el número de visitas a la heterogeneidad crece de manera rápida, es decir, el individuo centra sus trayectorias alrededor de dicho sitio de recurso y lo visita de manera preferencial. Ha adaptado su comportamiento al interactuar con su ambiente, emergiendo así el aprendizaje.

## 3.2. Aproximación de decorrelación

Ahora bien, obtener una solución exacta para el modelo resulta complicado ya que se consideran muchos factores que vuelven desafiante dicha solución como por ejemplo, un espacio heterogéneo, el uso de memoria y funciones en la ecuación de movimiento que consideran dos tiempos  $t$  y  $t'$ . Así, con el fin de encontrar ideas válidas sobre la dinámica de aprendizaje de nuestro caminante, consideremos una versión aproximada del modelo que transforma la Ec.(3.1) en una ecuación cerrada. Supongamos que a tiempos largos  $\mathbf{X}_{t'}$  y  $\mathbf{X}_t$  están decorrelacionadas:

$$\text{Prob}[\mathbf{X}_{t'} = \mathbf{n} \text{ and } \mathbf{X}_t \neq \mathbf{0}] \simeq P_{\mathbf{n}}(t')[1 - P_{\mathbf{0}}(t)]. \quad (3.2)$$

De forma similar,  $\text{Prob}[\mathbf{X}_{t'} = \mathbf{n} \text{ and } \mathbf{X}_t = \mathbf{0}] \simeq P_{\mathbf{n}}(t')P_{\mathbf{0}}(t)$ . Reemplazando estas expresiones en Ec.(3.1) y sustituyendo  $P_{\mathbf{n}}(t)$  y  $P_{\mathbf{n}}(t')$  por  $P_{\mathbf{n}}$  en el límite  $(t, t') \rightarrow \infty$ , obtenemos una ecuación para el NESS, si existe:

$$\begin{aligned} P_{\mathbf{n}} &= (1 - q) \sum_{\boldsymbol{\ell}} p(\boldsymbol{\ell}) P_{\mathbf{n} - \boldsymbol{\ell}} + q P_{\mathbf{n}} (1 - \gamma P_{\mathbf{0}}) \\ &+ \gamma P_{\mathbf{0}} [\delta_{\mathbf{n}, \mathbf{0}} - (1 - q) p(\mathbf{n})], \end{aligned} \quad (3.3)$$

y válida para cualquier número de dimensiones.

Denotemos como  $P_{\mathbf{0}} \equiv P_{\mathbf{0}}$  la probabilidad asintótica de ocupación de la inhomogeneidad, una cantidad aún por determinar. Usando la transformada discreta de Fourier  $\tilde{P}(\mathbf{k}) \equiv \sum_{\mathbf{n}} P_{\mathbf{n}} e^{-i\mathbf{k} \cdot \mathbf{n}}$  [98, 99], la ecuación (3.3) se puede reescribir como sigue:

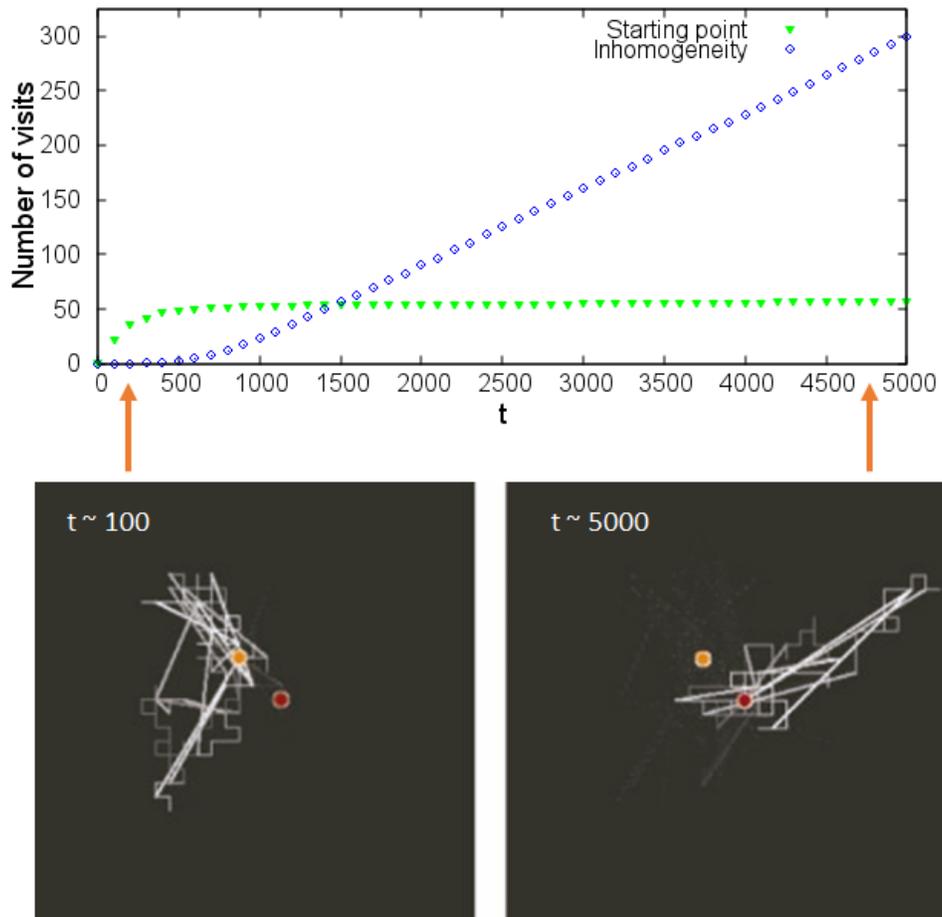


Figura 3.2: Número total de visitas a la posición inicial  $\mathbf{x}_0$  (triángulos verdes) y a la inhomogeneidad (rombos azules) como función del tiempo, para un caminante modelo  $2d$  con  $q = 0.2$  y  $\gamma = 0.9$ . A tiempos iniciales de la dinámica, el caminante se difunde lentamente alrededor de  $\mathbf{x}_0$  (círculo naranja en los recuadros). Después de que la inhomogeneidad (círculo rojo) es encontrada, las visitas a ésta se vuelven sostenidas, indicando una localización.

$$\tilde{P}(\mathbf{k}) = \frac{\gamma P_0 [1 - (1 - q)\tilde{p}(\mathbf{k})]}{(1 - q)[1 - \tilde{p}(\mathbf{k})] + q\gamma P_0}. \quad (3.4)$$

con  $\tilde{p}(\mathbf{k})$  la transformada discreta de Fourier de la distribución simétrica que define la longitud de los saltos aleatorios  $p(\ell)$ . Es claro que  $P_0 = 0$  es siempre solución de la ecuación anterior. Para encontrar soluciones distintas de la trivial, la constante  $P_0$  es determinada de forma auto-consistente por la transformada inversa de (3.4) evaluada en  $\mathbf{n} = \mathbf{0}$ , es decir,

$$P_0 = (2\pi)^{-d} \int_{\mathcal{B}} d^d \mathbf{k} \tilde{P}(\mathbf{k}) \quad (3.5)$$

donde  $\mathcal{B}$  es la primera zona de Brillouin:  $-\pi < k_i < \pi$  para  $i = 1, \dots, d$ , con  $d$  la dimensión de la rejilla en la que se mueve el caminante. Reordenando términos obtenemos que cualquier solución  $P_0 \neq 0$  viene dada por la ecuación trascendental:

$$\frac{1}{(2\pi)^d} \int_{\mathcal{B}} \frac{d^d \mathbf{k}}{(1 - q)[1 - \tilde{p}(\mathbf{k})] + q\gamma P_0} = \frac{1 - \gamma}{q\gamma(1 - \gamma P_0)}. \quad (3.6)$$

Fijando  $\gamma > 0$ , el modelo exhibe una transición de fase si existe una  $q$  crítica  $q_c \in (0, 1)$  tal que Ec.(3.6) no tenga ninguna raíz para  $q < q_c$  (en tal caso, sólo la solución trivial  $P_0 = 0$  existe). Así, considerando el valor mínimo que puede tomar la probabilidad de ocupación de la inhomogeneidad,  $P_0 = 0$ , y sustituyéndolo en (3.6) obtenemos una expresión para  $q_c$ , a saber:

$$q_c = \frac{(1 - \gamma)P_{no-return}}{\gamma + (1 - \gamma)P_{no-return}}, \quad (3.7)$$

donde

$$P_{no-return} = (2\pi)^d \left[ \int_{\mathcal{B}} d^d \mathbf{k} \frac{1}{1 - \tilde{p}(\mathbf{k})} \right]^{-1} \quad (3.8)$$

es la bien conocida probabilidad de que una caminata aleatoria Markoviana en una rejilla infinita nunca regrese a su sitio de partida [98].

Por lo tanto, una transición delocalización/localización puede existir a un  $q_c > 0$  (o  $\gamma_c > 0$  si es  $q$  quien se mantiene fijo) si  $P_{no-return} > 0$ , i.e., si el proceso Markoviano subyacente ( $q = 0$ ) nunca regresa a su sitio de partida, tal como una caminata aleatoria a primeros vecinos en  $d \geq 3$ . De forma análoga, un proceso recurrente como una caminata aleatoria en  $1d$  o  $2d$  tiene  $P_{no-return} = 0$  por lo que nuestro modelo (si la aproximación empleada resulta correcta) admite soluciones localizadas para cualquier  $q > 0$ . Debemos hacer énfasis en que para una caminata aleatoria pura ( $q = 0$ ), una sola inhomogeneidad con cualquier  $\gamma > 0$  no es suficiente para que el caminante se localice alrededor de ésta.

### 3.3. Dinámica a primeros vecinos

Se puede calcular la transformada inversa de (3.4) de forma exacta para el caso particular de visitas a primeros vecinos y en 1d.

Sabemos ya que para el caso de la dinámica a primeros vecinos,  $p(\ell) = (1/2)(\delta_{\ell,1} + \delta_{\ell,-1})$ , con  $\delta$  definida por la función  $\delta$  de Kronecker. Por lo tanto,  $\tilde{p}(k) = \cos(k)$ . Sustituyendo la expresión anterior en Ec.(3.4) tenemos que:

$$\tilde{P}(k) = \frac{\gamma P_0 [1 - (1 - q) \cos k]}{(1 - q)(1 - \cos k) + q\gamma P_0} = \gamma P_0 + \frac{q\gamma P_0 (1 - \gamma P_0)}{(1 - q)(1 - \cos k) + q\gamma P_0}. \quad (3.9)$$

La forma de la probabilidad de ocupación en el estado estacionario puede derivarse utilizando la transformada de Fourier inversa. Usando el hecho de que  $a^2 > 1$  tenemos que [100]:

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} dk \frac{\cos(kn)}{1 + a^2 - 2a \cos k} = \frac{1}{(a^2 - 1)a^{|n|}}, \quad (3.10)$$

escribiendo el denominador  $(1 - q)(1 - \cos k) + q\gamma P_0$  bajo la forma  $b(1 + a^2 - 2a \cos k)$ , y por simple comparación, obtenemos que:

$$2ab = 1 - q \quad (3.11)$$

$$b(1 + a^2) = 1 - q(1 - \gamma P_0) \quad (3.12)$$

lo cual nos lleva a una expresión directa para  $a$ :

$$a = 1 + \frac{\gamma q P_0}{1 - q} + \sqrt{\frac{\gamma q P_0}{1 - q} \left( 2 + \frac{\gamma q P_0}{1 - q} \right)}. \quad (3.13)$$

Utilizando las Ecs.(3.10)-(3.12), la inversa de la Ec.(3.9) está dada por:

$$P_n = \gamma P_0 \delta_{n,0} + \frac{q\gamma P_0 (1 - \gamma P_0)}{1 - q} \frac{2a}{(a^2 - 1)a^{|n|}}. \quad (3.14)$$

Ahora bien, como  $P_{n=0}$  debe ser igual a la constante  $P_0$  por autoconsistencia, deducimos pues de (3.14) que:

$$P_n = \gamma P_0 \delta_{n,0} + (1 - \gamma) P_0 a^{-|n|}, \quad (3.15)$$

Nos queda sólo encontrar una forma exacta para la probabilidad asintótica de ocupación de la inhomogeneidad ( $P_0$ ). Evaluemos, entonces, una vez más la Ec.(3.14) en  $n = 0$ , obteniendo:

$$2q\gamma(1 - \gamma P_0) = (1 - \gamma)(1 - q)(a - a^{-1}). \quad (3.16)$$

Ahora bien, sustituyendo la expresión para  $a$  dada por la Ec.(3.13) en la expresión  $(a - a^{-1})$  en la parte derecha de la Ec.(3.16) obtenemos que:

$$2q\gamma(1 - \gamma P_0) = (1 - \gamma)(1 - q) \left[ 2\sqrt{\frac{q\gamma P_0}{1 - q} \left( 2 + \frac{q\gamma P_0}{1 - q} \right)} \right]. \quad (3.17)$$

Despejando y aplicando cuadrados a ambos lados de la igualdad para deshacernos de la raíz obtenemos la siguiente ecuación cuadrática para  $P_0$ :

$$[q\gamma(1 - 2\gamma)]P_0^2 + [2((1 - q)(1 - \gamma)^2 + q\gamma^2)]P_0 - q\gamma = 0. \quad (3.18)$$

La cual tiene una sola raíz positiva, a saber:

$$P_0 = \frac{-(1 - q)(1 - \gamma)^2 - q\gamma^2}{q\gamma(1 - 2\gamma)} + \frac{\sqrt{[(1 - q)(1 - \gamma)^2 + q\gamma^2]^2 + (q\gamma)^2(1 - 2\gamma)}}{q\gamma(1 - 2\gamma)}, \quad (3.19)$$

para  $\gamma \neq 1/2$ . Cuando  $\gamma = 1/2$  la solución es simplemente  $P_0 = q$ .

La Ec.(3.19) muestra que  $P_0 > 0$  para cualquier  $0 < \gamma < 1$  y  $0 < q < 1$ : tal y como lo habíamos mencionado anteriormente *existen siempre soluciones localizadas en 1d* para cualesquiera intensidad de memoria y peso en la inhomogeneidad. A partir de la Ec.(3.15) vemos que la probabilidad de ocupación decae exponencialmente con la distancia al origen. Para comparar los resultados analíticos con las simulaciones de la dinámica; en vez de resolver numéricamente la Ec.(3.1), lo cual es complicado, realizamos simulaciones de Monte-Carlo para las reglas (a)-(b) del modelo.

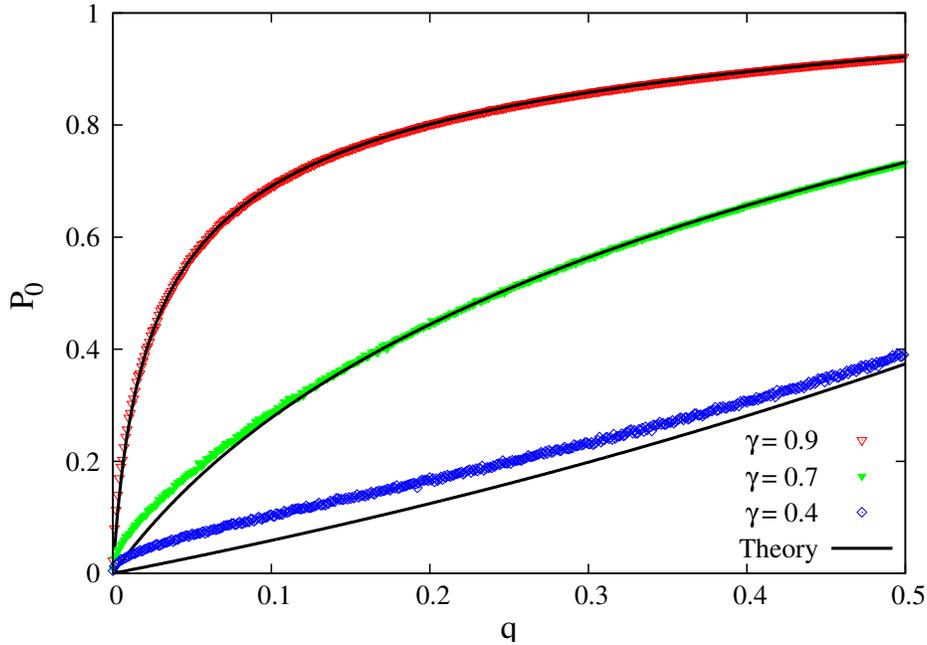


Figura 3.3:  $P_0$  como función de  $q$  en  $1d$  con saltos hacia primero vecinos. Las líneas sólidas están dadas por Ec.(3.19) y los símbolos por simulaciones de Monte Carlo de las reglas (a)-(b) (con  $t = 10^5$ ).

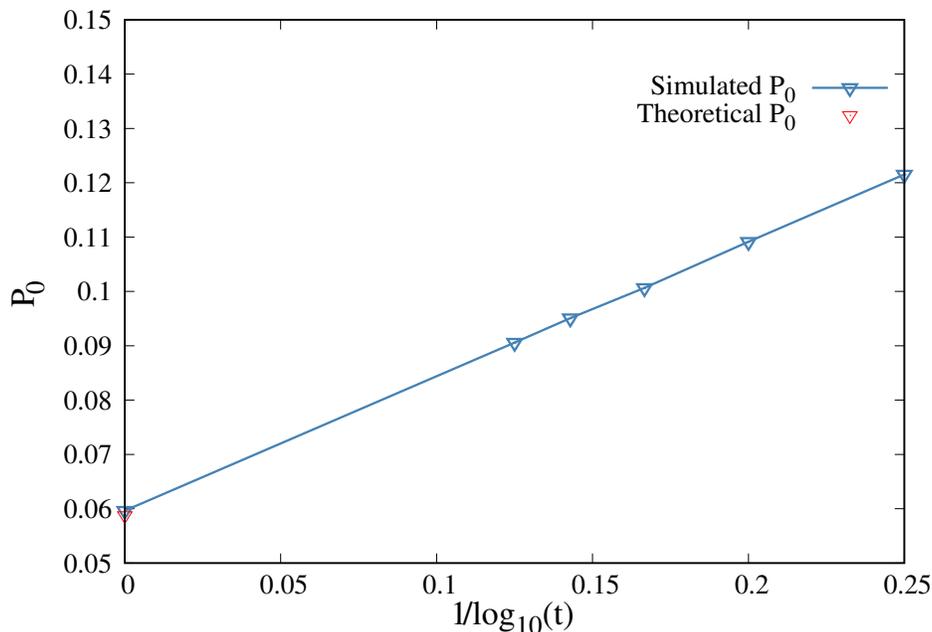


Figura 3.4: Extrapolación de  $P_0$  a  $t \rightarrow \infty$  tomando como base los valores aproximados de  $P_0$  calculados a tiempos finitos (triángulos azules). El valor teórico de  $P_0$  se muestra con un triángulo rojo sobre el eje y. Se toma, para mejor lectura de la gráfica, una escala de  $1/\log$  en el eje x.

La Fig.3.3 muestra el cambio de  $P_0$  como función de  $q$  para diferentes valores de  $\gamma$  siguiendo la Ec.(3.19) (línea sólida negra) y siguiendo los resultados de los experimentos numéricos (símbolos). El error que se observa para  $q$  y  $\gamma$  pequeñas se atribuye a que el régimen temporal asintótico es muy difícil de alcanzar en las simulaciones, ya que deberíamos esperar tiempos muy grandes para llegar a él. Analicemos más a detalle esto último: En la Fig.3.3 es claro que para el caso  $\gamma = 0.4$  el error para  $q$  pequeñas es significativo, alcanzando (en particular) alrededor del 33% para  $q = 0.1$ . Para comprobar que dicho mal ajuste se debe, precisamente, a la convergencia lenta y no a un error en el supuesto de decorrelación realizamos un experimento numérico extra presentado en la Fig.3.4. En esta figura se muestra la extrapolación de  $P_0$  a  $t \rightarrow \infty$  tomando como base los valores aproximados encontrados a tiempos finitos (triángulos azules). Podemos ver que la extrapolación lleva a un valor de  $P_0$  muy cercano a su valor teórico (triángulo rojo). Con esto podemos decir que la aproximación de decorrelación que utilizamos para llegar a las ecuaciones ya descritas, resulta muy buena, lo que sugiere que este supuesto de decorrelación es cierto para tiempos largos.

Ahora, analicemos uno de los casos en  $1d$  en el cual  $P_{no-return} > 0$  [98] lo cual implica una transición de fase a una  $q_c$  diferente de cero.

### 3.4. Dinámica con vuelos de Lévy

Recordemos que la probabilidad de no retorno a la posición inicial del caminante viene dada por [99]:

$$P_{no-return} = \frac{1}{\int_{\mathcal{B}} \frac{d^d \mathbf{k}}{(2\pi)^d} \frac{1}{1-\tilde{p}(\mathbf{k})}} \quad (3.20)$$

Así,  $P_{no-return}$  en Ec.(3.20) es cero o diferente de cero dependiendo si la integral en el denominador es infinita o finita, respectivamente. La divergencia de esta integral depende del comportamiento de  $\tilde{p}(\mathbf{k})$  a  $|\mathbf{k}|$  pequeñas. Para vuelos de Lévy simétricos y sin sesgo, el comportamiento a  $\mathbf{k}$  pequeñas está dado por:

$$\tilde{p}(\mathbf{k}) \simeq 1 - K_\mu |\mathbf{k}|^\mu \quad (3.21)$$

donde el índice de Lévy  $\mu \in (0, 2]$  y  $K_\mu$  una constante [101].

Para  $\mu < 2$ , el segundo momento de la distribución de saltos diverge debido al decaimiento en ley de potencia para  $p(\ell)$  para  $\ell$  grandes que va como [101]:

$$p(\ell) = C/|\ell|^{1+\mu}, \quad \ell = \pm 1, \pm 2, \pm 3, \dots, \quad (3.22)$$

con  $C$  una constante de normalización. El caso  $\mu = 2$  corresponde a caminatas aleatorias con difusión normal, es decir con pasos de longitud de varianza finita. Así, las caminatas aleatorias se caracterizan por  $\tilde{p}(\mathbf{k}) = 1 - D_0 |\mathbf{k}|^2$  para  $k$  pequeños, con  $D_0 = 1/2$  el coeficiente de difusión. De ahora en adelante, consideraremos el caso general con  $\mu \in (0, 2]$ , el cual incluye el caso  $\mu = 2$  que corresponde a la caminata aleatoria simple con saltos a primeros vecinos. Sustituyendo (3.21) en el denominador de la Ec.(3.20) es claro que la integral diverge si  $d < \mu$  y es finita si  $d > \mu$ . Por lo tanto,  $P_{no-return} = 0$  para  $d < \mu$ , mientras que es diferente de cero para  $d > \mu$ . Así, para vuelos de Lévy con índice  $0 < \mu \leq 2$ , la dimensión crítica es  $d_c = \mu$  que separa el comportamiento recurrente ( $d < \mu$ ) del transiente ( $d > \mu$ ). Para las caminatas aleatorias simples ( $\mu = 2$ ),  $d_c = 2$ .

Entonces, para analizar el caso en  $1d$  en el que existe una transición de fase a un  $q_c$  finito bastará con considerar una distribución de saltos dada por una ley de potencia definida en (4.3) con índice  $0 < \mu \leq 1$ . La Fig.3.5 muestra a  $P_0$  como función de  $q$  (manteniendo  $\mu$  fija en  $1/2$ ), dada por la aproximación numérica (línea sólida negra) de la Ec.(3.6) y por las simulaciones de Monte Carlo de la dinámica del caminante (símbolos), obteniéndose un buen acuerdo entre las comparaciones. La Fig.3.5 - Izquierda muestra las variaciones de  $P_0$  con  $q$  para  $\gamma = 0.9$  fijo y para distintos valores de  $t$ . Se hace claro que al aumentar la duración de la caminata ( $t$ ), el acuerdo entre la teoría y la simulación se hace cada vez mejor y el cálculo numérico del punto crítico  $q_c$  se hace más preciso, es decir, a tiempos más largos de la simulación, las variaciones de  $P_0$  alrededor del punto crítico se vuelven más abruptas. La línea naranja es el resultado de una prueba de robustez al modelo que se explicará más adelante. En la Fig.3.5 - Derecha se muestra la aproximación numérica de la Ec.(3.6) para distintos valores de  $\gamma$  y la comparación con sus respectivas simulaciones (símbolos). Resulta evidente que conforme disminuye el valor de  $\gamma$ , aumenta el valor de  $q_c$ . Además, conforme

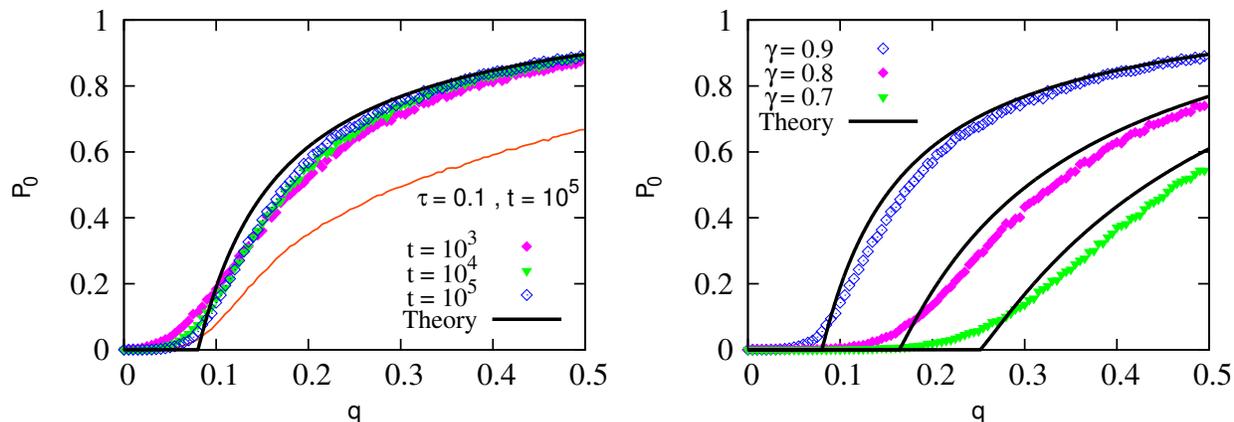


Figura 3.5: Transición de fase en  $1d$  para vuelo de Lévy con  $\mu = 1/2$ . **Izquierda:** La línea sólida negra está dada por (3.6) para el límite  $t = \infty$ , y los símbolos por las simulaciones a diferentes tiempos  $t$  ( $\gamma = 0.9$ ). La línea sólida naranja es una prueba de robustez que se hizo considerando un sitio de recurso consumible que se regenera a tasa  $\tau$ , este resultado se explica a detalle al final de esta sección. **Derecha:**  $P_0$  para  $\gamma = 0.9, 0.8$  y  $0.7$ , donde la teoría predice  $q_c \approx 0.0803, 0.1642$  y  $0.2519$ , respectivamente ( $t = 10^5$  en las simulaciones).

disminuye  $\gamma$ , el acuerdo entre teoría y simulación se hace menor, esto es (similar al caso de primeros vecinos) debido a la lenta convergencia al régimen temporal asintótico.

De la Ec.(3.7) realizamos un diagrama de fase en  $1d$  en el plano  $(\gamma, q)$  en la Fig.3.6. La curva punteada verde representa la línea de los puntos críticos  $q_c$  para  $\mu = 0.9$  conforme  $\gamma$  varía. Así, un proceso caracterizado por  $P_{no-return} \rightarrow 1$ , puede obtenerse tomando el límite  $\mu \rightarrow 0$  en  $1d$  o bien el límite  $d \rightarrow \infty$  (donde se espera que cualquier proceso se vuelva fuertemente transiente). Estos casos corresponden a la diagonal  $q_c = 1 - \gamma$ .

### 3.5. Comportamiento de la distribución $P_{\mathbf{n}}$ a $|\mathbf{n}|$ grandes

Ahora bien, una vez conocido  $P_0$ , el comportamiento de  $P_{\mathbf{n}}$  a  $|\mathbf{n}|$  grandes es sencillo de obtener. Para  $\mathbf{k}$  pequeñas la Ec.(3.4) da, para el caso de una caminata aleatoria con segundo momento finito y  $q > q_c$ ;

$$\tilde{P}(\mathbf{k}) \simeq \frac{q^*}{D|\mathbf{k}|^2 + q^*}, \quad \text{con } q^* = q\gamma P_0, \quad (3.23)$$

y  $D = (1 - q)D_0$  un reescalamiento de la constante de difusión.

La transformada inversa de (3.23) decae exponencialmente para  $|\mathbf{n}|$  grandes en todas las dimensiones espaciales [102, 103]. Particularmente, y como se mencionó ya en la sección 2.1 de este escrito, Evans y Majumdar proponen en [33] un modelo continuo de difusión con reinicio al origen a tasa  $r$ . Este problema está descrito por la ecuación maestra:

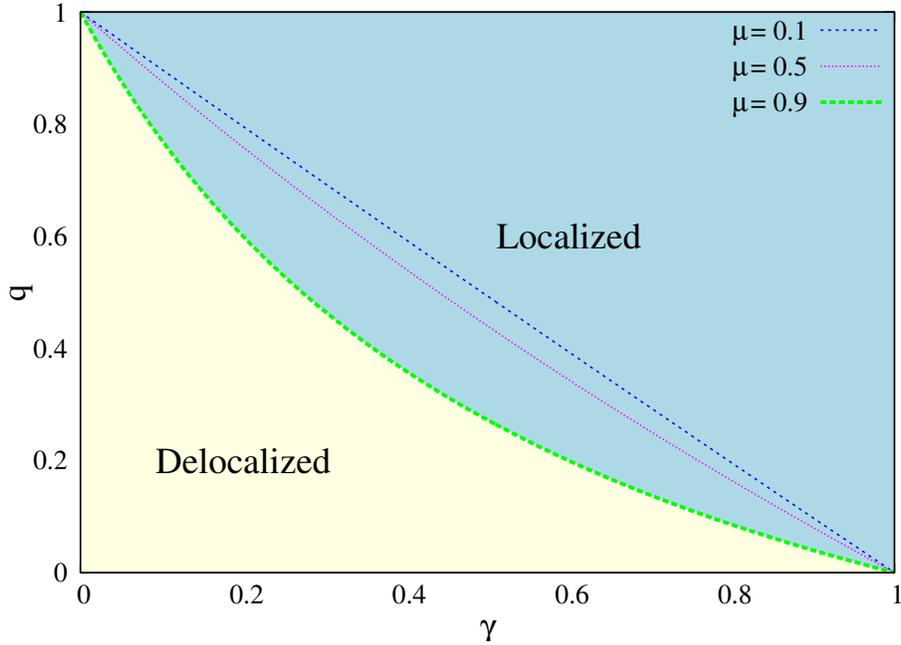


Figura 3.6: Diagrama de fase en 1d para diversos índices de Lévy.

$$\frac{\partial p(x, t|x_0)}{\partial t} = D \frac{\partial^2 p(x, t|x_0)}{\partial x^2} - rp(x, t|x_0) + r\delta(x - x_0) \quad (3.24)$$

con  $p(x, t|x_0)$  la probabilidad de que la partícula esté en la posición  $x$  al tiempo  $t$  dado que su posición inicial fue  $x_0$ ,  $D$  la constante de difusión y  $r$  la tasa de reinicio a la posición de inicio.

Es fácil probar que en el límite a tiempos largos la partícula alcanza una distribución estacionaria dada por:

$$p_{st}(x|x_0) = \frac{\alpha_0}{2} \exp(-\alpha_0|x - x_0|) \quad (3.25)$$

con  $\alpha_0 = \sqrt{r/D}$  una escala de longitud inversa correspondiente a la distancia típica difundida por la partícula entre reinicios. No es difícil probar que la transformada inversa de (3.25) viene dada justamente por:

$$\tilde{p}_{st}(k) = \frac{r}{D|k|^2 + r} \quad (3.26)$$

Así, la distribución dada por (3.23) sigue la forma del NESS dada por la difusión con reinicio estocástico al origen a tasa  $q^*$  [33, 91]. Por lo tanto, gracias al aprendizaje y a tiempos largos, el caminante tiene un comportamiento efectivo similar al de un caminante sin memoria que se reubica *solo* a la inhomogeneidad. La selección del punto de reinicio es una propiedad emergente del modelo y no es impuesta mediante la función delta en la Ec.(3.25) como en [33, 91]. El reinicio efectivo  $q^*$  es proporcional a  $P_0$  y se anula exactamente a  $q = q_c$ , donde el caminante ya no es capaz de adaptarse a su entorno.

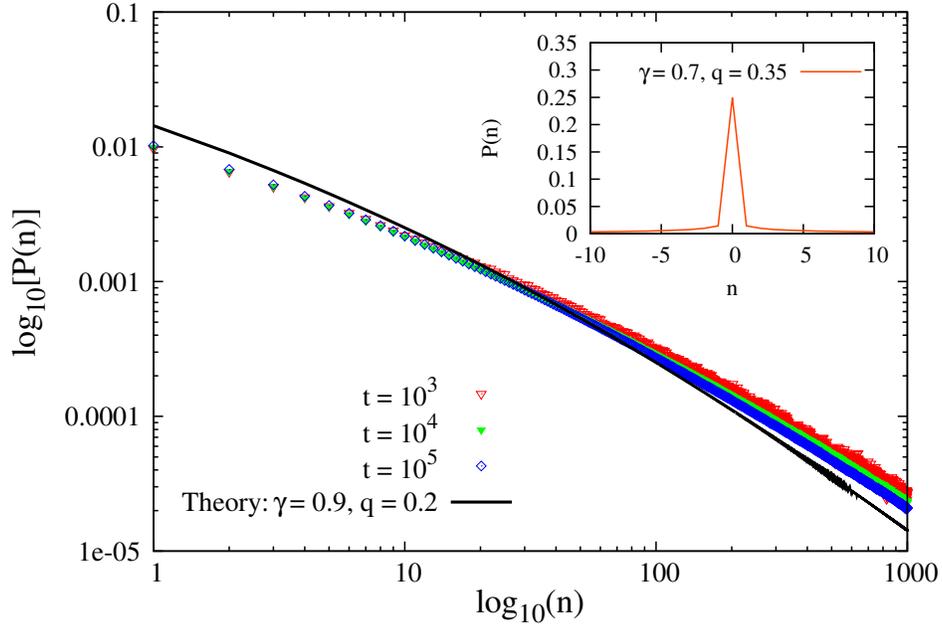


Figura 3.7: Probabilidad de ocupación a la inhomogeneidad. La línea sólida es la transformada inversa de Fourier de la Ec.(3.4) y los símbolos representan las simulaciones a varios tiempos. Escala log-log.

En el caso de vuelos de Lévy,  $D|\mathbf{k}|^2$  se reemplaza por  $K|\mathbf{k}|^\mu$  en (3.23): esta expresión también coincide con el NESS para vuelos de Lévy con reinicio al origen [90, 91]. En  $1d$  y para  $\mu \in (0, 2)$ , la inversión de  $\tilde{P}(k)$  nos da la distribución de Linnik [104]:

$$P_n \simeq \left\{ \frac{K}{q^* \pi} \sin\left(\frac{\pi\mu}{2}\right) \Gamma(\mu + 1) \right\} |n|^{-1-\mu} + R(n) \quad (3.27)$$

con  $|R(n)| < \{(K/q^*)^2 \Gamma(\mu + 1) / \pi \sin(\pi\mu/2)\} |n|^{-1-2\mu}$ . Así, el caminante se localiza conforme una ley de potencia con exponente  $-(1 + \mu)$  a  $|\mathbf{n}|$  grandes. En la Fig.3.7 se muestra la comparación entre la  $P_{\mathbf{n}}$  obtenida por la aproximación numérica de Ec.(3.4) (línea sólida negra) y las simulaciones de Monte Carlo para  $\mu = 0.5$  (símbolos) en escala log-log. Similar a lo que ya se ha expuesto en figuras anteriores, el ajuste entre la curva teórica y los experimentos numéricos se vuelve mejor conforme el tiempo que define la longitud de la caminata aumenta. Es claro que el decaimiento en la probabilidad de ocupación alrededor de la heterogeneidad sigue una ley de potencia, tal como la teoría lo predice. En el inset de la figura se observa la densidad de  $P(n)$  en escalas normales para el caso  $\gamma = 0.7$  y  $q = 0.35$ ,  $\mu$  y  $t$  se mantienen fijos en  $1/2$  y  $10^5$ , respectivamente.

La robustez del fenómeno de localización se puede probar incorporando una ausencia de recurso y regeneración en la regla (b). Asumamos que la inhomogeneidad se vacía al tiempo que el caminante la deja ( $\gamma \rightarrow 0$ ), y se regenera ( $0 \rightarrow \gamma$ ) con probabilidad  $\tau$  por unidad de tiempo. La Fig.3.5-Izquierda muestra la simulación de Monte Carlo para la curva  $P_0$  para  $\tau = 0.1$ , cuya forma es muy similar a las del modelo base, en el sentido de que el  $q_c$  de la

transición entre delocalización/localización es aproximadamente el mismo y, si bien la localización no es tan grande, está claramente presente. Con eso queda probada la robustez del fenómeno de localización.

## 3.6. Exponentes críticos

Para redondear los resultados analíticos hasta ahora expuestos, continuemos con la discusión de las propiedades de  $P_0$  y de la longitud de localización alrededor del punto crítico  $q_c$ . Empecemos con el análisis del parámetro de orden  $P_0$ .

### 3.6.1. Parámetro de orden.

Consideremos el valor crítico  $q_c$  (con  $\gamma$  fijo) que separa la fase delocalizada con  $P_0 = 0$  para  $q < q_c$  y la fase localizada con  $P_0 > 0$  para  $q > q_c$ . Hemos encontrado ya que este valor crítico está dado por la formula siguiente:

$$q_c = \frac{(1 - \gamma)P_{no-return}}{\gamma + (1 - \gamma)P_{no-return}} \quad (3.28)$$

donde  $P_{no-return}$  está dada por Ec.(3.20). Por lo discutido en la sección anterior tenemos que para vuelos de Lévy con índice  $0 < \mu \leq 2$  (incluyendo las caminatas aleatorias simples correspondientes a  $\mu = 2$ )  $q_c$  va como sigue:

$$q_c = \frac{(1 - \gamma)P_{no-return}}{\gamma + (1 - \gamma)P_{no-return}} > 0 \quad \text{para } d > \mu \quad (3.29)$$

$$= 0 \quad \text{para } d < \mu. \quad (3.30)$$

Ahora bien, nos interesa encontrar la forma en que  $P_0$  crece desde su valor mínimo (cero) conforme  $q$  crece por encima de  $q_c$ . Queremos probar que, en general, si  $q \rightarrow q_c^+$  entonces:

$$P_0 \sim (q - q_c)^\beta \quad (3.31)$$

donde el exponente  $\beta$  depende continuamente de  $\mu$  y  $d$  en el plano  $(\mu, d)$ . Mostraremos que:

$$\beta = \begin{cases} 1 & \text{para } d > 2\mu \\ \frac{\mu}{d-\mu} & \text{para } \mu < d < 2\mu \\ \frac{d}{\mu-d} & \text{para } d < \mu \end{cases} \quad (3.32)$$

Empecemos por recordar la expresión de auto-consistencia para  $P_0$  dada por Ec.(3.6):

$$\frac{1}{(2\pi)^d} \int_{\mathcal{B}} \frac{d^d \mathbf{k}}{(1 - q)[1 - \tilde{p}(\mathbf{k})] + q\gamma P_0} = \frac{1 - \gamma}{q\gamma(1 - \gamma P_0)}. \quad (3.33)$$

Es claro que para  $q = q_c$ ,  $P_0 = 0$  lo cual no lleva a la siguiente expresión:

$$\frac{1}{(2\pi)^d} \int_{\mathcal{B}} \frac{d^d \mathbf{k}}{(1 - q_c)[1 - \tilde{p}(\mathbf{k})]} = \frac{1 - \gamma}{q_c \gamma}, \quad (3.34)$$

lo cual nos lleva directamente a la expresión que ya conocíamos para  $q_c$  [Ec.(3.28)]. Para encontrar el comportamiento de  $P_0$  al incrementar  $q$  sobre  $q_c$  debemos considerar dos casos:

**Caso I:  $q_c > 0$ .** Como ya hemos visto, este caso corresponde al régimen transiente donde  $P_{no-return} > 0$ . Para vuelos de Lévy esto significa, como ya se ha mencionado, que  $d > d_c = \mu$ . Empecemos por restar (3.33) de (3.34), lo cual nos da:

$$\int_{\mathcal{B}} \frac{d^d \mathbf{k}}{(2\pi)^d} \frac{[q\gamma P_0 - (q - q_c)(1 - \tilde{p}(\mathbf{k}))]}{(1 - \tilde{p}(\mathbf{k}))[(1 - q)(1 - \tilde{p}(\mathbf{k})) + q\gamma P_0]} = \frac{(1 - q_c)(1 - \gamma)(q - q_c - q\gamma P_0)}{qq_c\gamma(1 - \gamma P_0)}. \quad (3.35)$$

Hagamos  $q = q_c + \epsilon$  con  $\epsilon \rightarrow 0$  y  $P_0 = \delta$  con  $\delta \rightarrow 0$ . El objetivo es encontrar cómo  $\delta$  escala con  $\epsilon$  en el régimen de  $\epsilon$  pequeño. En estos límites, la única contribución de la integral del lado izquierdo de (3.35) se da en rangos de  $\mathbf{k}$  pequeños, que es donde podemos reemplazar a  $\tilde{p}(\mathbf{k})$  por lo descrito en la Ec.(3.21). Así, la ecuación anterior puede simplificarse como sigue:

$$\delta I(\delta) + O(\delta) = A\epsilon \quad (3.36)$$

donde  $A = (1 - \gamma)(1 - q_c)K_\mu^2/(\gamma^2 q_c^3)$  es sólo una constante y  $I(\delta)$  es la integral

$$I(\delta) = \int_{\mathcal{B}} \frac{d^d \mathbf{k}}{(2\pi)^d} \frac{1}{|\mathbf{k}|^\mu [|\mathbf{k}|^\mu + b\delta]} \quad (3.37)$$

donde  $b = q_c\gamma/(K_\mu(1 - q_c))$  es también una constante. Ahora, debemos analizar la integral anterior en el límite  $\delta \rightarrow 0$ . Aquí hay de nuevo dos casos a considerar: (1)  $d > 2\mu$  y (2)  $\mu < d < 2\mu$ . Considerémoslos por separado.

1.  $d > 2\mu$ : En este caso, si hacemos  $\delta = 0$  la integral  $I(\delta)$  dada por la Ec.(3.37), la integral converge conforme  $k \rightarrow 0$ , siendo  $I(0)$  finito. Así, de la Ec.(3.36), tenemos que:

$$\delta \sim \epsilon \quad \Rightarrow \quad \beta = 1 \quad \text{para} \quad d > 2\mu. \quad (3.38)$$

2.  $\mu < d < 2\mu$ : En este caso, la integral  $I(0)$  de la Ec.(3.37) diverge. Por lo tanto, al extraer la singularidad y reescalando  $k \rightarrow \delta^{1/\mu}y$  en la Ec.(3.37) tenemos que:

$$I(\delta) \sim \delta^{\frac{d}{\mu}-2} \int_0^\infty \frac{dy y^{d-1-\mu}}{y^\mu + b}. \quad (3.39)$$

Nótese que la integral de la Ec.(3.39) converge en ambos límites  $y \rightarrow 0$  y  $y \rightarrow \infty$  para  $\mu < d < 2\mu$ . Así, sustituyendo la Ec.(3.39) en la Ec.(3.36) obtenemos que:

$$\delta \sim \epsilon^{\frac{\mu}{d-\mu}} \quad \Rightarrow \quad \beta = \frac{\mu}{d-\mu} \quad \text{para} \quad \mu < d < 2\mu. \quad (3.40)$$

**Caso II:  $\mathbf{q}_c = \mathbf{0}$ .** Este caso corresponde al régimen recurrente donde  $P_{no-return} = 0$ , haciendo  $q_c = 0$ . Como discutimos antes, para vuelos de Lévy con índice  $0 < \mu \leq 2$  eso sucede solo cuando  $d < d_c = \mu$ . En este caso analizaremos directamente la Ec.(3.33) sustituyendo  $q = \epsilon$  y  $P_0 = \delta$ . De nuevo, tomando sólo la contribución de la integral a  $\mathbf{k}$  pequeñas, obtenemos el orden principal:

$$\int_{\mathcal{B}} \frac{d^d \mathbf{k}}{(2\pi)^d} \frac{1}{|\mathbf{k}|^\mu + \epsilon \delta} \sim \frac{1}{\epsilon} \quad (3.41)$$

Reescalando  $k = (\epsilon \delta)^{1/\mu} y$  tenemos

$$(\epsilon \delta)^{\frac{d}{\mu}-1} \int_0^\infty \frac{dy y^{d-1}}{y^\mu + 1} \sim \frac{1}{\epsilon}. \quad (3.42)$$

La integral de la Ec.(3.42) converge en ambos límites  $y \rightarrow 0$  y  $y \rightarrow \infty$  para  $0 < d < \mu$ . Por lo tanto, de la Ec.(3.42) obtenemos

$$\delta \sim \epsilon^{\frac{d}{\mu-d}} \Rightarrow \beta = \frac{\mu}{\mu-d} \quad \text{para } 0 < d < \mu. \quad (3.43)$$

Esto completa la prueba de los resultados dados para el exponente  $\beta$  en la Ec.(3.32). Para el caso particular en el que  $p(\ell)$  tiene segundo momento finito, como en el caso de la dinámica de primeros vecinos,  $\beta$  queda como:

$$\beta = \begin{cases} 1 & \text{para } d \geq 4 \\ \frac{2}{d-2} & \text{para } 2 < d < 4 \\ \frac{d}{2-d} & \text{para } d < 2. \end{cases} \quad (3.44)$$

### 3.6.2. Longitud de localización.

Por otro lado la Ec.(3.23) para la distribución de la posición del caminante es similar, por un prefactor, a la función de correlación de un modelo Gaussiano con transición de fase de segundo orden cuyo parámetro de orden viene dado por:

$$C(\mathbf{k}) \propto 1/(|\mathbf{k}|^2 + \xi^{-2}) \quad (3.45)$$

donde la transformada inversa de (3.45) viene dada por [103]:

$$\tilde{C}(\mathbf{n}) \propto \frac{\exp(-|\mathbf{n}|/\xi)}{|\mathbf{n}|^{(d-2)/2}} \quad (3.46)$$

para  $|\mathbf{n}|$  grandes.

Así,

$$\xi = (D/q^*)^{1/2}, \quad q^* = q\gamma P_0 \quad (3.47)$$

es la longitud de localización. De (3.23) y (3.31) podemos deducir que  $\xi$  diverge siempre como

$$\xi \sim (q - q_c)^{-\nu} \quad (3.48)$$

cerca de  $q_c$ , con  $\nu = \beta/2$ . Por lo tanto, considerando la Ec.(3.44):

$$\nu = \begin{cases} 1/2 & \text{para } d \geq 4 \\ \frac{1}{d-2} & \text{para } 2 < d < 4 \\ \frac{d}{4-2d} & \text{para } d < 2. \end{cases} \quad (3.49)$$

Estos exponentes y dimensiones críticas son idénticas a las encontradas en la teoría de auto-consistencia (SCT) de la localización de Anderson [31, 105, 106]. En este problema el coeficiente de difusión que describe el transporte de ondas en un medio desordenado obedece una relación de auto-consistencia similar a la mostrada en la Ec.(3.6). Por lo tanto, nuestro modelo (o al menos los resultados derivados del supuesto de decorrelación) pertenecen a la misma clase de universalidad que el de la SCT de la localización de Anderson, la cual es también una aproximación al problema de la propagación de ondas.

### 3.7. Modelo mínimo en $2d$ y $3d$ .

Enfocándonos nuevamente en las caminatas aleatorias con varianza finita, presentaremos a continuación algunos resultados del comportamiento crítico de la transición de la localización en la dimensión crítica inferior  $d_c = 2$  y en  $d = 3$  que es la dimensión física donde la transición de fase puede observarse a una  $q_c$  finita si se considera una dinámica de primeros vecinos. Veamos:

**Caso  $2d$ :** Para el caso de caminatas aleatorias con segundo momento finito la dimensión crítica inferior es  $d_c = 2$ , lo cual es una consecuencia directa de la transición recurrencia/transitoriedad del proceso. El caso  $2d$  nos merece una atención especial, además de su relevancia ecológica resulta analíticamente interesante estudiar el comportamiento de la dinámica justo en su dimensión crítica. En  $d = 2$ , siguiendo la Ec.(3.49), el exponente de la longitud de correlación  $\nu$  se vuelve infinito, lo cual indica que la divergencia de  $\xi$  cerca de  $q_c = 0$  no sigue una ley de potencia. Para obtener el comportamiento preciso y correcto, uno debe resolver la ecuación de auto-consistencia (3.6) para  $P_0$  cuando  $q$  es muy cercano a cero. Para el caso de la caminata a primeros vecinos en una rejilla cuadrada,  $\tilde{p}(\mathbf{k}) \simeq 1 - |\mathbf{k}^2|/4$  para  $\mathbf{k}$  pequeños. Cambiando a coordenadas polares y tomando en cuenta que  $\rho \equiv |\mathbf{k}|$ , la Ec.(3.6) se ve como:

$$\int_0^R \frac{d\rho}{2\pi} \frac{\rho}{\frac{1-q}{4}\rho^2 + q\gamma P_0} \simeq \frac{1-\gamma}{q\gamma(1-\gamma P_0)} \quad (3.50)$$

para  $(q, P_0)$  pequeño y  $R$  una constante de orden 1. Integrando (3.50) obtenemos:

$$\ln(q\gamma P_0) \simeq -\frac{\pi(1-\gamma)(1-q)}{q\gamma(1-\gamma P_0)} \simeq -\frac{\pi(1-\gamma)}{q\gamma}. \quad (3.51)$$

Por lo tanto, como  $q \rightarrow 0$ ,  $P_0$  tiende a 0 con una singularidad esencial:

$$P_0(q) \sim \frac{1}{q\gamma} \exp \left[ -\frac{\pi(1-\gamma)}{\gamma q} \right]. \quad (3.52)$$

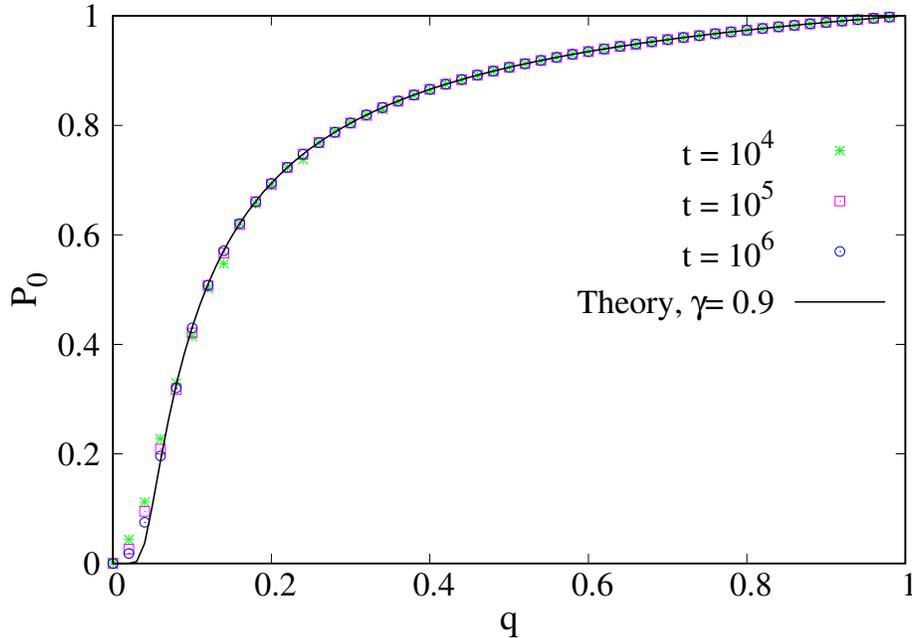


Figura 3.8:  $P_0$  como función de  $q$  para  $\gamma = 0.9$  y pasos aleatorios en una rejilla cuadrada  $2d$ . Los símbolos son el resultado de las simulaciones para las caminatas que inician en el origen y duran  $t$  pasos de tiempo, la línea sólida negra corresponde a la solución numérica de Eq. (3.6).

La longitud de correlación se deduce de (3.47), con una constante de difusión dada por  $D = (1 - q)/4$ . Se deduce pues que

$$\xi \sim \exp \left[ \frac{\pi(1 - \gamma)}{2\gamma q} \right]. \quad (3.53)$$

Así,  $\xi$  diverge más rápido que cualquier ley de potencia en el límite  $q \rightarrow 0$ . La teoría de escalamiento de la localización de Anderson en dos dimensiones predice una ley similar para la longitud de correlación como función del parámetro adimensional de conductividad característica [31].

En Fig. 3.8 comparamos la aproximación numérica de la ecuación de auto-consistencia (3.6) con las simulaciones de Monte Carlo, podemos notar que se obtiene un buen ajuste de la curva simulada (símbolos) a la curva teórica (línea negra). Esto justifica, nuevamente y en  $2d$ , el supuesto de decorrelación que se hizo desde un inicio para poder pasar de la Ec.(3.1) a la Ec.(3.3). Lo que parece ser una transición abrupta para un valor pequeño de  $q_c$  en la curva teórica, en realidad corresponde a un comportamiento suave donde  $P_0$  es muy pequeño pero no es cero. Cerca de  $q = 0$ , la convergencia de las simulaciones hacia el estado estacionario es muy lenta: esto debido al hecho de que  $\xi$  toma valores muy grandes en este límite, por lo tanto la dinámica lleva a que el caminante recorra distancias muy grandes antes de alcanzar el NESS. Este proceso toma mucho tiempo por la difusión logarítmica de la dinámica que

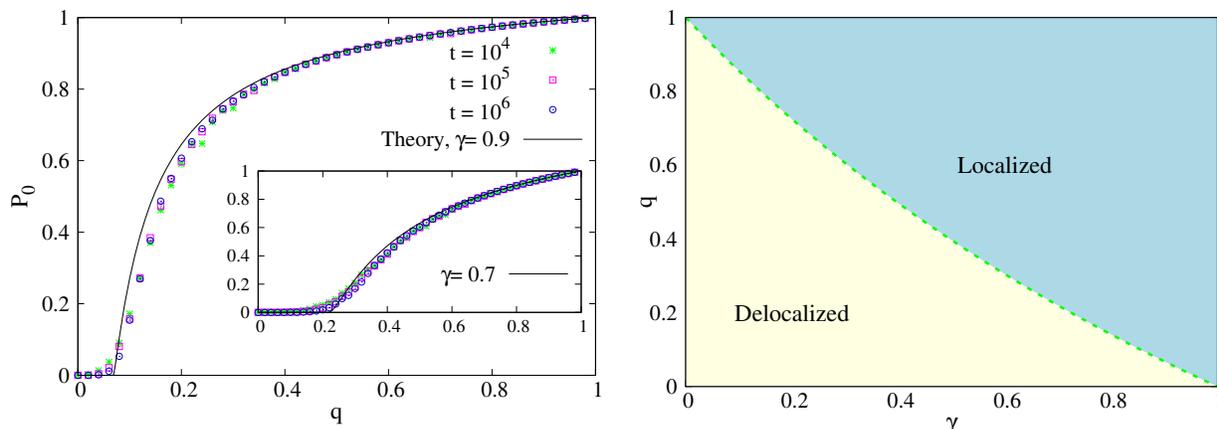


Figura 3.9: **Izquierda:** Lo mismo que en la Fig. 3.8 pero para el caso 3D en una rejilla cúbica. Inset:  $\gamma = 0.7$ . **Derecha:** Diagrama fase para el modelo mínimo en 3D con pasos aleatorios a primeros vecinos.

caracteriza a este modelo mínimo en un espacio homogéneo.

**Caso 3d:** En esta pequeña subsección presentaremos de manera breve algunos resultados numéricos para una caminata aleatoria a primeros vecinos en una rejilla cúbica que sigue las reglas dinámicas de la Ec.(3.6).

$d = 3$  es la dimensión física donde una transición de fase a un valor finito de  $q_c$  puede observarse en nuestro modelo con una dinámica de saltos aleatorios a primeros vecinos. en una rejilla cúbica, la Ec. (3.6) y las simulaciones de Monte Carlo exhiben una vez más un buen ajuste para todos los valores de  $q$ , como se muestra en la Fig. 3.9-Izquierda De la relación para  $q_c$  dada en (3.28) y el valor de la probabilidad de no retorno para una CAS en 3d  $P_{no\ return} = 0.6595\dots$  [99], se obtiene que  $q_c = 0.0682\dots$  para  $\gamma = 0.9$  y  $q_c = 0.2203\dots$  para  $\gamma = 0.7$  (inset). Existe pues una clara evidencia numérica de una transición de fase localización/delocalización para el caso 3d. Dar una estimación numérica precisa de  $q_c$  requeriría un análisis de escala de tiempo finito que va más allá del alcance y el objetivo de este trabajo. Sin embargo, sería interesante cuantificar las desviaciones del modelo de las predicciones de nuestra teoría auto-consistente, si las hubiera, y determinar si el modelo pertenece realmente a esta clase o a otra clase de universalidad. La Fig. 3.9-Derecha muestra el diagrama de fase del modelo mínimo en 3d, derivado de la Ec.(3.28), éste tiene un comportamiento parecido al ya mostrado para el caso de vuelos de Lévy en 1d donde el valor de  $q_c$  disminuye conforme aumenta el peso de la heterogeneidad ( $\gamma$ ) sabiendo ya que debajo de los valores críticos de  $q_c$  la dinámica se encuentra en una fase de delocalización y arriba de éstos se alcanza una fase de localización alrededor de la inhomogeneidad.

En la siguiente sección, y la última de este capítulo, expondremos una modificación del modelo mínimo presentado hasta ahora que presenta, también, una transición de fase delocalización/localización y tiene propiedades analíticas interesantes.

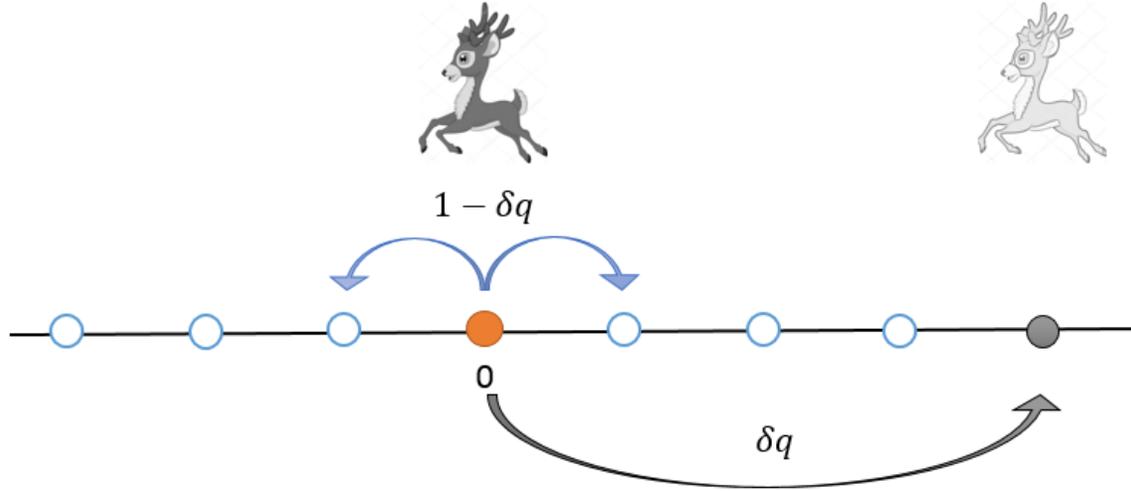


Figura 3.10: Reglas de la modificación al modelo mínimo, cuando  $\mathbf{X}_t = \mathbf{0}$ . Si  $\mathbf{X}_t \neq \mathbf{0}$ , el movimiento será como el ya expuesto en la Fig. 3.1-Arriba.

### 3.8. Una simple modificación al modelo mínimo.

El segundo modelo que presentaremos a continuación está también inspirado en el fenómeno ecológico del forrajeo animal pero ahora consideraremos una respuesta diferente del individuo al encontrar un sitio de recursos. Supongamos que el caminante se desplaza en una rejilla  $d$ -dimensional en la que se encuentra un único sitio de interés (s.p.g. supondremos que éste está en  $\mathbf{n} = \mathbf{0}$ ), este sitio tiene una probabilidad de estadía cero (como en el resto de los sitios) pero favorece el movimiento aleatorio, de tal suerte que cuando el animal se encuentra en él es más probable que elija dar un paso aleatorio que reubicarse a un sitio previamente visitado. En otras palabras, una vez que el animal se da cuenta de que en ese sitio hay alimento, cambia su comportamiento hacia un aumento de la exploración local a expensas de usar su memoria. Esto se modela disminuyendo la tasa de reubicación en  $\mathbf{0}$  por un factor  $\delta < 1$  (Fig.3.10).

De manera formal: si definimos  $r_{\mathbf{n}} = q[1 - (1 - \delta)\delta_{\mathbf{n},0}]^1$  como nuestra nueva tasa de reubicación, la ecuación de movimiento que define la nueva dinámica ya expuesta arriba, viene dada por:

$$\begin{aligned}
 P_{\mathbf{n}}(t+1) &= \sum_{\ell} (1 - r_{\mathbf{n}-\ell}) P_{\mathbf{n}-\ell}(t) p(\ell) \\
 &+ \frac{q}{t+1} \sum_{t'=0}^t \text{Prob}[\mathbf{X}_{t'} = \mathbf{n} \text{ and } \mathbf{X}_t \neq \mathbf{0}] \\
 &+ \frac{\delta q}{t+1} \sum_{t'=0}^t \text{Prob}[\mathbf{X}_{t'} = \mathbf{n} \text{ and } \mathbf{X}_t = \mathbf{0}]
 \end{aligned} \tag{3.54}$$

<sup>1</sup>Donde  $\delta$  no debe confundirse con el símbolo para la delta de Kroneker  $\delta_{\mathbf{n},0}$

en cualquier dimensión  $d$ . Recordando que  $P_{\mathbf{n}}(t)$  es la probabilidad de ocupar el sitio  $\mathbf{n}$  de la rejilla al tiempo  $t$  y  $p(\ell)$  la probabilidad de realizar un salto aleatorio de longitud  $\ell$ . La posición que depende de la probabilidad de reubicación está dada por  $r_{\mathbf{n}} = q$  para  $\mathbf{n} \neq \mathbf{0}$  y  $r_{\mathbf{n}} = \delta q$  para  $\mathbf{n} = \mathbf{0}$ . Así como en el modelo anterior, utilizamos la aproximación de decorrelación para obtener los resultados analíticos de interés. Aproximamos pues,  $\text{Prob}[\mathbf{X}_{t'} = \mathbf{n} \text{ and } \mathbf{X}_t \neq \mathbf{0}]$  por  $P_{\mathbf{n}}(t')[1 - P_{\mathbf{0}}(t)]$ , y  $\text{Prob}[\mathbf{X}_{t'} = \mathbf{n} \text{ and } \mathbf{X}_t = \mathbf{0}]$  por  $P_{\mathbf{n}}(t')P_{\mathbf{0}}(t)$ . La validez de este supuesto será puesta a prueba con simulaciones de Monte Carlo. Tomando de manera conjunta los límites  $t \rightarrow \infty$  y  $t' \rightarrow \infty$ , y tomando en cuenta que  $\lim_{t \rightarrow \infty} P_{\mathbf{n}}(t) = P_{\mathbf{n}}$ , obtenemos una ecuación que describe el NESS:

$$P_{\mathbf{n}} = (1 - q) \sum_{\ell} P_{\mathbf{n}-\ell} p(\ell) + P_{\mathbf{0}} p(\mathbf{n}) q (1 - \delta) + q P_{\mathbf{n}} [1 - P_{\mathbf{0}} (1 - \delta)]. \quad (3.55)$$

Haciendo  $P_{\mathbf{0}} \equiv P_0$  y aplicando la transformada discreta de Fourier a esta ecuación, obtenemos:

$$\tilde{P}(\mathbf{k}) = \frac{P_0 q (1 - \delta) \tilde{p}(\mathbf{k})}{(1 - q) [1 - \tilde{p}(\mathbf{k})] + q (1 - \delta) P_0}. \quad (3.56)$$

Además de la solución trivial  $P_0 = 0$ , se pueden obtener otras soluciones sustituyendo (3.56) en la relación de auto-correlación general (3.5). Aquellas soluciones en las que  $P_0 > 0$  siguen la ecuación

$$\frac{1}{(2\pi)^d} \int_{\mathcal{B}} \frac{d\mathbf{k}}{(1 - q) [1 - \tilde{p}(\mathbf{k})] + q (1 - \delta) P_0} = \frac{1 - \delta q}{q (1 - \delta) [1 - q + q (1 - \delta) P_0]}. \quad (3.57)$$

Una vez más, fijando  $\delta$ , el punto crítico  $q_c$  que caracteriza el comienzo del comportamiento localizado se obtiene al fijar  $P_0 = 0$  en la ecuación de arriba. Reordenando los términos tenemos que:

$$q_c = \frac{P_{no\text{-}return}}{(1 - \delta) + \delta P_{no\text{-}return}}. \quad (3.58)$$

Así como en el modelo mínimo expuesto en las secciones anteriores, la existencia de la transición de fase a un valor finito de  $q_c > 0$  es posible cuando el proceso Markoviano subyacente (para  $q = 0$ ) es transiente, esto es  $P_{no\text{-}return} \neq 0$ , mientras  $q_c = 0$  para los procesos recurrentes, esto es  $P_{no\text{-}return} = 0$ . Las ecuaciones (3.56) y (3.57) tienen la misma estructura que las ecuaciones (3.4) y (3.6) del modelo anterior, por lo tanto la transición de localización pertenece a la misma clase. En particular, el comportamiento a gran escala del NESS obedece (3.23) con  $q^* = q(1 - \delta)P_0$  y la divergencia de la longitud de la localización está dada por las leyes de escala descritas en (3.49).

### 3.8.1. Expresión analítica en $1d$

Así como en el modelo mínimo anterior [1], es posible deducir analíticamente una solución aproximada para la caminata aleatorio a primeros vecinos en  $1d$  que sigue la dinámica ya expuesta en esta sección. Consideraremos la distribución de pasos a primeros vecinos dada por:

$$p(\ell) = \frac{1}{2} [\delta_{\ell,1} + \delta_{\ell,-1}], \quad (3.59)$$

(el símbolo de Kroneker  $\delta_{n,0}$  no debe confundirse con el parámetro  $\delta$ ) cuya transformada de Fourier es, como ya se ha mencionado antes,  $\tilde{p}(k) = \cos(k)$ . El estado estacionario (3.56) queda pues como sigue:

$$\tilde{P}(k) = \frac{P_0 q (1 - \delta) \cos k}{(1 - q)(1 - \cos k) + q(1 - \delta)P_0} = A + \frac{B}{(1 - q)(1 - \cos k) + q(1 - \delta)P_0} \quad (3.60)$$

con  $A = [q(1 - \delta)P_0]/(q - 1)$  y  $B = A [q - 1 - q(1 - \delta)P_0]$ . Siguiendo el mismo desarrollo ya expuesto para el modelo mínimo original obtenemos la siguiente ecuación para la probabilidad de que el caminante ocupe el sitio  $n$ .

$$P_n = \frac{q(1 - \delta)P_0}{q - 1} \delta_{n,0} + \frac{q(1 - \delta)P_0 [1 - q + q(1 - \delta)P_0]}{(1 - q)^2} \frac{2a}{(a^2 - 1)a^{|n|}}. \quad (3.61)$$

con  $a$  definida por:

$$a = 1 + u + \sqrt{u(2 + u)} \quad (3.62)$$

y  $u = [q(1 - \delta)P_0]/(1 - q)$ . Fijando  $n = 0$  en la Ec. (3.61), obtenemos una ecuación de auto-consistencia para la probabilidad, aún no conocida,  $P_0$  de ocupar el origen,

$$2q(1 - \delta)[q - 1 - q(1 - \delta)P_0] = (\delta q - 1)(1 - q)(a - a^{-1}). \quad (3.63)$$

Sustituyendo el valor para  $a$  dado en (3.62) en la igualdad de arriba, obtenemos una ecuación cuadrática para  $P_0$ :

$$a_1 P_0^2 + b_1 P_0 + c_1 = 0 \quad (3.64)$$

con,

$$a_1 = q^2(1 - \delta)^2(1 - q)[1 + q(1 - 2\delta)] \quad (3.65)$$

$$b_1 = 2q(1 - \delta)(1 - q)^2[1 + q(1 - 2\delta)] \quad (3.66)$$

$$c_1 = -q^2(1 - \delta)^2(1 - q)^2. \quad (3.67)$$

La Ec. (3.64) tiene una sola raíz positiva dada por

$$P_0 = \frac{1}{q(1 - \delta)} \left[ \sqrt{\frac{(1 - q)(1 - \delta q)^2}{1 + q(1 - 2\delta)}} + q - 1 \right]. \quad (3.68)$$

Es sencillo probar que  $0 \leq P_0 \leq 1$  para cualquier  $(q, \delta) \in [0, 1]^2$ . Fijando  $\delta$ , y expandiendo (3.68) para  $q$  pequeñas obtenemos que:

$$P_0 \simeq \frac{1 - \delta}{2} q \rightarrow 0, \quad (3.69)$$

indicándonos que el NESS desaparece para  $q = 0$ . En este límite, se reduce a una CAS con saltos aleatorios a primeros vecinos, la cual, como bien se sabe, ocupa el origen con probabilidad cero para  $t = \infty$ . De manera contraria, cuando  $q \rightarrow 1$  se deduce de (3.68) que

$$P_0 \simeq \sqrt{\frac{1 - q}{2(1 - \delta)}} \quad (3.70)$$

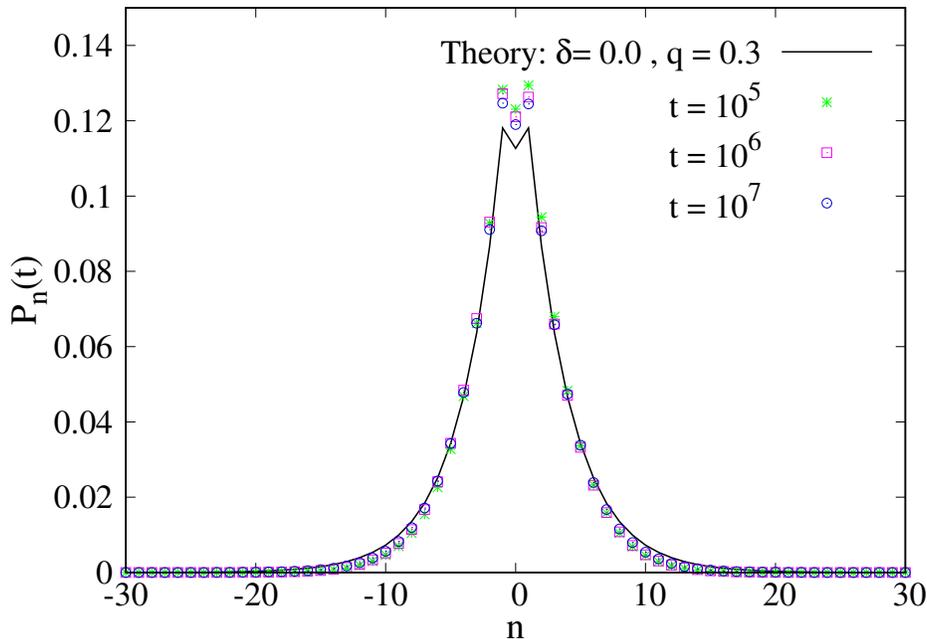


Figura 3.11: Probabilidad de ocupación del sitio  $n$   $P_n$  como función de la posición  $n$  en  $1D$  para el modelo- $\delta$  definido por (3.61). Los símbolos son los resultados de las simulaciones de Monte Carlo para caminatas que inician en el origen y realizan  $t$  saltos.

para cualquier  $\delta < 1$  fijo. Por lo tanto  $P_0 \rightarrow 0$  en el límite de un uso muy frecuente de memoria, lo cual parece sorprendente. El caminante que inicia en el origen, tarde o temprano salta a alguno de sus sitios vecinos más cercanos ( $n = 1$  o  $-1$ ) el cual se refuerza luego de varios pasos, dejando el origen desocupado. Evaluando la Ec. (3.61) en  $n = \pm 1$  y usando (3.70), se obtiene que

$$P_1 = P_{-1} = \frac{1}{2}, \quad (3.71)$$

en el límite  $q = 1$ , mientras que  $P_n = 0$  para  $|n| > 1$  debido a la divergencia de la constante  $a$ . Por lo tanto, el individuo se localiza por completo en  $n = 1$  y  $n = -1$  cuando  $q \rightarrow 1$ , y no en el origen como en el modelo mínimo original. Si el sitio de partida no es el origen, el individuo permanecerá en ese sitio por siempre, como una consecuencia trivial de la dinámica. La teoría de auto-consistencia presentada hasta ahora no es válida para aquellas trayectorias que no tienen la oportunidad de visitar el origen. La Fig. 3.11 muestra el NESS obtenido para  $q = 0.3$  y  $\delta = 0$ , y donde es ya más que claro que, asintóticamente,  $P_{n=0} < P_{n=1}$ . Se obtiene un ajuste excelente sobre todo el rango de posiciones  $n$  entre la teoría y las simulaciones de Monte Carlo incluso para tiempos no tan largos de la caminata.

Se deduce de todo lo ya hecho arriba que, contrario a lo que pasa para el modelo original,  $P_0$  no crece de manera monótona con  $q$  y que debe existir un máximo para un valor específico  $q_\delta^*$  de la probabilidad de reubicación. La Fig. 3.12-Izquierda muestra la solución (3.68) como función de  $q$  para varios valores de  $\delta$ . El valor en donde  $P_0$  alcanza su máximo,  $q_\delta^*$ , crece  $\delta$ : además, la localización se debilita conforme  $\delta$  crece.

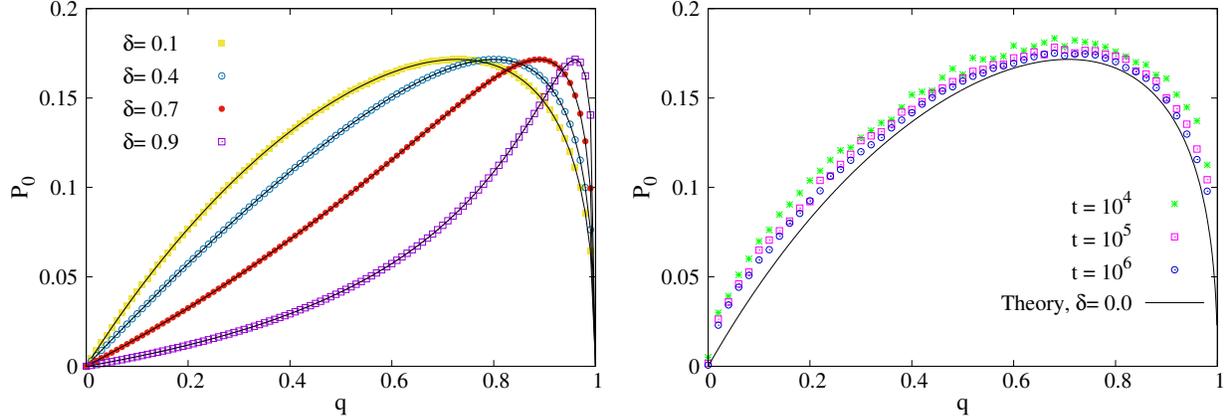


Figura 3.12: **Izquierda:**  $P_0$  como función de  $q$  en  $1d$  para varios valores de  $\delta$  tal como se define en la Ec. (3.68). **Derecha:** Misma cantidad pero para  $\delta = 0$  (línea sólida); los símbolos representan los resultados de las simulaciones.

Es de particular interés el caso  $\delta = 0$ : en esta situación el caminante suprime por completo su uso de memoria cuando está en el origen (lo cual no significa que no recuerde esta posición para futuras reubicaciones) y realiza, siempre, saltos aleatorios hacia sus vecinos más cercanos  $n = 1$  o  $n = -1$ . La Ec. (3.68) con  $\delta = 0$  se reduce a

$$P_0 = \frac{1}{q} \left[ \sqrt{\frac{1-q}{1+q}} + q - 1 \right], \quad (3.72)$$

la cual alcanza su máximo en  $q = q_0^* = 1/\sqrt{2} = 0.707106\dots$ . En este valor  $q_0^*$ ,  $P_0 = (\sqrt{2}-1)^2 = 0.171572\dots$ , y es este también el valor máximo que  $P_0$  puede alcanzar para cualquier  $\delta \neq 0$  (ver Fig. 3.12-Izquierda). Esta probabilidad es significativamente menor que los valores cercanos a la unidad que se obtenían para el modelo mínimo original. Se obtiene un ajuste cuantitativo aceptable entre las simulaciones de Monte Carlo y la Ec.(3.72), ver la Fig. 3.12-Derecha, sugiriendo que la aproximación de decorrelación que se utilizó para obtener todos estos resultados analíticos debe ser correcta ya que conforme el tiempo de simulación  $t$  aumenta, los resultados numéricos se aproximan lentamente a la curva asintótica esperada. Esta lenta convergencia se atribuye a la dinámica logarítmica que domina el modelo de visita preferencial, como ya se ha discutido con anterioridad.

El caso en el que  $q$  se fija y  $\delta$  es cercano a la unidad resulta igual de interés. Una expansión de la Ec. (3.68) con  $\delta = 1 - \epsilon$  y  $\epsilon \ll 1$  no da que  $P_0 \propto \epsilon$ . Así, cuando la diferencia entre las tasas de reubicación entre el origen y otros sitios tiende a cero, el NESS desaparece. Esto va de acuerdo con el comportamiento encontrado en el modelo inicial en un espacio homogéneo expuesto en [85], donde el desplazamiento cuadrático medio crece como  $\ln t$  para cualquier valor de  $q$ , implicando un desvanecimiento de la densidad cuando  $t = \infty$ .

### 3.8.2. Análisis para el caso $2d$ y $3d$

Seguiremos ahora con un estudio de este modelo mínimo modificado con dinámica a primeros vecinos en dimensiones grandes. Comencemos con  $d = 2$  que, como ya sabemos, es la dimensión crítica para la transición de un proceso recurrente/transiente. En esta dimensión crítica,  $q_c$  sigue siendo 0 y la longitud de correlación puede obtenerse siguiendo la misma ruta ya marcada por la Ec. (3.53) para el modelo mínimo original. Empezando de la Ec. (3.57) podemos obtener, luego de una simple álgebra,

$$P_0 \sim \frac{1}{q(1-\delta)} \exp \left[ -\frac{\pi}{q(1-\delta)} \right], \quad (3.73)$$

para  $q \ll 1$ . El comportamiento de la Ec. (3.56) para valores pequeños de  $k$  nos permite identificar la longitud de correlación como  $\xi = \{D/[q(1-\delta)P_0]\}^{1/2}$ . Por lo tanto, a tasas de reubicación pequeñas,

$$\xi \sim \exp \left[ \frac{\pi}{2(1-\delta)q} \right]. \quad (3.74)$$

La Fig. 3.13-Izq muestra  $P_0$  vs.  $q$  para  $\delta = 0$ . Las simulaciones numéricas exhiben el mismo comportamiento cualitativo que el previsto por la teoría, pero los efectos de tiempo finito son muy fuertes y una comparación cuantitativa se vuelve prácticamente imposible.

Podemos notar, sin embargo, que la tasa de reubicación característica  $q_{char}$ , la cual describe una singularidad esencial, es significativamente mayor en la Fig. 3.13-Izquierda que en la Fig. 3.8 para el modelo mínimo original. Este comportamiento puede ser entendido desde la Ec. (3.74), la cual predice que  $q_{char} = \pi/[2(1-\delta)] = \pi/2$  (para  $\delta = 0$ ) y comparándola con la tasa de reubicación característica para el modelo original la cual está definida por  $q_{char} = \pi(1-\gamma)/(2\gamma) = \pi/18$  (para  $\gamma = 0.9$ ).

Para el caso de las caminatas aleatorias con dinámica a primeros vecinos que siguen las reglas de movimiento de este modelo mínimo modificado en una rejilla cúbica en  $3D$ ,  $q_c = 0.6595\dots$  de la Ec. (3.58) con  $\delta = 0$ . El comportamiento analítico del valor  $P_0(q)$  se muestra en la Fig. 3.13-Derecha, junto con los resultados de los experimentos numéricos. Una vez más el ajuste es sólo cualitativo, presumiblemente por la convergencia extremadamente lenta hacia el NESS. Notemos también que las probabilidades de ocupación son pequeñas ( $< 0.06$  para todo  $q$ ), lo que hace su estimación numérica aún más difícil. La Fig. 3.13-Abajo muestra el diagrama de fase para el modelo mínimo modificado en  $3D$  predicho por la Ec. (3.58). Contrario a lo que pasa para el diagrama de fase del modelo mínimo original en  $3d$  (ver Fig. 3.9), en este modelo mínimo modificado, el valor crítico de la tasa de reubicación aumenta junto con el valor de  $\delta$ .

Hasta aquí hemos mostrado un par de modelos mínimos que presentan una transición de fase entre los estados delocalización/localización que depende de la tasa de uso de memoria un individuo que explora un ambiente heterogéneo simple. Ambos modelos muestran la emergencia de un aprendizaje espacial del individuo al interactuar con su ambiente.

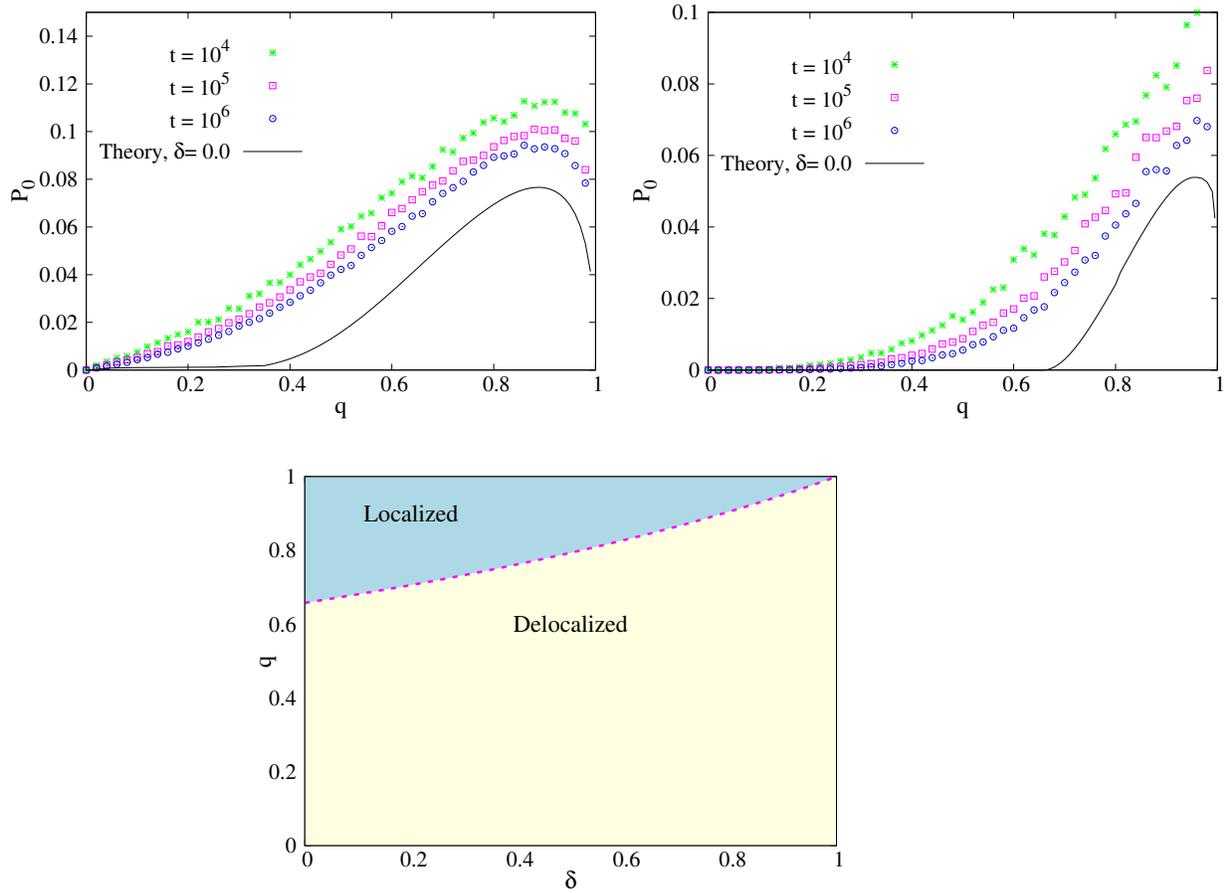


Figura 3.13: **Izquierda:**  $P_0$  como función de  $q$  para  $\delta = 0$  para el modelo mínimo modificado con saltos a primeros vecinos en una rejilla cuadrada en  $2D$ . **Derecha:** caso en  $3D$ . Los símbolos representan las simulaciones de Monte Carlo y las líneas sólidas son la solución numérica de la ecuación de auto-consistencia (3.57). **Abajo:** diagrama de fase para el caso  $3D$ .

En el siguiente capítulo presentamos los resultados obtenidos al hacer ciertas modificaciones al modelo simple original, arriba expuesto, con el fin de emular el forrajeo y aprendizaje colectivos considerando un *enjambre* de caminantes y una red de transferencia de información. Dicho capítulo se basa en [2].



## Capítulo 4

# Enjambre de caminantes y redes de transferencia de información

Los grupos animales en movimiento, donde los individuos interactúan de forma local, son propensos a exhibir comportamientos colectivos [43, 69, 107, 108] a través de los cuales el grupo puede reunir información de manera precisa sobre el estado actual de su ambiente [47, 109–111]. Esta “inteligencia colectiva” puede permitir una respuesta rápida, tanto individual como grupal, a señales externas. Los efectos de las interacciones en el movimiento colectivo [107, 112] y su impacto en el éxito de forrajeo [93] han sido bien estudiadas en sistemas de agentes sin memoria, hemos presentado en la sección 2.2 algunos de estos modelos. Cuando las reglas dinámicas son locales en el tiempo sólo dependen del estado actual del sistema y no de su historia previa. En contraste, los mecanismos bajo los cuales el comportamiento colectivo emerge de interacciones de individuos con memoria son menos entendidos. En este capítulo nos gustaría cuantificar, por medio de un modelo basado en agentes, cómo el uso de memoria y las interacciones entre miembros del mismo grupo puede contribuir a mejorar el éxito de forrajeo del enjambre completo.

Definimos aquí el aprendizaje colectivo como un cambio coordinado de comportamiento en el grupo, como consecuencia de las experiencias individuales y la transferencia de información entre miembros del grupo [113]. Durante este aprendizaje colectivo los individuos con experiencias diferentes pueden adquirir información valiosa a través de las interacciones con sus pares, lo cual posiblemente puede resultar en un éxito de forrajeo mayor al que se obtendría si se consideraran sólo individuos aislados. La estructura de la red social bajo la cual se transfiere información puede ser relevante en este proceso de aprendizaje colectivo, ya que ciertos individuos son más importantes que otros para transmitir información de interés sobre sitios de recursos [67].

Desarrollaremos un modelo simple basado en agentes que muestra que el aprendizaje colectivo puede contribuir de manera significativa al éxito del forrajeo grupal en ambientes complejos, entendiendo por ambientes complejos aquellos que están compuestos de muchos sitios de recursos cada uno de ellos con valor propio y distinto a los demás. Cuando los agentes pueden usar su memoria para visitar lugares ubicados más allá de su rango de percepción, buscamos determinar si comportamientos colectivos como la agregación espacial

pueden emerger de una población dispersa en el ambiente y que no tiene información inicial sobre los recursos, y si dicha agregación se produce alrededor de los sitios más importantes (con mayores recursos). Nos interesa también estudiar los efectos que distintas topologías en las redes de comunicación pueden tener en el aprendizaje colectivo.

## 4.1. Modelo

El modelo colectivo presentado aquí es una extensión de aquel que corresponde a un sólo caminante con aprendizaje por reforzamiento, expuesto de diferentes formas en [1, 21, 85, 86, 114] y en el capítulo 3 de este mismo escrito. Recordando, en el modelo mínimo del capítulo anterior asumimos que el movimiento del individuo está gobernado por la combinación de dos tipos básicos de desplazamiento: uno que responde a una caminata aleatoria simple y otro que se refiere a regresos preferenciales a sitios ya visitados en el pasado. En este último, el caminante (suponiendo que no existe interacción alguna con sus pares) escoge un sitio particular para reubicarse con una probabilidad proporcional a la cantidad de tiempo total ya gastado en dicho sitio. Por lo tanto, los sitios que han sido visitados con más frecuencia tienen una mayor probabilidad de volver a ser visitados en el futuro, es decir, son linealmente reforzados [1]. Ya que esta dinámica preferencial depende de la historia, esto implica que el caminante tiene la capacidad de recordar los sitios que ha visitado y la cantidad de veces que lo ha hecho. Para agregar la interacción entre varios individuos, vamos a asumir que la memoria se comparte a través de una constante comunicación. Así, en el modo de memoria, un caminante puede reubicarse a un sitio ya visitado por él mismo, o bien, a un sitio previamente ocupado por algún otro miembro del grupo, es decir, utiliza su propia experiencia o la experiencia de alguno de sus compañeros. Otro aspecto importante a considerar es la configuración del ambiente: cada sitio de recurso será caracterizado por un peso o probabilidad  $\gamma_i$ , con lo que un caminante localizado en el sitio  $i$  se quedará en él al siguiente paso de tiempo con probabilidad  $\gamma_i$ , o se moverá con probabilidad  $1 - \gamma_i$ . Esta característica es similar a la que ya habíamos presentado en el capítulo anterior para un ambiente con un solo sitio de recurso, con la diferencia de que ahora el ambiente estará formado por múltiples sitios de interés con diversos pesos  $\gamma_i$ . Veamos esto más a detalle:

### 4.1.1. Ambiente

Consideramos una rejilla discreta, de  $L$  sitios en una dimensión ( $1d$ ) y de  $L \times L$  sitios en dos dimensiones ( $2d$ ), con espacios unitarios y fronteras reflejantes. La coordenada del sitio  $n$  será un entero ( $1d$ ) o un par de enteros en  $2d$ . En la rejilla los recursos están distribuidos aleatoriamente con densidad  $\delta < 1$ , así, en promedio, hay  $M = \delta L$  sitios de interés en  $1d$ . Para cada sitio de recurso  $i$  es asignado un peso fijo  $\gamma_i$  que es un número aleatorio uniformemente distribuido en el intervalo  $(0, \gamma_{max})$ , con  $\gamma_{max} < 1$  un parámetro dado. Sea  $\mathbb{T} = \{(T_i, \gamma_i)\}_{i=1}^M$  el conjunto que contiene las posiciones  $T_i$  y los pesos  $\gamma_i$  de todos los sitios de interés.

### 4.1.2. Caminantes

Consideramos  $N$  caminantes con posiciones iniciales aleatorias en la rejilla y conectados entre sí por una red completa de transferencia de información, es decir, cada caminante puede comunicarse con cualquier otro miembro del grupo (abajo explicaremos los casos en los que se consideran otros tipos de redes de comunicación). El tiempo es una variable discreta  $t$ . Durante el paso  $t \rightarrow t + 1$ , cada individuo  $l = 1, \dots, N$  actualiza su posición  $\mathbf{X}_t^{(l)}$  como sigue:

- (i) *Memoria propia*: Si no está en un sitio de interés, con probabilidad  $q(1 - \rho)$  el caminante regresa a un sitio que ya haya visitado en el pasado, esto es:  $\mathbf{X}_{t+1}^{(l)} = \mathbf{X}_{t'}^{(l)}$ , donde  $t'$  es un entero aleatorio tomado uniformemente del intervalo  $[0, t]$ . Esto implica que la probabilidad de escoger un sitio particular (recurso o no) es proporcional al número de pasos de tiempo previamente gastados en ese sitio. Este es el esquema de reubicación preferencial lineal estudiado ya en el capítulo 3 de este mismo escrito.
- (ii) *Transferencia de información*: Si el caminante no está en un sitio de recurso, con probabilidad  $q\rho$  el individuo escoge al azar otro caminante ( $m$ ) y se reubica en un sitio ya visitado por ese individuo siguiendo la misma regla preferencial:  $\mathbf{X}_{t+1}^{(l)} = \mathbf{X}_{t'}^{(m)}$ , con  $t'$  un entero aleatorio entre  $[0, t]$ .
- (iii) *Movimiento aleatorio*: Si no está en un sitio de interés, con probabilidad  $1 - q$  el caminante se mueve a alguno de sus sitios vecinos más cercanos con igual probabilidad cada uno:  $\mathbf{X}_{t+1}^{(l)} = \mathbf{X}_t^{(l)} + \ell_t$  donde  $\ell_t = \pm 1$  en  $1d$ .
- (iv) *Explotación*: Si está en un sitio de interés (existe  $i$  tal que  $\mathbf{X}_t^{(l)} = T_i$ ), el caminante  $l$  se queda en ese sitio con probabilidad  $\gamma_i$ , y con probabilidad complementaria  $1 - \gamma_i$  usa alguna de las reglas de movimiento (i)-(iii) con sus respectivas probabilidades.

Así, en un paso de tiempo, un individuo que no se queda en alguno de los sitios de interés usa la memoria con probabilidad  $q$  y realiza pasos aleatorios con probabilidad  $1 - q$ . Si elige usar la memoria, con probabilidad  $1 - \rho$ , el caminante se remite a su propia experiencia, y con probabilidad  $\rho$ , a la de otro miembro del grupo. Los parámetros  $q$  (tasa de uso de memoria) y  $\rho$  (tasa de transferencia de información o frecuencia con la cual los individuos se comunican entre sí) son críticos para el aprendizaje. La Fig. 4.1 presenta las reglas (i)-(iv) arriba expuestas.

### 4.1.3. Otras redes de interacción

Hemos asumido que los caminantes pueden obtener información de cualquier otro miembro del grupo. Esto resulta realista para aquellas sociedades animales en las que existe una gran fluidez en los patrones de asociación [115]. Pero, aún estos patrones de asociación fluidos pueden ser representados por redes con diferente conectividad [116]. En una red arbitraria, cada caminante (nodo  $l$ ), tiene un conjunto fijo de conexiones con otros nodos. En este caso, en la regla (ii) uno de estos nodos ( $m$ ) es escogido aleatoriamente para actualizarse. Como un ejemplo representativo, estudiaremos más adelante el caso de las redes aleatorias de Erdős-Renyi (ER), donde la conexión entre cualesquiera dos pares de nodos existe con probabilidad

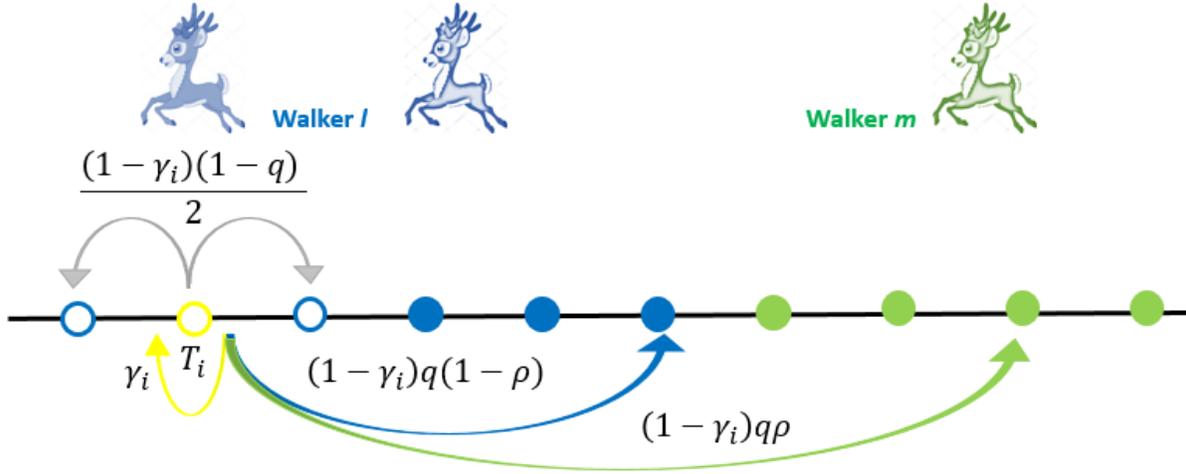


Figura 4.1: Dinámica que considera la explotación (flecha amarilla), el movimiento aleatorio (flecha gris), movimiento basado en memoria (flecha azul) y transferencia de información (flecha verde), para un caminante  $l$  localizado en el sitio de recurso  $T_i$ . Si el individuo no se encuentra en un sitio de recurso, las mismas reglas se aplican considerando  $\gamma_i = 0$ .

$s < 1$  ( $s = 1$ , corresponde a una red completa) [117, 118]. También exploramos el caso de las redes aleatorias libres de escala, construidas con el método dado en [119] donde el número de nodos con  $k$  conexiones es  $N_k \propto k^{-\alpha}$ . El exponente de la ley de potencia  $\alpha$  puede tomar cualquier valor mayor que 1. Las redes libres de escala son de interés porque contienen nodos altamente conectados, una propiedad que no tienen las redes ER. Estos nodos imitan a los individuos centrales que juegan roles importantes en las redes sociales animales reales [68, 120, 121]. Por ejemplo, en [122] se utiliza como objeto de estudio a los chimpancés y se encuentra que aquellos individuos con un estatus social alto son centrales en la red de acicalamiento y juegan un rol esencial para mantener la cohesión del grupo. La importancia de dichos individuos centrales puede ser una propiedad característica en las redes afiliativas en animales. Si bien la red social real no presenta propiedades libres de escala, el rol de los nodos centrales de la red original pueden emularse por medio de los nodos altamente conectados de una red libre de escala.

## 4.2. Resultados

### 4.2.1. Localización selectiva

Los primeros resultados que se muestran a continuación fueron obtenidos para interacciones en una red completa. En la Fig. 4.2 mostramos las posiciones iniciales y finales (después de  $t$  pasos de tiempo, con  $t$  grande) en cuatro simulaciones numéricas representativas. Los paneles superiores e inferiores corresponden a espacios en una y dos dimensiones, mientras que la columna izquierda y derecha corresponden a los casos  $\rho = 0$  y  $\rho = 0.5$  (respectivamente). En esta figura también se muestran las posiciones de los sitios de recurso con los

cinco mayores pesos  $\gamma_i$ . De manera notable, para  $\rho > 0$ , muchos de los caminantes tienden a agregarse alrededor de estos mejores sitios de recurso, creando una distribución mucho menos uniforme que la inicial, especialmente en  $1d$ . En contraste, cuando  $\rho = 0$ , la distribución de los individuos permanece bastante uniforme a través de la dinámica. Esta preferencia por ocupar algunos de los mejores sitios de recurso sobre una gran cantidad de sitios (tanto de recurso como no), puede ser referida como una **localización selectiva** inducida por un intercambio de información.

La selección de recursos por los caminantes puede ser cuantificada a través de la probabilidad de ocupación final de cada sitio de recurso  $T_i$ , después de un tiempo largo de simulación, cuando una distribución estable ha sido alcanzada: denotaremos esto por  $P_{T_i}$  y es definida por la probabilidad de que un individuo escogido al azar ocupe el sitio  $T_i$ . Esta cantidad se muestra en las Figuras 4.3(a)-(b) como función del peso  $\gamma_i$ , donde  $\rho$  se fija en 0.5 (cuadros verdes). Uno puede observar que  $P_{T_i}$  es muy pequeña y prácticamente independiente de  $\gamma_i$  en la mayor parte del intervalo  $(0, \gamma_{max})$ , mientras que crece rápidamente cuando  $\gamma_i$  se aproxima a la cota superior  $\gamma_{max}$ . En  $1d$ , hay una probabilidad muy grande, de alrededor de  $1/3$ , de estar en el sitio de recurso más pesado (de entre 50 sitios de recurso y 2000 sitios disponibles). Nuestros caminantes son fuertemente atraídos por los recursos más abundantes y por lo tanto bastante selectivos. En contraste, cuando los individuos no interactúan entre ellos ( $\rho = 0$ , curvas con círculos azules), esta selectividad prácticamente desaparece, a pesar del hecho de que los caminantes siguen usando la memoria a la misma tasa  $q$ . En este caso cada individuo se localiza alrededor de aquel sitio de recurso cercano a su posición inicial. Cuando los caminantes son simples buscadores aleatorios sin memoria ( $q = 0$ , triángulos rosas) la selectividad desaparece por completo. Por lo tanto, la emergencia de la localización selectiva es, aquí, solamente posible en sociedades conectadas y con memoria.

Para investigar cómo los individuos que interactúan escogen entre varios sitios de recursos con pesos variados, comparamos el rango calculado (o percibido) contra el rango real de los sitios de recurso. El rango real corresponde a los recursos ordenados respecto a sus pesos  $\gamma_i$  (el mejor sitio tiene rango 1) y el rango calculado corresponde a los sitios de recurso ordenados respecto su probabilidad de ocupación final  $P_{T_i}$ . Si el enjambre es capaz de escoger específicamente los mejores sitios de recurso, esperamos una gran correlación entre ambas cantidades sólo para los mejores sitios de interés y una correlación menor conforme el rango crece. Como se muestra en Fig. 4.3(c)-(d), nuestros caminantes no son solo capaces de encontrar y explotar estos mejores recursos, sino que además calculan su rango de manera bastante precisa. En nuestro ejemplo, el 10% de los mejores sitios de interés (los primeros 5 en  $1d$  y los primeros 10 en  $2d$ ), yacen en la identidad, lo cual significa que éstos son siempre identificados de manera correcta por el enjambre. Sin embargo, las fluctuaciones crecen conforme el rango lo hace. En particular, en  $1d$ , los sitios de recurso pobres (con  $\gamma_i$  baja) a veces son considerados como importantes (frecuentemente ocupados), dichos sitios se identifican como puntos que yacen por bajo de la diagonal.

La Fig. 4.4(a) muestra las variaciones con  $q$  de la probabilidad de ocupación final  $P_{T_{best}}$  definida como la probabilidad de que un caminante elegido al azar ocupe el *mejor* sitio de recurso (con el mayor  $\gamma_i$ ). Esto cuantifica el éxito de búsqueda del enjambre.  $P_{T_{best}}$  presenta

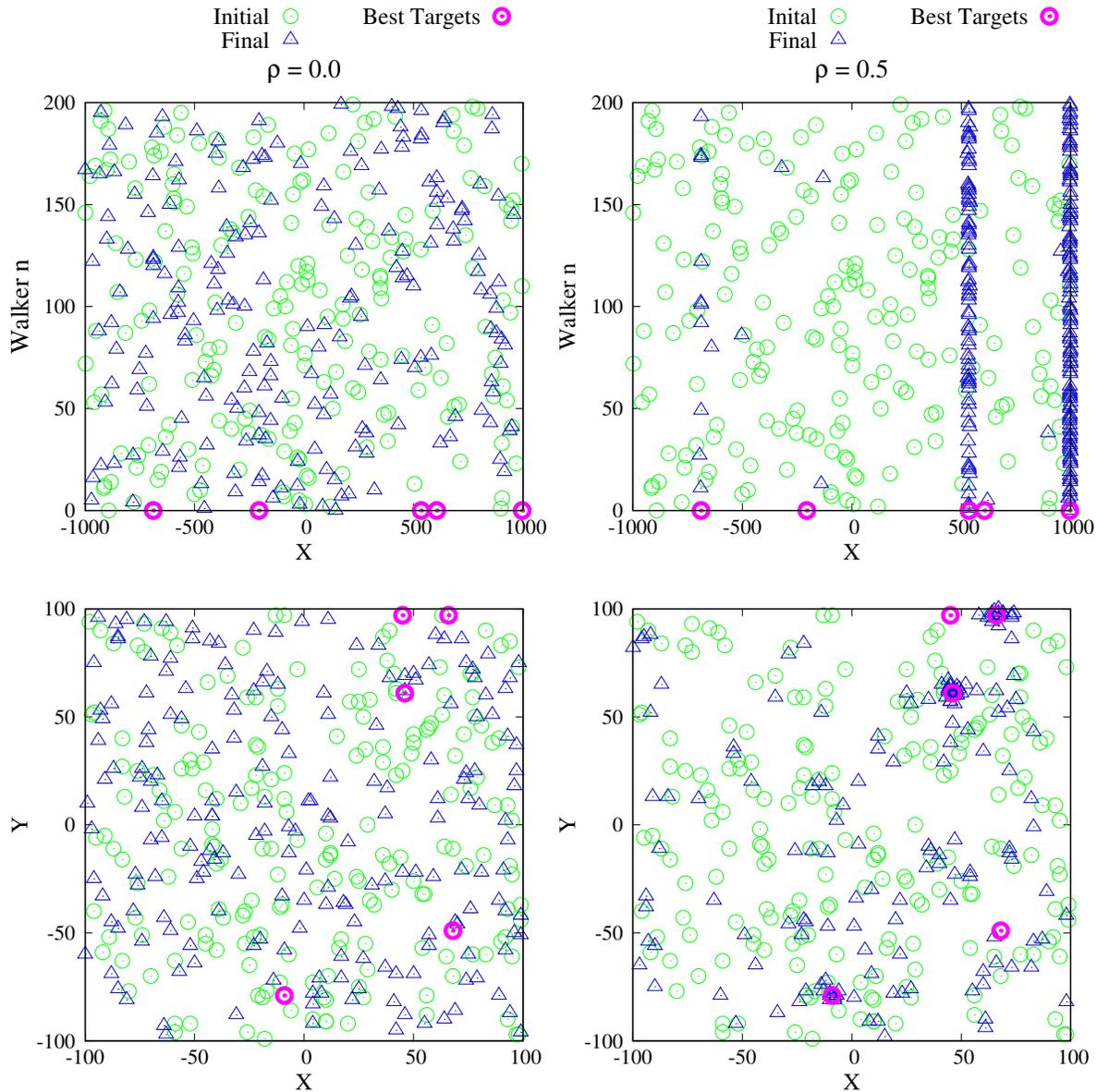


Figura 4.2: Posiciones iniciales (círculos verdes) y finales (triángulos azules) de 200 caminantes después de  $t$  iteraciones de las reglas dinámicas (i)-(iv). Los mejores cinco sitios de recurso (sobre  $M$ ) son representados por los círculos rosas. En todos los casos,  $q = 0.1$  y  $\gamma_{max} = 0.9$ . En la columna de la izquierda  $\rho = 0.0$  (sin comunicación) y en la columna de la derecha  $\rho = 0.5$ . Paneles superiores: sistema unidimensional de longitud  $L = 2000$  con  $M = 50$  y  $t = 5 \times 10^4$ . Paneles inferiores: sistema bidimensional con  $200 \times 200$  sitios,  $M = 100$  y  $t = 5 \times 10^5$ .

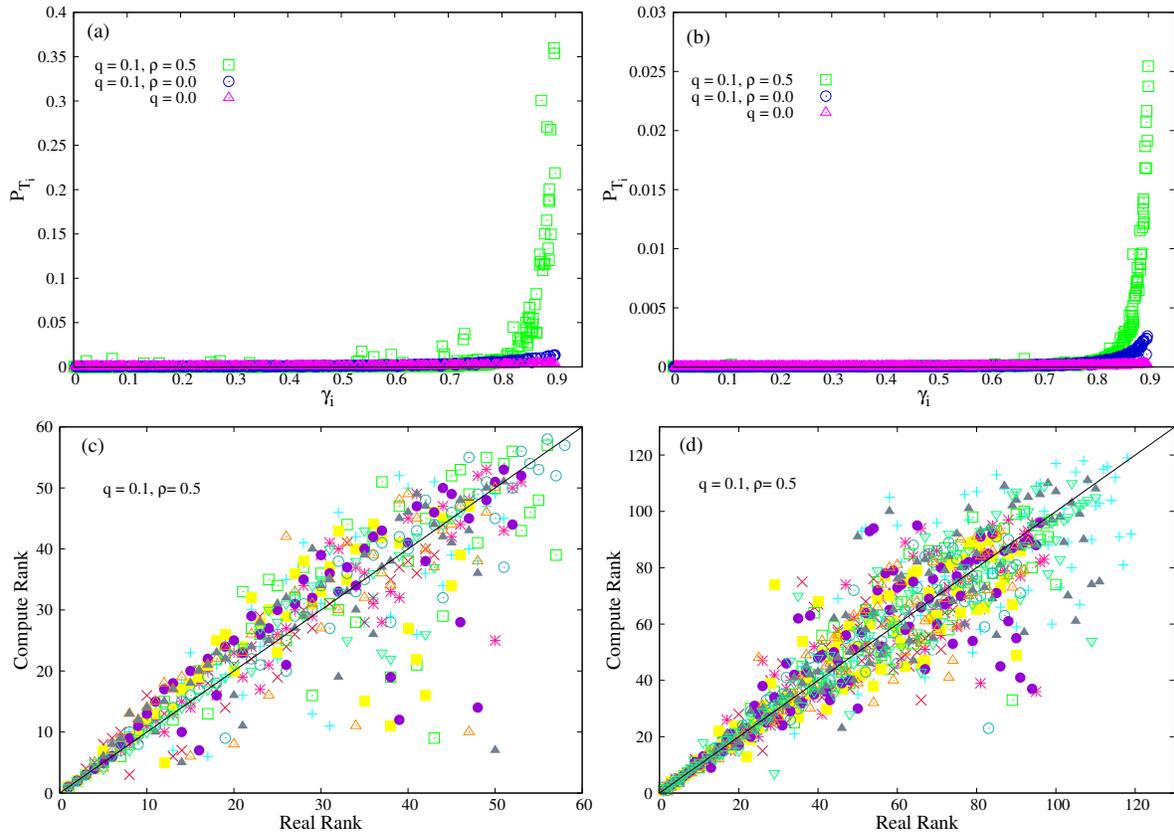


Figura 4.3: Simulaciones con  $N = 200$  y  $\gamma_{max} = 0.9$ . (a)  $P_{T_i}$  en 1d contra el valor  $\gamma_i$  correspondiente a cada sitio de interés. Cada punto representa un sitio de recursos y  $P_{T_i}$  se obtuvo del promedio sobre 1000 caminatas independientes en un mismo ambiente. Los datos se muestran para 10 ambientes diferentes. Los otros parámetros son  $t = 5 \times 10^4$ ,  $L = 2000$  y  $\delta = 0.025$  ( $M = 50$ ). (b) Misma cantidad en un espacio 2d con  $200 \times 200$  sitios y  $\delta = 0.0025$  ( $M = 100$ ) con  $t = 5 \times 10^5$ . (c) Rango percibido de un sitio de recurso (respecto a  $P_{T_i}$ , donde el rango 1 corresponde al  $P_{T_i}$  más alto) contra el rango real (respecto a  $\gamma_i$ ) para 10 diferentes ambientes unidimensionales, con los mismos parámetros que en (a). Cada punto representa un sitio de recursos en un ambiente particular y cada símbolo representa un ambiente diferente. (d) Misma gráfica en 2d, con los mismo parámetros que en (b).

un máximo para un valor particular de  $q$ . En  $1d$  un valor de  $q = 0.1$  es suficiente para que emerja el aprendizaje colectivo, o un  $P_{T_{\text{best}}}$  grande. Escogimos fijar, entonces,  $q = 0.1$  en el caso  $2d$  por consistencia y para una sencilla comparación. Valores un poco más grandes ( $q \sim 0.2$ ) llevan a una localización más fuerte en  $2d$ . La Fig. 4.4(b) presenta  $P_{T_{\text{best}}}$  como función del parámetro de interacción  $\rho$ . Un comportamiento monótonamente creciente puede observarse para valores de  $q > 0$ , alcanzando un valor “asintótico” alrededor de  $\rho = 0.5$ . En lo siguiente,  $\rho$  se fijará en éste valor. Cuando la memoria se suprime,  $P_{T_{\text{best}}}$  se anula para todos los valores de  $\rho$ .

### 4.2.2. Cohesión del enjambre

Los resultados anteriores sugieren que muchos de los individuos se agrupan alrededor del mejor sitio de recurso. Para encontrar más evidencia a favor de esta observación, enfoquémonos en estudiar la evolución temporal de la cohesión espacial del grupo. Para hacer esto, definamos la vecindad de un caminante por el segmento de longitud  $2r$  con  $r = L/(2N)$  en  $1d$  y por un disco de radio  $r = L/(2\sqrt{N})$  para el caso  $2d$ , centrados ambos en la posición del individuo. La longitud  $r$  representa la mitad de la distancia media entre dos individuos vecinos si éstos estuvieran distribuidos al azar en el espacio. Para cada caminante, al tiempo  $t$ , contamos el número de otros miembros del grupo que se encuentran en su vecindad, y tomamos el promedio ( $N_r$ ) sobre todos los caminantes y sobre mil simulaciones independientes. Un promedio sobre 10 configuraciones espaciales diferentes también se tomaron en cuenta.

La Fig. 4.4(c) muestra  $N_r$  como función del tiempo en  $2d$ . Se puede notar que no hay agregación cuando los caminantes no poseen memoria ( $q = 0$ ) o bien, cuando sí poseen memoria pero son incapaces de intercambiar información ( $q > 0$  y  $\rho = 0$ ); en estos casos, cada individuo tiene en promedio un solo vecino ( $N_r \approx 1$ ) para todo  $t$ . En contraste, cuando los agentes poseen memoria y la capacidad de comunicarse ( $q > 0$  y  $\rho > 0$ )  $N_r$  crece con el tiempo y alcanza valores considerablemente grandes, sobre todo en  $1d$  (Fig. 4.4 - Inset). Este resultado, junto con los ya presentados para esta sección, nos indican que los individuos se agrupan alrededor de los mejores sitios de recurso. Está agrupación selectiva es una consecuencia del aprendizaje colectivo y nos indica un gran éxito de búsqueda a nivel del grupo.

### 4.2.3. Efecto de las redes de interacción

Ahora bien, investiguemos el efecto de la conectividad de la red de transferencia de información entre los individuos en la localización selectiva. Para una red de interacción ER, la Fig. 4.4 muestra a  $P_{T_{\text{best}}}$  como función del número de conexiones promedio por caminante (grado promedio),  $\langle k \rangle = s(N - 1)$ , donde  $s \in [0, 1]$  es el parámetro de conectividad. La probabilidad de ocupación final del mejor sitio de recurso es muy baja para valores pequeños de  $\langle k \rangle$  pero crece rápidamente y se satura, esto sucede para  $1d$  y  $2d$ . Tan pronto como los caminantes tienen en promedio seis o siete conexiones,  $P_{T_{\text{best}}}$  alcanza el mismo valor que el de para una red completa ( $\langle k \rangle = 199$  en este caso). Notemos que para  $\langle k \rangle \approx 6$  el componente gigante de la red está ya formado, ya que el umbral de percolación para las redes

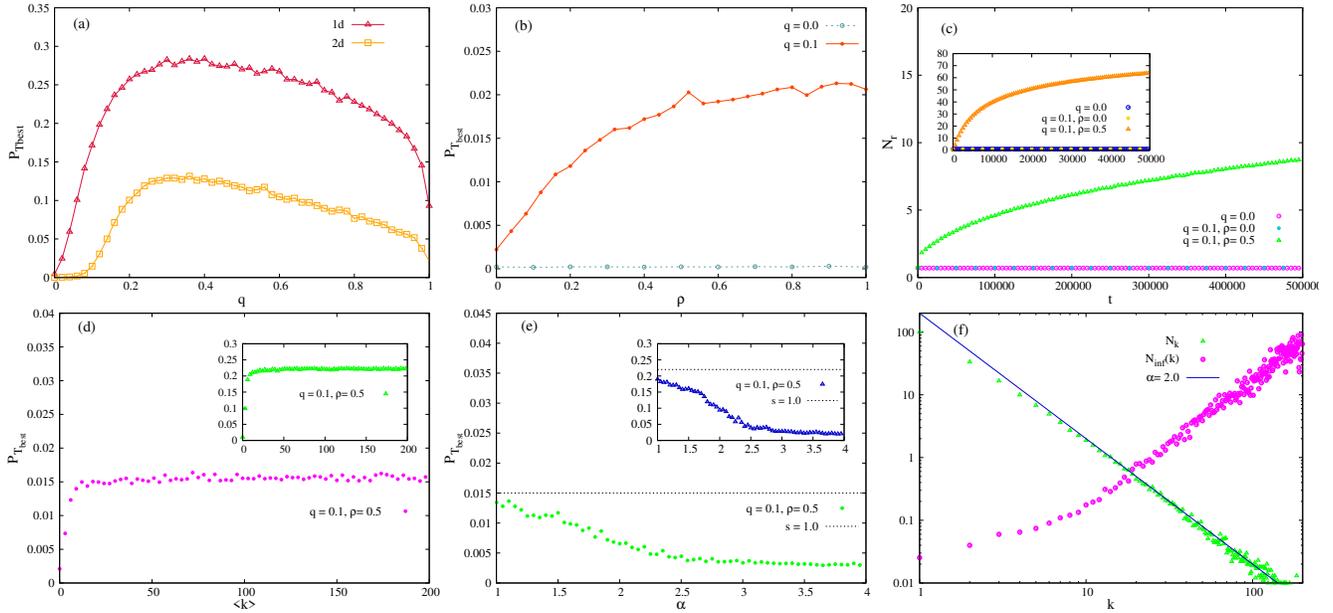


Figura 4.4: Simulaciones con  $N = 200$  individuos,  $M = 50$  sitios de recursos en  $1d$ ,  $M = 100$  en  $2d$  y  $\gamma_{max} = 0.9$ : (a)  $P_{T_{best}}$  como función de la tasa de uso de memoria  $q$  en una y dos dimensiones. (b)  $P_{T_{best}}$  como función del parámetro de interacción  $\rho$  en dos dimensiones para una  $q$  fija. (c) Promedio del número de vecinos a distancia  $r$  de un individuo en  $2d$  ( $r = 7$ ). Inset: Caso  $1d$  con  $r = 5$ . (d)  $P_{T_{best}}$  como función del grado promedio  $\langle k \rangle$  en una red de interacción ER para un ambiente  $2d$ . Inset: Caso  $1d$ . (e)  $P_{T_{best}}$  en  $2d$  como función del exponente de distribución de grado  $\alpha$  es una red de interacción libre de escala. La línea punteada representa el valor de una gráfica de interacción completa. Inset: Caso  $1d$ . (f)  $N_{inf}(k)$  es el número promedio de individuos que alcanzan por primera vez uno de los mejores sitios de recurso (en  $2d$ ) por utilizar la experiencia de un nodo vecino de grado  $k$ . La red es una red libre de escala con exponente  $\alpha = 2$  y la distribución de grado  $N_k$  se muestra en triángulos verdes. Notemos la gran importancia de los nodos altamente conectados para transferir información. En todos los casos se muestra un promedio sobre diez ambientes diferentes.

aleatorias ocurre en  $\langle k \rangle_c = 1$  [117, 118]. Por lo tanto, el aprendizaje colectivo requiere muchas menos conexiones que las de una red completa, pero una red apenas arriba del umbral de percolación tampoco es suficiente.

Los resultados para una red libre de escala con exponente de distribución de grado  $\alpha$  (vea la ref. [119] para los detalles de la construcción de la red) se muestran en la Fig. 4.4(e). La probabilidad de ocupación final del mejor sitio de recurso alcanza su máximo para  $\alpha = 1$  (el valor mínimo de  $\alpha$  debido a la condición de normalización de la distribución de grado), que es igual a la de la gráfica completa.  $P_{T_{\text{best}}}$  decrece con  $\alpha$ , i.e., conforme la red se hace menos heterogénea. Notemos, sin embargo, que la conectividad media también decrece con  $\alpha$  en este modelo de red. Para  $\alpha > 3$ , las redes son menos conectadas (más dispersas) y exhiben una menor localización. La Fig. 4.4(f) muestra que los nodos más conectados (“hubs”) son críticos para la transferencia de información: el número de caminantes que usan el conocimiento de un nodo de grado  $k$  para llegar por primera vez a uno de los mejores sitios de recurso,  $N_{\text{inf}}(k)$ , es una función que crece rápidamente con  $k$ . Así, los nodos más conectados influyen de manera crucial en las decisiones exitosas de muchos de los otros nodos.

Ahora bien, en la sección 3.4 de este mismo escrito se estudió a profundidad el efecto que los Vuelos de Lévy (VL) tiene sobre el aprendizaje para un individuo aislado que explora un ambiente que posee un solo sitio de recurso. Desde el punto de vista ecológico, se ha probado que muchas especies animales tienden a explorar el espacio a su disposición realizando VL [13–15], es decir, sus trayectorias se encuentran caracterizadas por pasos cortos y menos frecuentes pasos largos que los llevan de una vez a sitios del espacio alejados de sus posiciones actuales. Más aún, se ha demostrado que para muchas de estas especies esta estrategia de búsqueda resulta ser óptima [13, 15, 123, 124]. Interesa pues averiguar si dicha estrategia de búsqueda que permite a varias especies animales explorar y explotar el ambiente a su disposición de manera eficiente tiene los mismos efectos dentro de un grupo formado por individuos que poseen la capacidad de uso de memoria y la habilidad de comunicarse entre ellos. Modificamos pues la regla de movimiento (iii). En la siguiente sección se explica esto a detenimiento.

### 4.3. Vuelos de Lévy

En esta sección abordaremos una leve modificación a las dinámica colectiva abordada en la sección previa. Recalamos que la construcción del ambiente de búsqueda, el número de individuos, la forma en la que usan la memoria, y la manera en la que se comunican entre ellos, se mantendrán como hasta ahora. Las reglas de actualización serán las mismas a excepción de la regla de movimiento (iii). ésta se modificará como sigue:

- (iii) *Movimiento aleatorio*: Si el individuo  $l$  no está en un sitio de recurso, con probabilidad  $(1 - q)$  se desplaza una longitud aleatoria  $\ell_t$  escogida de una distribución simétrica  $p(\ell)$ . Consideraremos específicamente una distribución de Lévy definida como:

$$p(\ell) = \frac{C}{|\ell|^{1+\mu}}, \quad \ell = \pm 1, \pm 2, \pm 3, \dots \quad (4.1)$$

con  $C$  una constante de normalización y  $0 \leq \mu \leq 3$ .

Entonces,  $\mathbf{X}_{t+1}^{(l)} = \mathbf{X}_t^{(l)} + \ell_t$ . Para el caso  $2d$  cada una de las coordenadas se actualizará de manera independiente.

Antes de exponer los resultados obtenidos de los experimentos numéricos para esta nueva regla dinámica, recordemos que el caso de primeros vecinos ha sido ya estudiado en [2] y expuesto en las secciones previas de este capítulo 4. El enjambre cuyos individuos se desplazan sólo a sus sitios inmediatos más cercanos exhibe aprendizaje colectivo debido al uso combinado de memoria y comunicación entre miembros del grupo. El enjambre es capaz de distinguir los mejores sitios de recursos y localizarse alrededor de ellos.

A menos que se mencione lo contrario, los resultados que mostraremos a continuación se obtuvieron para una red de interacción completa y un ambiente en dos dimensiones. En la Fig. 4.5 se muestra la probabilidad de ocupación final de cada sitio de interés  $T_i$ , denotada por  $P_{T_i}$ , como función de sus pesos  $\gamma_i$ . Las Figs. 4.5 (a)-(b) muestran el cambio de  $P_{T_i}$  para un enjambre que sigue una dinámica de Vuelos de Lévy en un ambiente uni y bidimensional, respectivamente. Uno puede observar que  $P_{T_i}$  es mayor para aquellos  $\gamma_i$  cercanos a  $\gamma_{max}$  y decae rápidamente con  $\gamma_i$ . En  $1d$  existe una probabilidad mayor a  $1/2$  de estar en el mejor sitio de recurso ( $\rho = 0.5$ , triángulos verdes). Aquí la localización selectiva es bastante fuerte, incluso más que la mostrada en las secciones previas y en [2] para una dinámica de primeros vecinos. Algo similar ocurre para el caso  $2d$ . Una cosa importante para recalcar es que cuando los agentes del enjambre no tienen la habilidad de comunicarse entre ellos ( $\rho = 0$ , cuadros rosas) la explotación de los mejores recursos sigue ocurriendo pero con una fuerza menor y no es tan selectiva como cuando la transferencia de comunicación está presente, algo totalmente diferente a lo que ocurre para la dinámica a primeros vecinos donde la localización selectiva desaparece por completo en ausencia de la comunicación entre agentes [2]. Cuando los caminantes no poseen memoria ( $q = 0$ , círculos azules) la localización se desvanece en su totalidad. En este caso los individuos no son capaces de aprender, siquiera, de su propia experiencia. Por lo tanto, la localización selectiva en un enjambre gobernado por VL es posible cuando los individuos del mismo poseen memoria individual y se ve reforzada por la transferencia de información.

Las Figs. 4.5 (c)-(d) muestran el cambio de  $P_{T_i}$  con  $\gamma_i$  para una dinámica de primeros vecinos y de VL, respectivamente. Es fácil ver que además de tener una mayor localización en los mejores sitios de recurso, los VL también mejoran la selectividad, es decir, la pendiente de la curva es mayor que la de primeros vecinos. Aquellos sitios de recursos que eran importantes para el enjambre a primeros vecinos ya no lo son más para el enjambre con VL. Para entender esta pequeña pero significativa diferencia presentaremos una comparación más profunda entre estos dos tipos de dinámicas midiendo diferentes cantidades que nos dan información sobre la efectividad del enjambre. Fijamos el parámetro de transferencia de información en  $\rho = 0.5$ .

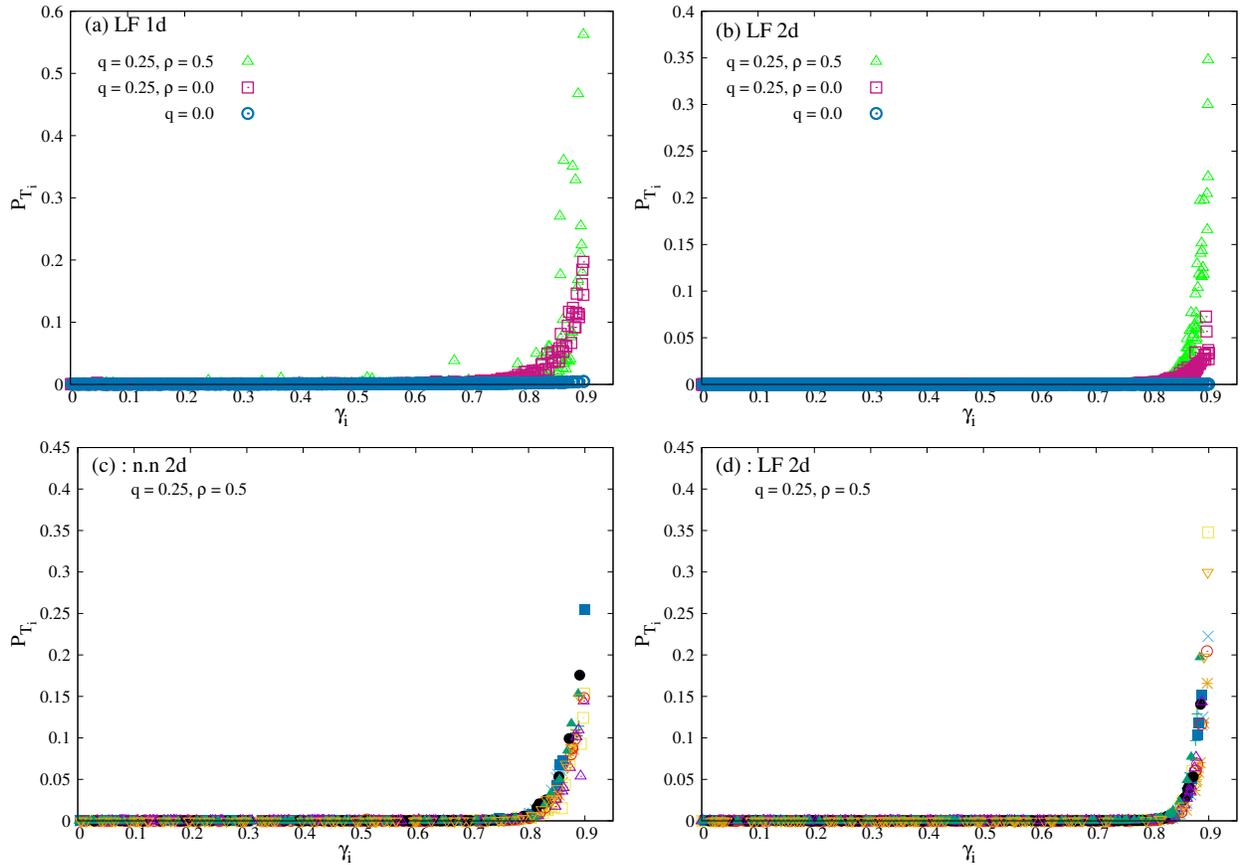


Figura 4.5: Simulaciones con  $N = 200$  individuos y  $\gamma_{max} = 0.9$ : (a) Probabilidad de ocupación final a cada sitio de recurso  $P_{T_i}$  en  $1d$  para una dinámica de VL con  $\mu = 0.5$  como función de los pesos  $\gamma_i$  y para distintos valores de  $q$  y  $\rho$ . Los otros parámetros son  $t = 10^4$ ,  $L = 2000$  y  $\delta = 0.025$  ( $M=50$ ). (b) Misma curva que (a) pero para un ambiente  $2d$  con  $L = 200 \times 200$  sitios,  $\delta = 0.0025$  ( $M=100$ ) y  $t = 10^5$ . (c) Probabilidad de ocupación final a cada sitio de interés  $P_{T_i}$  en  $2d$  para una dinámica de primeros vecinos como función del peso  $\gamma_i$ . (d) Misma curva que en (c) pero para VL con  $\mu = 0.5$ . En todos los casos cada punto representa un sitio de recurso y  $P_{T_i}$  se obtiene de promediar sobre 1000 caminatas independientes en el mismo ambiente. Se muestran los datos de diez configuraciones espaciales diferentes.

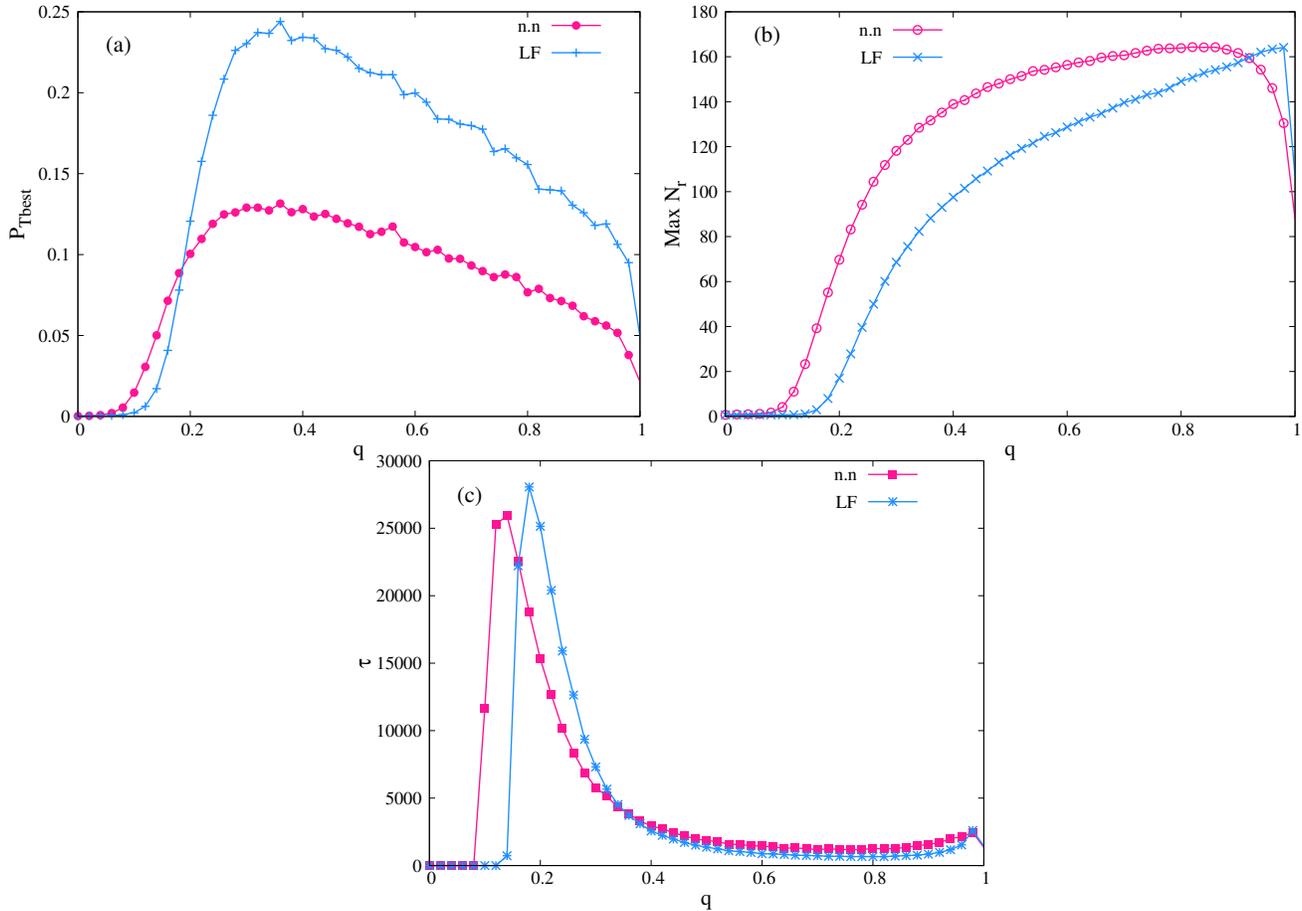


Figura 4.6: (a) Probabilidad de ocupación del mejor sitio de recurso  $P_{T_{\text{best}}}$  como función del parámetro de memoria  $q$ . (b) Cohesión máxima del enjambre como función del parámetro de memoria  $q$ . (c) Tiempo medio que le toma al enjambre alcanzar su máxima cohesión,  $\tau$ , como función del parámetro de memoria  $q$ . Todas las curvas son un promedio sobre mil caminatas independientes en el mismo ambiente y sobre diez ambientes diferentes. Los valores para los parámetros restantes son los mismos que los considerados en las Figs. 4.5 (c)-(d).

La Fig. 4.6 (a) muestra las variaciones con  $q$  de la probabilidad de ocupación del mejor sitio de recurso  $P_{T_{\text{best}}}$  definida como la probabilidad de que un caminante elegido al azar ocupe el mejor sitio de recurso (con el mayor peso  $\gamma_i$ ) para ambas dinámicas; a primeros vecinos (curva rosa) y VL (curva azul). Uno puede ver que el mínimo de memoria necesaria para que el enjambre muestre aprendizaje colectivo es menor en la dinámica a primeros vecinos que en la de VL pero cuando  $q \gtrsim 0.2$  el enjambre de VL presenta un mejor desempeño en la búsqueda global. En ambos casos un uso excesivo de memoria lleva a una menor localización y ésta tiende a desaparecer conforme  $q \rightarrow 1$  (en este régimen los individuos tienden a quedarse “atrapados” en los primeros sitios que visitan, reduciendo así su movilidad y por lo tanto su capacidad de exploración).

La Fig. 4.6 (b) muestra la máxima cohesión alcanzada por el enjambre durante la búsqueda, denotada por  $\text{Max } N_r$  y que representa el valor asintótico de la cohesión del grupo a

tiempos grandes, como función del parámetro de uso de memoria  $q$ . Aquí, al contrario de lo que pasa en el panel anterior, la cohesión del enjambre a primeros vecinos es mayor que la del enjambre de VL. Esto es esperado ya que la dinámica de primeros vecinos, con un uso de memoria frecuente, lleva a una exploración del espacio menos difusiva que la que puede obtenerse con una dinámica de VL. En ambos casos la cohesión máxima decae de manera abrupta cuando  $q \approx 1$ .

En la Fig. 4.6 (c) podemos ver el cambio del tiempo típico que el enjambre tarda en alcanzar su máxima cohesión con  $q$ . Denotaremos esta magnitud por  $\tau$ , el cálculo de este valor es sencillo; una vez obtenida la  $\text{Max } N_r$  calculamos, en la dinámica temporal, el  $t$  en el que el grupo alcanza la mitad de su máxima cohesión y ese valor de  $t$  será nuestro valor típico  $\tau$ . Podemos observar que ambas curvas alcanzan un máximo, para diferentes valores de  $q$ , y después este máximo en  $\tau$  decae lentamente hasta alcanzar un valor asintótico que es más o menos el mismo para ambas curvas.

Si vemos todos estos resultados a la vez y tomamos la información que cada uno de ellos nos proporciona podemos hacer interpretaciones cualitativas importantes. Aún y cuando el enjambre de primeros vecinos converge de manera “más rápida” que en enjambre de VL (un valor de  $\text{Max } N_r$  más grande y un  $\tau$  menor) éste no se localiza de manera tan efectiva alrededor del mejor sitio de recurso (menor  $P_{T_{\text{best}}}$ ). Esto significa que hay una menor cantidad de individuos, comparada con la del enjambre de VL, que están explotando el mejor sitio de recurso. Por otro lado, aún y cuando los VL no tienden a fortalecer la cohesión del enjambre, sí fomenta la localización alrededor del mejor sitio de recurso. Podemos decir, pues, que los VL resultan más efectivos que una dinámica simple a primeros vecinos para encontrar y explotar, de manera grupal, el mejor sitio de recurso.

## 4.4. Decaimiento de Memoria

Ahora bien, hasta el momento hemos supuesto que cada uno de nuestros individuos posee una memoria infinita, es decir, son capaces de recordar cada uno de los sitios visitados en el pasado sin importar hace cuánto tiempo éstos hayan sido visitados y por lo tanto, cuando la información se transfiere de un agente a otro se hace suponiendo que la memoria compartida es también infinita. Esto es un supuesto bastante fuera de la realidad ya que si bien muchas especies animales poseen memoria, ésta tiende a decaer con el paso del tiempo. Es decir, los recuerdos se difuminan en el pasado y así es más probable que el animal elija volver a un sitio visitado recientemente que a uno visitado en un pasado lejano. Ejemplos como el anterior son los expuestos por Briese y French para los leones tamarinos dorados y el tití de orejas negras [125], ambas especies de primates que tienen distintos rangos de retención de información y en los cuales el decaimiento de memoria agiliza el éxito de forrajeo pues estos animales actualizan de manera constante los detalles que poseen de su medio cambiante. Está también el caso bien estudiado por Merkle *et al.* para los bisontes canadienses [16], en este estudio teórico-experimental se propone un modelo de forrajeo con decaimiento de memoria que va como una ley de potencia para ajustar los datos de campo obtenidos de 33 bisontes.

En este estudio se concluye que un justo medio entre memoria y olvido provee al animal con la capacidad de forrajear su ambiente de manera óptima. Estos resultados en ecología y muchos otros que tratan el decaimiento de memoria [126–128] consideran sólo el análisis sobre individuos aislados pero no en el colectivo. Así pues, con el fin de estudiar el efecto que el decaimiento de memoria puede tener sobre el éxito de forrajeo grupal, en esta sección consideraremos este nuevo supuesto para modificar la dinámica de nuestro enjambre.

Supongamos entonces que los individuos del enjambre tienen un deterioro (decaimiento) en su memoria, esto significa que tienen una tendencia a olvidar los lugares visitados en el pasado lejano. Entonces, cuando eligen hacer uso de su experiencia para reubicarse en el espacio o cuando eligen comunicarse entre ellos, los sitios visitados recientemente tiene una mayor probabilidad de volver a ser visitados que aquellos que lo fueron en un pasado lejano. Para incluir esta nueva asunción en nuestra dinámica nos inspiraremos en la regla de decaimiento expuesta en [129] para modificar nuestras reglas de movimiento (i) y (ii) como sigue:

- (i) *Memoria propia*: Si el agente  $l$  no se encuentra en un sitio de recurso, con probabilidad  $q(1 - \rho)$  el caminante regresa a uno de los sitios ya visitado en el pasado, esto es:  $\mathbf{X}_{t+1}^{(l)} = \mathbf{X}_{t'}^{(l)}$  donde  $t'$  es un entero aleatorio en el intervalo  $[0, t]$  escogido de acuerdo a una distribución de probabilidad  $p_t(t')$  dada *a priori* y definida como:

$$p_t(t') = \frac{F_t(t')}{C(t)} \quad (4.2)$$

con  $C(t) = \sum_{t'=0}^t F_t(t')$  un factor de normalización y  $F_t(t')$  la función del kernel de memoria. Aquí consideraremos un kernel de memoria de ley de potencia dado por:

$$F_t(t') = (t - t' + 1)^{-\beta} \quad (4.3)$$

Si  $\beta = 0$  recuperamos el modelo de visitas preferenciales lineal estudiado anteriormente [2]. Así si  $\beta > 0$  es más probable que el individuo escoja sitios visitados recientemente para reubicarse en lugar de aquellos que fueron visitados en un pasado lejano. Este es el esquema de revisita preferencial en ley de potencia mencionado arriba. La Fig. 4.7 muestra la comparación entre estos dos casos ( $\beta = 0$  y  $\beta > 0$ ).

- (ii) *Transferencia de información*: Si el individuo no está en un sitio de recurso, con probabilidad  $q\rho$  el caminante ( $l$ ) escoge de manera aleatoria a uno de sus vecinos en la red de comunicación ( $m$ ) y se reubica en un sitio visitado por este vecino siguiendo la misma regla de visita preferencial explicada arriba:  $\mathbf{X}_{t+1}^l = \mathbf{X}_{t'}^{(m)}$  donde  $t'$  es también un entero aleatorio en el intervalo  $[0, t]$  escogido de acuerdo a la distribución de probabilidad  $p_t(t')$  definida en (4.2) y (4.3).

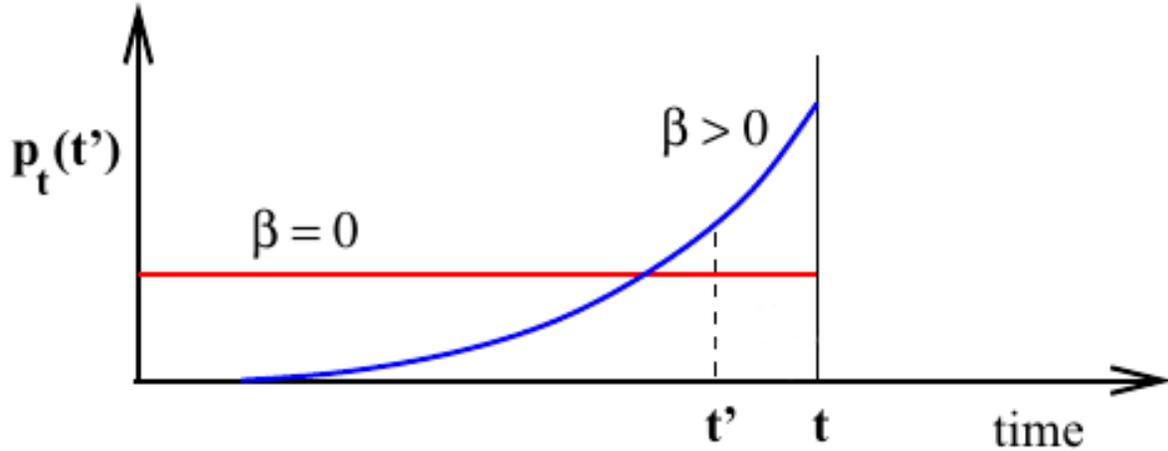


Figura 4.7: Distribución de probabilidad para  $t'$  para el modelo de visita preferencial lineal ( $\beta = 0$ ) y modelo con decaimiento de memoria ( $\beta > 0$ ). Esta figura fue tomada de [129].

Considerando estas nuevas modificaciones a nuestra dinámica de movimiento, para cada uno de los agentes del enjambre, obtuvimos los siguientes resultados: La Fig. 4.8 presenta la probabilidad de ocupación de cada sitio de recurso como función de sus pesos  $\gamma_i$ , se fijó  $\beta \approx 1$  (este valor de  $\beta$  marca el intermedio entre memoria perfecta y decaimiento rápido, por otro lado, resultados previos muestran que este valor de  $\beta$  maximiza la efectividad de la búsqueda en un individuo aislado [130].) En las Figs. 4.8 (a)-(b) podemos ver el valor de  $P_{T_i}$  para enjambres con dinámica de primeros vecinos y enjambres de VL, respectivamente. Aquí, similar a lo ya expuesto en las secciones anteriores y en [2] para el caso de memoria infinita,  $P_{T_i}$  crece de manera abrupta cuando  $\gamma_i$  es cercano a  $\gamma_{max}$  cuando la memoria y comunicación están presentes en el grupo (triángulos grises). Nuestros individuos son atraídos hacia los mejores sitios de recurso y hacen una localización bastante selectiva. Cuando la comunicación se suprime ( $\rho = 0$ , círculos azules),  $P_{T_i}$  es muy pequeña para todos los  $\gamma_i$  en el enjambre de primeros vecinos, esto es porque sin comunicación los individuos sólo aprenden de su propia experiencia y se localizan alrededor de aquellos sitios de recursos cercanos a sus posiciones iniciales. Pero para el enjambre de VL la ausencia de comunicación no impide la selectividad. Cuando el enjambre se comunica, sin embargo, la localización selectiva es dos veces más grande que cuando la comunicación se suprime. Cuando los individuos no poseen memoria ( $q = 0$ , cruces verdes) la localización desaparece por completo en ambas dinámicas.

El efecto del decaimiento de memoria se hace mucho más visible en las Figs. 4.8 (c)-(d) que comparan las curvas  $P_{T_i}$  para los enjambres con VL para el caso de memoria infinita (c) y decaimiento de memoria (d) (el caso a primeros vecinos no se compara en esta figura pero el estudio se hará más adelante). Podemos ver que el decaimiento de memoria mejora de manera significativa la localización, aumentándola a poco más del doble, y que la rapidez con la que  $P_{T_i}$  decae es un poco más rápida para el caso  $\beta = 1$  que para el caso  $\beta = 0$ . Con esto podemos decir que el enjambre con decaimiento de memoria es mucho más selectivo que el enjambre con memoria infinita. Parece ser, pues, que darle a los individuos la posibilidad de olvidar los sitios que han sido visitados en un pasado muy lejano, le proporciona al grupo una mucho mejor posibilidad de encontrar y explotar los mejores sitios de recursos.

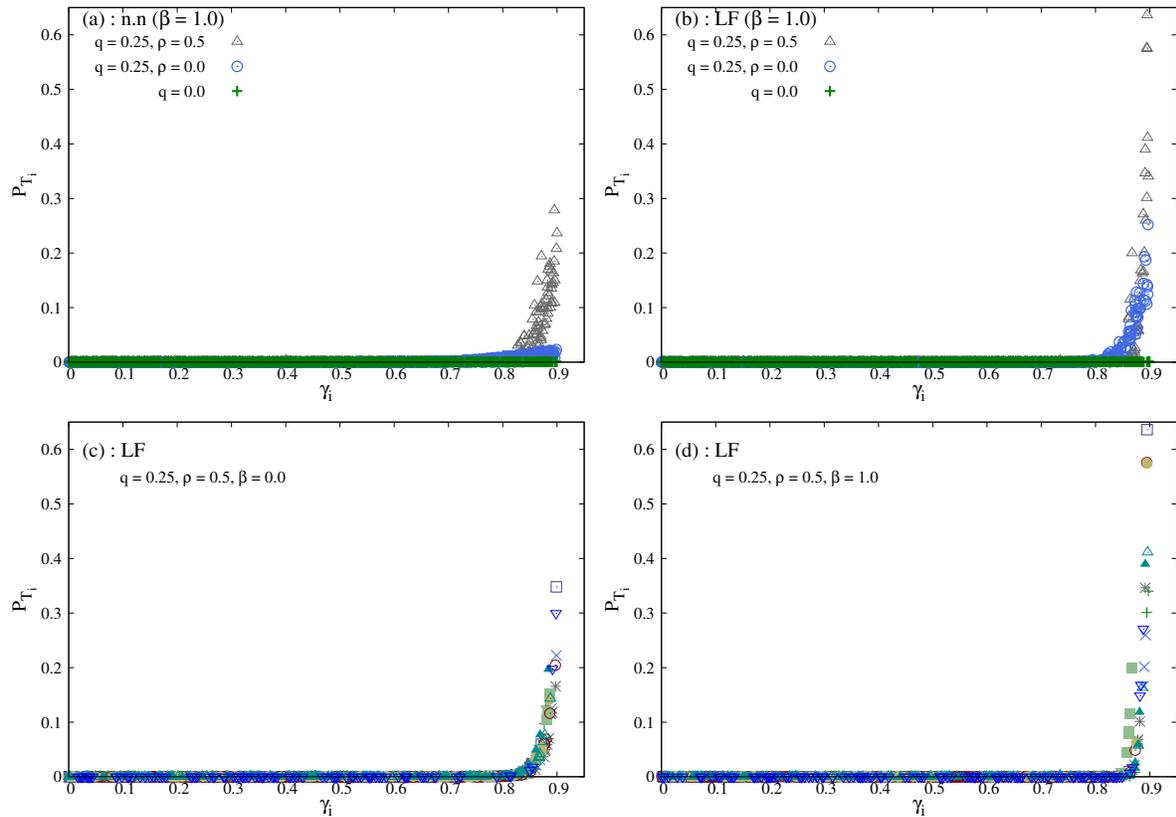


Figura 4.8: Decaimiento de memoria con  $\beta \approx 1$  en  $2d$ . Simulaciones con  $N = 200$  individuos y  $\gamma_{max} = 0.9$ . (a) Probabilidad de ocupación de cada sitios de recurso  $P_{T_i}$  para la dinámica a primeros vecinos como función de los pesos  $\gamma_i$  y para distintos valores de  $q$  y  $\rho$ . (b) Misma curva que en (a) pero para un enjambre de Vuelos de Lévy con  $\mu = 0.5$ . (c) Misma curva que en (a) considerando solo el caso  $q = 0.25, \rho = 0.5$ . (d) Misma curva que en (b) considerando solo el caso  $q = 0.25, \rho = 0.5$ . En todos los casos cada punto representa un sitio de recurso y  $P_{T_i}$  se obtiene de promediar sobre 1000 caminatas aleatorias independientes en un mismo ambiente. Se muestran los resultados para 10 ambientes diferentes.

Para comparar de manera más directa y cuantitativa el impacto que el decaimiento de memoria tiene sobre el éxito de búsqueda y explotación de los recursos en el ambiente del enjambre hicimos los mismos experimentos numéricos para  $P_{T_{\text{best}}}$ ,  $\text{Max } N_r$  y  $\tau$  variando los valores de  $\beta$  tanto para los enjambres con dinámica a primeros vecinos y VL. Se fijaron los valores de  $q = 0.25$  y  $\rho = 0.5$ . La Fig. 4.9 muestra la probabilidad de ocupación final del mejor sitio de recurso ( $P_{T_{\text{best}}}$ ) - Arriba como función de  $\beta$  para primeros vecinos (curva amarilla) y VL (curva verde). Para ambas dinámicas, valores intermedios de  $\beta$  mejoran la localización, es decir, una alta capacidad de memoria ( $\beta \rightarrow 0$ ) lleva a los individuos y al grupo reforzar aquellos sitios visitados en el inicio y que no necesariamente resultan sitios ricos en recursos, por otro lado una muy baja capacidad de memoria ( $\beta \rightarrow 3$ ) se traduce en un olvido rápido y por lo tanto en una falta de reforzamiento de aquellos sitios que no hayan sido visitados en el presente o pasado muy cercano. Así, ambas curvas crecen con  $\beta$  hasta alcanzar un máximo ( $\beta \approx 1$  para primeros vecinos y  $\beta \approx 1.5$  para VL), luego de ese óptimo la curva decae de manera lenta de tal suerte que un valor elevado del índice de decaimiento de memoria lleva a una localización menor. Es importante notar que la mejora en la localización debido al decaimiento de memoria es mucho más alta y significativa para el enjambre VL que para el enjambre a primeros vecinos, esto se puede observar en el hecho de que la curva para VL crece mucho más rápido que la de primeros vecinos y, de hecho, alcanza a duplicar su valor inicial ( $\beta = 0$ ) cuando  $\beta$  llega a valores intermedios. La localización es siempre mayor para VL que para primeros vecinos.

La Fig. 4.9 (b) muestra las variaciones de la cohesión del enjambre  $\text{Max } N_r$  con  $\beta$ . De nuevo, ambas curvas crecen hasta alcanzar un máximo y luego tienden a decrecer lentamente. El enjambre a primeros vecinos alcanza una cohesión mayor que el de VL pero es importante notar que estos máximos no se alcanzan en los mismos valores de  $\beta$  que  $P_{T_{\text{best}}}$ , lo cual significa que no podemos asociar directamente el hecho de que una mayor cohesión lleva a una mejor localización y explotación del mejor sitio de recurso.

La Fig. 4.9 (c) presenta el cambio de  $\tau$  (el tiempo típico que tarda el enjambre para agruparse alrededor de los mejores sitios de recurso) con el índice de decaimiento de memoria  $\beta$ . Ambas curvas decrecen fuertemente conforme  $\beta$  crece y ya no varían mucho luego de  $\beta \approx 1$ . Es importante destacar que cuando los individuos del enjambre (VL o primeros vecinos) poseen una capacidad de memoria infinita ( $\beta = 0$ ), el grupo tarda mucho tiempo ( $\tau$  muy grande) en encontrar y agruparse alrededor de los mejores recursos mientras que el olvido acelera de manera drástica esta convergencia. Notemos también que la curva para el enjambre de primeros vecinos está siempre por debajo de la VL, esto es un comportamiento esperado ya que a una menor difusión al enjambre le cuesta mucho menos cohesionarse y converger.

Así pues, interpretando los tres resultados anteriormente discutidos, podemos concluir que un valor “óptimo” de decaimiento de memoria debería de realizar el compromiso adecuado entre una localización robusta y una convergencia rápida, ese valor óptimo resulta ser  $\beta \approx 1.5$  para enjambre con VL y  $\beta \approx 1$  para enjambres a primeros vecinos.

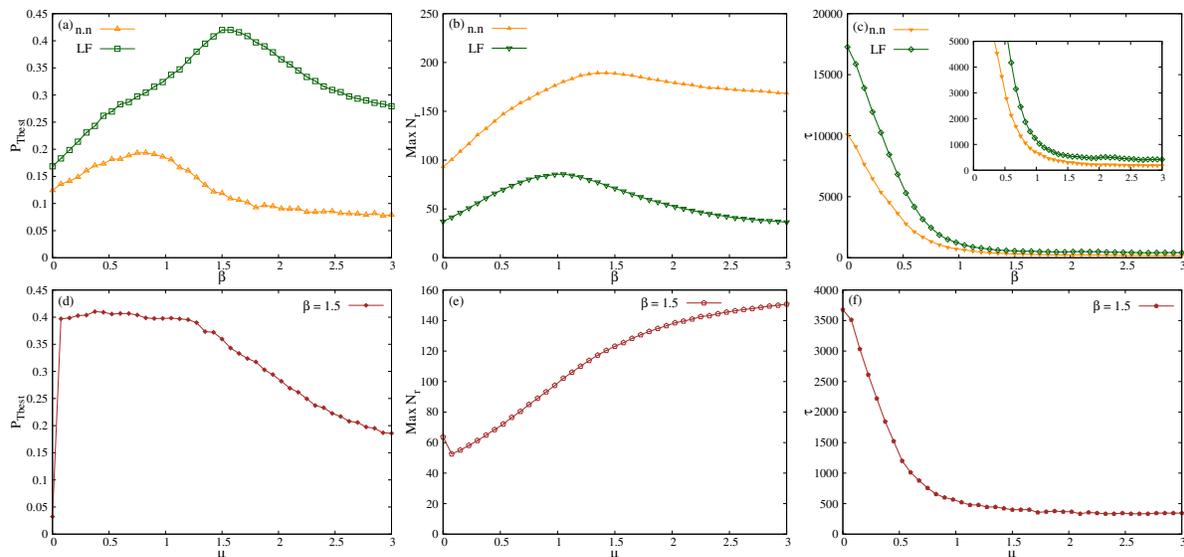


Figura 4.9: Decaimiento de memoria en  $2d$ . (a) Probabilidad de ocupación del mejor sitio de recurso como función del parámetro de decaimiento de memoria  $\beta$ . (b) Máxima cohesión alcanzada por el enjambre durante la búsqueda como función de  $\beta$ . (c) Tiempo medio que tarda el enjambre en alcanzar su máxima cohesión como función de  $\beta$ . (d)-(e)-(f) Mismas cantidades mostradas en (a)-(b)-(c) respectivamente pero como función del índice de Lévy  $\mu$ . Todas las curvas son el promedio de 10 ambientes diferentes. Los valores de los parámetros restantes son los mismos considerados en la Fig.4.8 (c)-(d), respectivamente.

Con el propósito de comparar la efectividad del enjambre durante la búsqueda dependiendo de con qué tanta frecuencia sus individuos dan pasos largos, medimos el cambio de estas mismas tres cantidades pero ahora como función del índice del Vuelo de Lévy  $\mu$  (ver Ec.(4.3)), fijando  $\beta = 1.5$  que es el valor aproximado en el cual el enjambre alcanza su mejor nivel de explotación al mejor recurso. Un valor de  $\mu$  cercano a cero nos dará una dinámica superdifusiva en la cual cada individuo escoge siempre (para su movimiento aleatorio) dar pasos largos. Por otro lado si  $\mu$  es mayor que tres, la dinámica que obtenemos es muy similar a la de primeros vecinos en la cuál se favorece la exploración local y la reubicaciones a distancias largas son mucho menos frecuentes.

La Fig. 4.9 (d) muestra las variaciones de  $P_{T_{best}}$  como función de  $\mu$ , aquí la curva crece abruptamente para  $\mu$  apenas mayor a cero manteniendo valores grandes hasta  $\mu \approx 1.5$  cuando empieza a decrecer conforme  $\mu$  aumenta. Este resultado es consistente con lo mostrado en la Fig. 4.9 (a). Una dinámica con valores de intermedios de  $\mu$  junto con un decaimiento en la memoria y la comunicación, mejora la capacidad del enjambre de encontrar y explotar el mejor sitio de recurso.

La Fig. 4.9 (e) presenta los cambios de  $Max N_r$  con el índice de Lévy  $\mu$ . Esta curva crece con  $\mu$ , es decir, conforme los individuos se vuelven menos superdifusivos, el enjambre alcanza una mayor cohesión, lo cual hace sentido, pues a un enjambre con individuos poco difusivos le cuesta menos trabajo cohesionarse pues la exploración a pasos largos es mucho menos

frecuente que la exploración local.

Por último, la Fig. 4.9 (f) muestra el tiempo típico medio que tarda el enjambre en alcanzar su máxima cohesión ( $\tau$ ) como función de  $\mu$ . Aquí  $\tau$  decrece conforme  $\mu$  crece hasta alcanzar un valor asintótico en  $\mu \approx 1.5$ . Esto confirma el resultado expuesto anteriormente; a un enjambre poco difusivo le tomará mucho menos tiempo alcanzar su máxima cohesión pero esto no implica que dicho agrupamiento se haga alrededor del mejor sitio de recurso.

Interpretando estos resultados a la vez, podemos concluir que aún y cuando el enjambre de primeros vecinos converge mucho más rápido que el de VL, y aún cuando el decaimiento de memoria mejora la búsqueda y explotación en ambos, el enjambre de primeros vecinos no se localiza de manera tan efectiva en el mejor sitio de recurso. Una dinámica de VL parece ser la estrategia “óptima” de búsqueda que maximiza el éxito de forrajeo grupal.

Como conclusión general podemos decir que, con nuestros modelos simple y colectivo, hemos mostrado la importancia que tiene el uso de la memoria junto con el movimiento aleatorio para la búsqueda eficiente y exitosa de recursos. Nuestros modelos, relativamente sencillos, muestran un aprendizaje emergente, tanto en los individuos aislados como en el colectivo, gracias al uso de memoria y la comunicación. Mostramos que una dinámica de VL optimiza la búsqueda y explotación de los recursos y que, tal como se ha mostrado en trabajos previos [13–15], esta dinámica resulta óptima no sólo a nivel individual sino también grupal. Hemos probado también la importancia que tiene que los individuos tengan la capacidad de olvidar, ya que esto les permite reforzar sitios visitados en el pasado cercano y actualizar de manera constante la información relevante que les permita localizarse en los mejores sitios de recurso. Podemos decir, hasta aquí, que la capacidad de recordar sitios previamente visitados y volver a ellos aún y cuando éstos se encuentren lejos de la posición actual del caminante, es clave para que el individuo modifique sus trayectorias (comportamiento) al interactuar con su medio y sea capaz de reconocer sitios de los cuales puede sacar provecho.

Con el fin de ver si estos simples ingredientes son suficientes para reproducir y predecir las trayectorias de animales reales en el siguiente capítulo mostramos un conjunto de modelos bayesianos jerárquicos para aproximar las trayectorias de 21 Ciervos Canadienses.

# Capítulo 5

## Modelos Bayesianos Jerárquicos para Ciervos Canadienses

En este capítulo, presentamos los resultados obtenidos al ajustar cuatro modelos bayesianos distintos a los datos de movimiento de 21 ciervos canadienses. Recapitulamos todos los conceptos y definiciones sobre la teoría de modelos bayesianos que se harán necesarios para entender a plenitud la sección de resultados de este capítulo.

Antes que nada, creemos prudente puntualizar la notación que usaremos de aquí en adelante. Introducida por primera vez en el artículo seminal de Gelfand y Smith [131], los corchetes se han vuelto una de las notaciones preferidas para los ecólogos y estadísticos que utilizan métodos bayesianos, ya que permiten escribir los modelos complejos multidimensionales de una forma compacta. Usaremos la notación  $[z]$  para referirnos a la probabilidad de una variable aleatoria  $z$  si  $z$  es discreta y a la densidad de probabilidad de  $z$  si  $z$  es continua. Así,  $[z]$  denota “ $z$  se distribuye como”. No incluiremos argumentos adicionales dentro de los corchetes cuando nos refiramos a distribuciones generales. Si estamos escribiendo sobre distribuciones específicas y nos queremos referir a los parámetros de ésta, usaremos  $[z|\alpha, \beta]$  para denotar la probabilidad o la densidad de probabilidad condicionada a  $\alpha$  y  $\beta$ . También nombraremos a menudo distribuciones específicas; por ejemplo, escribiremos  $\text{gamma}(z|\alpha, \beta)$  para denotar que la variable  $z$  sigue una distribución gamma con parámetros  $\alpha$  y  $\beta$ . Usaremos  $z \sim \text{gamma}(\alpha, \beta)$  para lo mismo. Y cuando nos refiramos específicamente a la probabilidad de  $z$  (excluyendo la densidad de probabilidad) usaremos la notación  $\text{Pr}(z)$  [95].

Empezaremos este capítulo con una sección completa para la función likelihood. Si bien es posible entender el análisis bayesiano sin entrar en detalle sobre esta función, nos parece importante agregar una sección para ella por la siguiente razón: La función likelihood o función de verosimilitud es el enlace fundamental entre los modelos y los datos dentro del esquema bayesiano [95].

## 5.1. Función de verosimilitud

Como ya mencionamos antes, podemos definir las distribuciones de probabilidad en términos de la variable aleatoria  $z$  y un conjunto de parámetros. Esta amplia definición tiene sentido ya que el enfoque bayesiano ve todas aquellas cosas no observadas como variables aleatorias, lo cual requiere un esquema igualmente amplio para tratar muchos tipos de cantidades [95].

Para explicar la función de verosimilitud (likelihood) empezaremos por acotar los argumentos en las distribuciones de probabilidad para incluir los datos puntuales  $y$  y los parámetros  $\theta$  en un modelo ecológico. Cuando los parámetros son valores fijos y los datos son variables aleatorias, se puede calcular la probabilidad de una observación (para datos discretos) o la densidad de probabilidad de una observación (datos continuos) utilizando una distribución que provenga de algún modelo con parámetros  $\theta$ . Pero el caso usual en la investigación de ecología cuantitativa es justo lo contrario; los parámetros son desconocidos y son los datos los que están fijos. Es decir, se tiene un conjunto de observaciones y se quiere conocer la información que esos datos nos pueden dar sobre los parámetros de un modelo dado que tiene como objetivo describir algún fenómeno ecológico. Se necesita pues una forma de evaluar la evidencia en los datos fijos para los valores arbitrarios de los parámetros, esto se hace por medio de la función de likelihood definida como:

$$L(\theta|y) = [y|\theta] \quad (5.1)$$

La ecuación anterior dice simplemente que el likelihood del parámetro  $\theta$  dada la observación  $y$ , es igual a la probabilidad (o la densidad de probabilidad) de la observación condicionada al parámetro. Para  $n$  observaciones independientes,

$$L(\theta|\mathbf{y}) = \prod_{i=1}^n [y_i|\theta] \quad (5.2)$$

$$\log(L(\theta|\mathbf{y})) = \sum_{i=1}^n \log[y_i|\theta] \quad (5.3)$$

La Ec.(5.1) define la llamada función likelihood. Dentro del enfoque bayesiano se refieren a esta misma ecuación como el modelo de datos o la función de verosimilitud [95].

La principal diferencia entre una distribución de probabilidad y una función likelihood es que el parámetro se encuentra fijo y los datos son variables aleatorias en una distribución de probabilidad, mientras que en la función likelihood, los datos se encuentran fijos y los parámetros son variables. La relación entre las funciones likelihood y las distribuciones de probabilidad pueden verse de manera más sencilla graficándolas (Fig.5.1). Si mantenemos el parámetro  $\theta$  constante y graficamos la función densidad de probabilidad  $[y|\theta]$  como función de distintos valores de  $y$  una variable continua, el área bajo la curva es igual a uno (Fig.5.1-A). Sin embargo, si fijamos  $y$  en un valor constante y graficamos la misma función de densidad de probabilidad pero ahora como función de  $\theta$  (Fig.5.1-B) obtenemos el *perfil del likelihood* [79], donde el área bajo la curva *no es* igual a uno. Sólo un punto de estas curvas se comparte,

mostrando que la probabilidad y el likelihood son el mismo sólo cuando el parámetro  $\theta$  es tratado como una cantidad fija [95]. Un comportamiento análogo puede verse cuando los datos se toman como variables aleatorias discretas (Fig.5.1-C-D).

Las unidades del eje  $y$  para el perfil del likelihood son arbitrarios y pueden escalarse por cualquier cantidad [132, 133]. De manera frecuente, el perfil del likelihood se escala de tal suerte que el pico de la curva sea igual a uno, lo cual se logra dividiendo todos los likelihoods entre el likelihood en el que el perfil alcanza su máximo (el likelihood máximo). Esto da un escalamiento conveniente, pero no cambia la relación entre el likelihood y la probabilidad [95]. Las gráficas mostradas en la Fig.5.1 ilustran una importante pero sutil diferencia entre las funciones likelihood y las distribuciones de probabilidad. Decir que  $\theta$  no es un valor fijo, nos permite calcular  $L(\theta|y)$  variando el valor de dicho parámetro, pero esto no significa que  $\theta$  deba ser tratado como una variable aleatoria en el esquema de la función likelihood. No es una variable aleatoria pues las variables aleatorias están definidas como cantidades regidas por distribuciones de probabilidad, y las funciones likelihood no definen la probabilidad o la densidad de probabilidad de  $\theta$  [95].

Una vez definida la función likelihood (o simplemente likelihood) y entendida la diferencia que guarda con la probabilidad o la densidad de probabilidad, en la siguiente sección expondremos los principios básicos de la inferencia bayesiana. El propósito es usar las reglas de probabilidad básicas (regla condicional de la probabilidad y la ley de probabilidad total) para mostrar cómo es que el Teorema de Bayes se aplica.

## 5.2. Modelos Bayesianos

El mundo puede dividirse en aquéllas cantidades que son observadas y aquéllas que no lo son. Las cantidades no observadas incluyen parámetros en los modelos, estados latentes predichos por los modelos, datos faltantes, o datos que aún no hemos observado (varios ejemplos de esto se mostraron ya en la sección 2.3 de este mismo escrito). Nos gustaría pues aprender de todas estas cantidades no observadas utilizando observaciones. El esquema bayesiano utilizado para alcanzar ese conocimiento es aplicado del mismo modo sin importar las especificaciones del problema de estudio o la naturaleza de las cantidades no observadas [95].

La característica principal del análisis bayesiano, y la que lo diferencia de todos los otros tipos de análisis estadísticos, es que éste trata a todas las cantidades no observadas como variables aleatorias. Ya que el comportamiento de las variables aleatorias se encuentra directamente definido por las distribuciones de probabilidad, se sigue que las cantidades no observadas pueden ser caracterizadas por dichas distribuciones de probabilidad. El análisis bayesiano usa las reglas de la probabilidad para descubrir las características de las distribuciones de probabilidad de las cantidades no observadas [95].

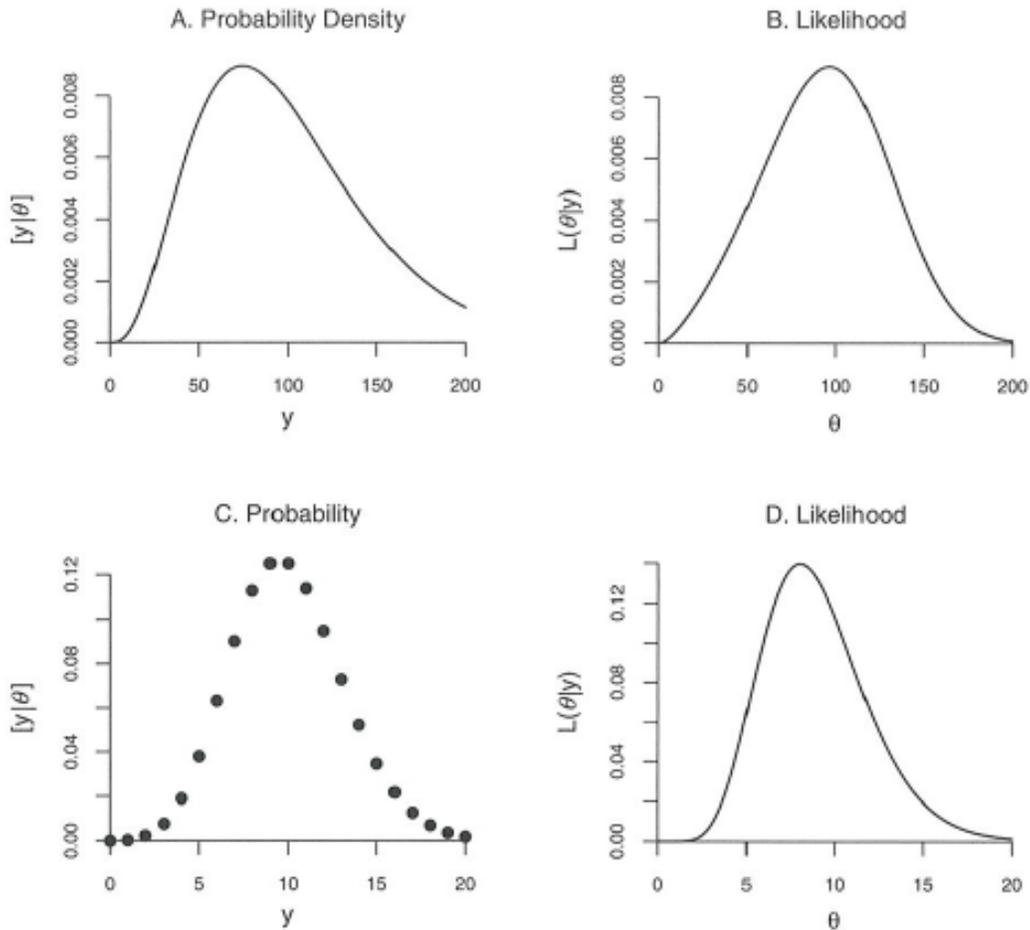


Figura 5.1: Relaciones entre las distribuciones de probabilidad para datos continuos (**A,B**) y discretos (**C,D**). Se asume que se tiene datos continuos estrictamente positivos con media fija en  $\theta = 100$  y varianza  $\sigma^2 = 2500$ . **A:** Función de densidad de probabilidad gamma graficada sobre  $y = 0, \dots, 200$  y condicionada a los parámetros fijos  $\alpha = \theta^2/\sigma^2$  y  $\beta = \theta/\sigma^2$ . La integral de esta función tomada de cero a  $\infty$  es igual a uno. Si asumimos que tenemos un dato fijo  $y = 75$  y una varianza fija en  $\sigma^2 = 2500$ , **B:** Función de densidad de probabilidad gamma sobre  $\theta = 0, \dots, 200$ . Esta curva es el perfil del likelihood y no integra uno. Un sólo punto se comparte entre las dos curvas,  $[y|\theta, \sigma^2]$ , donde  $y = 75, \theta = 100$  y  $\sigma^2 = 2500$ , ilustrando que la probabilidad y el likelihood son iguales sólo cuando los parámetros son tratados como valores fijos. **C:** Se asume una media fija  $\theta = 10$  para una función de probabilidad Poisson de los datos  $y = 1, 2, \dots, 20$ . La suma de las probabilidades es uno. **D:** El perfil del likelihood graficado como una función de probabilidad Poisson, donde la media  $\theta$  varía de cero a veinte,  $y = 8$  es fijo. La integral sobre la curva no es uno. Figura tomada de [95].

## El teorema de Bayes

El problema básico en ecología, y muchas otras áreas de la ciencia, es entender procesos que no podemos observar basándonos en cantidades que sí podemos observar. Representamos estos procesos no observados como modelos compuestos por parámetros y estados latentes que denotaremos por  $\theta$ . Hacemos observaciones  $y$  para aprender sobre  $\theta$ . Antes de que los datos sean observados, los trataremos como variables aleatorias. La probabilidad observar los datos condicionados a  $\theta$  está dada por la distribución de probabilidad  $[y|\theta]$ . Ya que  $\theta$  es también una variable aleatoria, está definida por su distribución  $[\theta]$ . Nos gustaría entonces descubrir la distribución de probabilidad de la cantidad no observada  $\theta$  condicionada a los datos observados  $[\theta|y]$ . Utilizando las reglas básicas de la probabilidad condicional para dos variables aleatorias, tenemos:

$$[\theta|y] = \frac{[\theta, y]}{[y]} \quad (5.4)$$

$$[y|\theta] = \frac{[\theta, y]}{[\theta]} \quad (5.5)$$

Resolviendo la Ec.(5.5) tenemos que:

$$[\theta, y] = [y|\theta][\theta] \quad (5.6)$$

y sustituyendo (5.6) en (5.4) tenemos:

$$[\theta|y] = \frac{[y|\theta][\theta]}{[y]} \quad (5.7)$$

Ya que  $y$  es condicional a  $\theta$ , la ley de la probabilidad total para parámetros discretamente valuados muestra que:

$$[y] = \sum_{\theta} [y|\theta][\theta] \quad (5.8)$$

donde la suma se hace sobre todos los posibles valores de  $\theta$ . Para parámetros que son continuos se tiene que:

$$[y] = \int [y|\theta][\theta]d\theta \quad (5.9)$$

Sustituyendo (5.9) en (5.7), obtenemos el teorema de Bayes para parámetros continuos,

$$[\theta|y] = \frac{[y|\theta][\theta]}{\int [y|\theta][\theta]d\theta} \quad (5.10)$$

Lo análogo para parámetros discretos se obtiene al sustituir (5.8) en (5.7). Dentro de la jerga utilizada en el enfoque bayesiano, la Ec.(5.10) es también conocida como la distribución posterior.

El teorema de Bayes nos da las bases para estimar la probabilidad de distribución de las cantidades no observadas  $\theta$  por medio de los datos (variables observadas)  $y$  [95]. Entender la inferencia bayesiana y por qué ésta funciona, requiere del entendimiento de cada una de sus componentes, los cuales explicaremos brevemente a continuación para el caso de parámetros continuos.

El *likelihood*  $[y|\theta]$  juega un rol clave en el análisis bayesiano haciendo un puente entre los datos no observados  $\theta$  y los observados  $y$ . Esto nos permite responder la pregunta central de la ciencia: ¿cuál es la probabilidad de que observemos los datos si nuestro modelo retrata de manera adecuada los procesos que dieron lugar a los datos? [95]. En la sección anterior expusimos de manera breve la definición e interpretación de la función likelihood.

La *distribución previa* de las cantidades no observadas  $[\theta]$ , representa nuestro conocimiento sobre  $\theta$  antes de que los datos sean colectados. La distribución previa puede ser informativa, reflejando el conocimiento obtenido en investigaciones anteriores, o puede ser “no informativa”, reflejando una falta de información sobre  $\theta$  antes de recolectar los datos que ahora se poseen y sobre los cuales se hará el análisis [95]. Es importante mencionar que la elección de las distribuciones previas es una decisión que puede afectar en la inferencia y hay que tener en cuenta dos cosas importantes [95]:

1. No existe tal cosa como *previas no informativas*, pero ciertas previas influyen en la distribución posterior más que otras.
2. Las previas informativas, cuando están debidamente justificadas, pueden ser bastante útiles en el enfoque bayesiano.

El objetivo del análisis bayesiano es el de proveer información que pueda resultar útil para análisis subsecuentes; la posterior obtenida en un estudio, podría convertirse en la previa de un estudio futuro. Por lo tanto, las previas “no informativas” son provisionales [74]. Siempre debemos preferir el uso de previas informativas bien construidas pero cabe reconocer que esto no siempre es posible.

El producto del likelihood y la previa es la distribución conjunta (ver Ec.(5.6)). En la sección previa mencionamos que el producto  $[y|\theta][\theta]$  no define una distribución de probabilidad para  $\theta$  ya que el área bajo la curva de  $[y|\theta][\theta]$  respecto a  $[\theta]$  no es igual a uno. Por otro lado, la distribución marginal de los datos  $[y]$  (definida en la Ec.(5.9)) es el área bajo curva de la distribución conjunta. Así pues, dividiendo cada punto de la distribución conjunta  $[y|\theta][\theta]$  por  $\int [y|\theta][\theta]d\theta$  normalizamos la curva con respecto a  $\theta$ , obteniendo la *distribución posterior*  $[\theta|y]$ . La distribución posterior es la función de densidad de probabilidad real que reúne todos los requisitos de estas funciones, incluyendo que  $\int [\theta|y]d\theta = 1$  [95, 99]. Al dividir la distribución conjunta por  $\int [y|\theta][\theta]d\theta$  aseguramos que la distribución posterior integre uno, por lo cual  $[y]$  será definida como la constante de normalización.

Antes de que los datos sean recolectados,  $y$  es una variable aleatoria, y la cantidad  $\int [y|\theta][\theta]d\theta$  es una distribución marginal, la cual también es a veces llamada la distribución predictiva previa (PPD por sus siglas en inglés) y nos dice lo que conocemos de los datos

antes de que éstos sean recolectados. Sin embargo, una vez que los datos han sido obtenidos,  $\int [y|\theta][\theta]d\theta$  pasa a ser un valor conocido. Esto significa que:

$$[\theta|y] \propto [\theta, y] \quad (5.11)$$

$$\propto [y|\theta][\theta] \quad (5.12)$$

Se puede usar la Ec.(5.12) para aprender acerca de la distribución posterior a partir de la distribución conjunta cuando no podemos calcular directamente  $[y]$ , como es comúnmente en el uso de esta herramienta estadística. La Ec.(5.12) define un modelo bayesiano simple, pues representa la distribución conjunta de las cantidades no observadas y observadas como producto del likelihood y las distribuciones previas [95]. Se puede hacer el mismo desarrollo para parámetros discretos, sustituyendo las integrales por sumas.

Como mencionamos brevemente en capítulos anteriores, el análisis bayesiano muchas veces se enfoca en el nivel individual. Sin embargo, muchos estudios en ecología están relacionados con el análisis a nivel poblacional, es decir, para varios individuos. Por ejemplo, en un conjunto de individuos observados ¿hay evidencia de respuestas de comportamiento consistentes a las variables ambientales? Para obtener un análisis a nivel poblacional, se hace necesario utilizar un modelo jerárquico con efectos aleatorios para los individuos que se agrupan a nivel de la población [71]. En la siguiente sección definiremos los modelos bayesianos jerárquicos y su potencial uso en la ecología, que es nuestra área de interés.

### 5.3. Modelos Jerárquicos

Como hemos visto en la sección anterior, un modelo es bayesiano si las cantidades no observadas las podemos entender como variables aleatorias cuyas distribuciones de probabilidad son estimadas a partir de las observaciones y el conocimiento previo. Así, un modelo bayesiano es simple si representa la probabilidad conjunta de esas variables aleatorias como producto del likelihood y las distribuciones previas. Por ejemplo,

$$[ \overbrace{\theta_1, \theta_2, z}^{\text{no observadas}} \mid \underbrace{\mathbf{y}}_{\text{observada}} ] \propto \underbrace{[\theta_1, \theta_2, z, \mathbf{y}]}_{\text{conjunta}} \quad (5.13)$$

$$\propto \underbrace{[\mathbf{y}|\theta_1, \theta_2, z]}_{\text{likelihood}} \underbrace{[\theta_1][\theta_2][z]}_{\text{previas}} \quad (5.14)$$

es un modelo bayesiano simple que toma en cuenta las cantidades no observadas  $\theta_1$ ,  $\theta_2$  y  $z$ , y las cantidades observadas  $\mathbf{y}$  [95]. Este modelo no es jerárquico ya que no existen condiciones más allá de la dependencia de los datos en las cantidades no observadas. Esto significa que cada una de los valores que aparecen en lado derecho de la condicional del likelihood se les asigna una distribución previa. Como ya vimos antes, la distribución posterior es proporcional a la distribución conjunta porque se ha omitido el denominador del teorema de Bayes,

la distribución marginal de los datos ( $\int \int [\mathbf{y}|\theta_1, \theta_2, z][\theta_1][\theta_2][z]d\theta_1d\theta_2dz$ ), que es un valor conocido luego de que se han observado los datos.

Existen varios enfoques en el área de modelos jerárquicos bayesianos, la más utilizada es la jerarquización por hiper-parámetros. Decimos pues que un modelo bayesiano es jerárquico por hiper-parámetros siempre que usemos reglas de probabilidad para factorizar la distribución conjunta como un producto de distribuciones condicionales [95]. Es decir, siempre que alguna de las variables no observadas a nivel individual esté condicionada a otra variable no observada a nivel poblacional cuya previa es, justamente, una distribución común para todos los individuos. Por ejemplo,

$$[\theta_1, \theta_2, z|\mathbf{y}] \propto [\theta_1, \theta_2, z, \mathbf{y}] \quad (5.15)$$

$$\propto [\mathbf{y}|\theta_1, z][z|\theta_2][\theta_1][\theta_2] \quad (5.16)$$

es jerárquico ya que los datos sólo están condicionados a las variables no observables  $\theta_1$  y  $z$  (que son cantidades no observables a nivel individual), pero  $z$  depende a su vez de otra variable no observada a nivel poblacional definida por  $\theta_2$  que es independiente de  $\theta_1$ . Cabe notar que la cantidad no observada  $z$  no tiene una distribución previa ya que está condicionada a una variable que sí tiene una distribución previa,  $\theta_2$ , esta previa es común para todos los individuos. Esta factorización de la distribución conjunta en productos de distribuciones condicionales no es arbitraria sino que está basada en el conocimiento previo del proceso ecológico que desea modelarse [95].

Una vez definido de manera más o menos intuitiva lo que es un modelo jerárquico bayesiano, a continuación presentaremos y desarrollaremos el ejemplo clásico de un modelo jerárquico, el cual que nos ayudará en secciones posteriores para definir las herramientas numéricas que utilizamos para obtener los resultados de este capítulo.

Asumamos que observamos  $y_{ij}$  datos continuos con  $i$  réplicas ( $i = 1, \dots, n_j$ ) para el individuo  $j$  ( $j = 1, \dots, J$ ). Un ejemplo de esto podría ser una medida fisiológica como la temperatura corporal recolectada en distintos  $n_j$  tiempos, que se suponen independientes, para  $J$  organismos individuales escogidos aleatoriamente de una población lo suficientemente grande. Si se está interesado en una inferencia a nivel poblacional sobre estos datos, donde el nivel poblacional se refiere a la temperatura media  $\mu$  (por ejemplo), no se hace prudente tratar las medidas a nivel individual  $y_{ij}$  como muestras unitarias, ya que existe un potencial muestreo desbalanceado puesto que  $n_j$  no es necesariamente igual para todos los individuos  $J$ . Así, si ajustamos el modelo con un likelihood basado solamente en el hecho de que  $y_{ij} \sim N(\mu, \sigma^2)$  (con  $N(\mu, \sigma^2)$  la distribución normal de media  $\mu$  y varianza  $\sigma^2$ ), podemos obtener una inferencia errónea o engañosa para  $\mu$  que es, de primera instancia, una medida bastante precisa y que no acepta rangos de variabilidad amplios.

Los modelos jerárquicos están diseñados para explicar estas llamadas medidas repetidas en cada individuo de una manera que aborda explícitamente la incertidumbre de la medición, la pseudo-replicación y diferentes cantidades de incertidumbre a nivel individual debido a los

diferentes tamaños de muestra o a las propias variaciones intrínsecas que puede haber entre individuos de una misma población. Los modelos jerárquicos proporcionan un mecanismo formal para incorporar efectos aleatorios porque, en este caso, tratan los parámetros de nivel individual como dicho efecto aleatorio. Para formular un modelo jerárquico para esta situación, permitimos que los datos originales surjan de una distribución con parámetros de nivel individual y permitamos que esos parámetros, a su vez, surjan de un proceso a nivel de población. Especificamos el modelo jerárquico como:

$$y_{ij} \sim N(\mu_j, \sigma_j^2) \quad (5.17)$$

$$\mu_j \sim N(\mu, \sigma^2) \quad (5.18)$$

$$\sigma_j^2 \sim \text{IG}(q, r) \quad (5.19)$$

$$\mu \sim N(\mu_0, \sigma_0^2) \quad (5.20)$$

$$\sigma^2 \sim \text{IG}(q, r) \quad (5.21)$$

para  $i = 1, \dots, n_j$ ,  $n_j$  el número de observaciones del individuo  $j$ ,  $\text{IG}(q, r)$  la distribución gamma inversa de parámetros  $q, r$ , donde  $\mu_j$  representa los efectos aleatorias a nivel individual (para  $j = 1, \dots, J$ ) y  $\sigma_j^2$  retrata la variación, también, a nivel individual (debida quizá a errores de medición o a la variabilidad natural entre individuos, i.e., a las diferencias intrínsecas de comportamiento entre individuos). Los parámetros a nivel poblacional  $\mu$  y  $\sigma^2$  son usualmente de mayor interés y cada uno tiene previas definidas por distribuciones normales y gamma inversa, respectivamente. En el esquema de modelos jerárquicos se hace referencia a los parámetros latentes a nivel individual  $\mu_j$  (o  $\sigma_j$ ) para  $j = 1, \dots, J$  como los “procesos”. Así, los datos surgen de la condicionalidad de un modelo que depende de los procesos latentes, los cuales a su vez vienen de una distribución que depende de un conjunto de parámetros [73].

La distribución posterior conjunta para el modelo jerárquico definido en Ecs.(5.17)-(5.21) es:

$$[\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \mu, \sigma^2 | \{y_{ij}, \forall i, j\}] \propto \left( \prod_{j=1}^J \left( \prod_{i=1}^{n_j} [y_{ij} | \mu_j, \sigma_j^2] \right) [\mu_j | \mu, \sigma^2] [\sigma_j^2] \right) [\mu] [\sigma^2] \quad (5.22)$$

donde  $\boldsymbol{\mu} \equiv (u_1, \dots, \mu_J)'$ ,  $\boldsymbol{\sigma}^2 \equiv (\sigma_1^2, \dots, \sigma_J^2)$ , y  $\{y_{ij}, \forall i, j\}$  corresponde las entradas del conjunto de datos  $y_{ij}$  para  $i = 1, \dots, n_j$  y  $j = 1, \dots, J$  [73].

Los modelos jerárquicos son frecuentemente aplicados en ecología para lidiar con cuatro retos comunes:

1. Representar la variación entre individuos que surge, por ejemplo, de las diferencias genéticas, las distintas posiciones espaciales o bien de las experiencias diversas que cada uno de ellos ha adquirido.
2. Estudiar fenómenos que operan a más de una escala espacial o nivel de organización ecológico.

3. Estimar incertidumbres que pueden deberse a la modelación del proceso, como esas incertidumbres que vienen de las observaciones imperfectas del proceso.
4. Entender los cambios en los estados de los sistemas ecológicos que no pueden observarse de manera directa. Estos estados surgen de procesos “ocultos”.

Un ejemplo que engloba la mayoría de estas características fue ya expuesto de manera breve en la sección 2.3 de este mismo escrito.

Ahora bien, resolver analíticamente estos modelos bayesianos (tanto a nivel individual como jerárquico) es, generalmente, imposible. Es por eso que varias herramientas numéricas han sido desarrolladas para aproximar los ajustes que proporciona el análisis bayesiano al conjunto de datos que deseamos explicar. En la siguiente sección presentaremos de manera intuitiva y breve el principal algoritmo numérico que se utiliza para el análisis bayesiano, las Cadenas de Markov Monte Carlo (MCMC por sus siglas en inglés).

## 5.4. Monte Carlo de Cadenas de Markov

Como vimos en la sección 5.3 de esta tesis, dividir la distribución conjunta por la distribución marginal de los datos, normaliza la distribución conjunta lo que la lleva a cumplir todas las características para ser una función de densidad de probabilidad. Esta normalización es lo que permite al modelo bayesiano tratar todas las variables no observadas como variables aleatorias que vienen de una distribución de probabilidad [74, 95].

Integrar la distribución conjunta analíticamente es imposible excepto para los problemas más sencillos, se hace ya difícil para tres parámetros, no digamos para modelos que contengan hasta 30 de ellos. Es aquí donde la aproximación de Monte Carlo de Cadenas de Markov (MCMC por sus siglas en inglés) nos permiten, de manera numérica, encontrar las distribuciones posteriores marginales para cada uno de los valores no conocidos (parámetros), evitando así cualquier integración formal. La idea esencial del MCMC es que podemos aprender de los valores no conocidos haciendo muestreos aleatorios de las posteriores marginales. La acumulación de muchas de estas muestras forma el perfil de la distribución posterior de la misma manera en la que tomando muchas muestras de un conjunto de datos da lugar a un histograma. Estas múltiples muestras nos permiten calcular cantidades de interés como la media, la desviación estándar y los cuantiles. Más aún, podemos obtener de manera más o menos sencilla las posteriores de cualquier cantidad que se derive de los valores desconocidos [74, 95].

Ahora bien, leyendo lo anterior es esperado que surja la siguiente pregunta: ¿cómo podemos obtener muestras aleatorias de algo que no conocemos? Aquí cabe recordar que la distribución posterior de los parámetros no es del todo desconocida, está parcialmente definida por los datos que influyen en estos valores no conocidos por medio del producto de la función likelihood y las previas. Recordemos que el teorema de Bayes nos dice que para los datos en un vector  $\mathbf{y}$  y para cualquier número de parámetros desconocidos y cantidades latentes en el vector  $\boldsymbol{\theta}$ ,

$$\underbrace{[\boldsymbol{\theta}|\mathbf{y}]}_{\text{posterior}} \propto \underbrace{[\mathbf{y}|\boldsymbol{\theta}]}_{\text{likelihood}} \underbrace{[\boldsymbol{\theta}]}_{\text{previa}}. \quad (5.23)$$

El problema que el MCMC resuelve es el de encontrar la constante de proporcionalidad (normalización) específica que lleve a la distribución posterior a ser una función de densidad de probabilidad que integre uno [74].

Antes de explicar cómo es que esta normalización ocurre, sería útil recordar cierta terminología y ponerla en contexto. La palabra “cadena” en el MCMC describe una sucesión de valores acumulados de muestras aleatorias de la distribución posterior. Estas muestras se guardan en un vector  $\boldsymbol{\theta}$ , con elementos  $\theta^{(k)}$ , con  $k = 1, \dots, K$  donde  $K$  es el número total de muestras. El algoritmo es de Markov porque cada elemento  $\theta^{(k)}$  depende sólo del valor de la muestra inmediata anterior de la cadena (paso anterior)  $\theta^{(k-1)}$  y es independiente de los valores anteriores a ese. El algoritmo se llama Monte Carlo porque los elementos de la cadena se acumulan durante  $K$  pasos aleatorios de una distribución, algo similar a escoger número de una ruleta de casino. Por ahora pensemos en esta distribución como una manera de generar números aleatorios que están dentro del soporte del parámetro que estamos tratando de estimar [74, 95].

De manera heurística, el algoritmo MCMC funciona como sigue para el caso de un sólo parámetro (más adelante generalizaremos a múltiples parámetros): Dado el conjunto de datos  $\mathbf{y}$  que nos permitirán estimar  $\theta$ , definimos una previa para el parámetro, a saber,  $[\theta]$  que será una distribución numérica con argumentos que pueden ser o no informativos, dependiendo si tenemos o no conocimiento previo del comportamiento de dicho parámetro. Escogemos un valor que iniciará la cadena, el valor  $\theta$  al tiempo  $k = 1$  será  $\theta^{(1)}$ . Esta elección puede ser arbitraria, pero es preferible escoger un valor que responda a algún conocimiento previo del parámetro, si lo sabemos, si no escapa de nuestro conocimiento, bastará con tomar un valor aleatorio muestreado de la previa. Luego, escogemos un nuevo valor propuesto de  $\theta$ ,  $\theta^{(*)}$ , de una distribución propuesta apropiada para el soporte de  $\theta$  (dicha distribución propuesta puede depender del valor inicial  $\theta^{(1)}$ ). Por el momento sigamos pensando que esta distribución propuesta es arbitraria [74, 95].

La esencia del MCMC es: Se hace una sucesión de elección de valores de la distribución propuesta. Utilizamos la distribución conjunta para escoger valores a cada extracción de tal suerte que la frecuencia de los valores de  $\theta$  acumulados en la cadena sean proporcionales a su probabilidad en la distribución conjunta, es decir, usamos la distribución conjunta para decidir la frecuencia de los valores que se acumulan en la cadena. Un gran número de valores de  $\theta$  que son proporcionales a la conjunta, definen la posterior de la misma manera en la que un gran número de observaciones definen una distribución en un histograma [95].

Para el caso de múltiples parámetros  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)$  el MCMC funciona de manera similar pero extendida, es decir, cada uno de los  $m$  valores no conocidos tiene su propia cadena con sus propios valores iniciales. El MCMC itera sobre cada uno de los parámetros, tratándolos como si fueran los únicos valores no conocidos, así como en el caso de un sólo parámetro.

Así, a cada paso de la cadena escogemos un valor para cada parámetro asumiendo que los otros son conocidos, se repite el proceso de manera independiente para cada uno de ellos [95].

Mostraremos a continuación algunas especificaciones que hemos omitido en nuestra explicación heurística y que nos parecen importantes para terminar de entender cómo funciona el algoritmo MCMC.

## Muestreo de la distribución posterior

Hay dos formas de obtener muestras de la distribución posterior: la primera es el llamado *muestreo aceptar-rechazar* donde algunas propuestas se mantienen y otras se descartan, el segundo tipo es el llamado *muestreo de Gibbs* y en él todas las propuestas se mantienen ya que la distribución propuesta es particularmente bien escogida. En este trabajo abordaremos sólo el muestreo aceptar-rechazar pues es el que utilizaremos en las secciones posteriores de resultados. En [74, 134] el lector puede encontrar, si le es de interés, las ideas básicas relacionadas con el muestreo de Gibbs.

El muestreo aceptar-rechazar que se compone de los siguientes tres pasos [74, 95] en el caso de un sólo parámetro:

1. Se propone un nuevo valor  $\theta^{(*)}$  para el parámetro.
2. Se calcula la probabilidad de aceptar el valor propuesto utilizando una tasa de transición basada en la distribución conjunta evaluada en el valor propuesto  $\theta^{(*)}$  y el valor previo  $\theta^{(k-1)}$ .
3. Se acepta el valor propuesto ( $\theta^{(k)} = \theta^{(*)}$ ) con probabilidad calculada en el paso anterior. Si no se acepta el valor propuesto entonces  $\theta^{(k)} = \theta^{(k-1)}$

Cómo implementar estos pasos se describe en los siguientes apartados.

### La distribución propuesta

Como hemos dicho antes, la implementación del MCMC inicia con un vector de valores para el parámetro  $\theta$  que está vacío excepto por el primer elemento  $\theta^{(1)}$ , que se asigna de manera informada o aleatoria. Se llenan los elementos restantes del vector tomando muestras propuestas de una distribución, a la que se le llama, *la distribución propuesta*. La distribución propuesta es un concepto clave en el MCMC.

En el muestreo aceptar-rechazar, existen dos tipos de distribuciones propuesta: la independiente y la dependiente. El valor propuesto tomado de una distribución independiente no depende en ningún sentido del valor actual de la cadena. De forma inversa, los valores propuestos de una distribución dependiente, dependen del valor actual de la cadena. Aquí nos enfocaremos en las distribuciones propuesta dependientes pues son las más utilizadas en las implementaciones del MCMC [95]. Cabe mencionar que la distribución posterior resulta

“independiente” de la distribución propuesta pero la distribución propuesta sí puede tener un impacto directo en el tiempo de convergencia del MCMC [74, 95].

La idea de una distribución propuesta dependiente es simple. Asumamos que  $\theta^{(k)}$  es el valor actual de la cadena, y queremos generar una propuesta  $\theta^{(*)}$  para el siguiente valor de la cadena  $\theta^{(k+1)}$ . Se toma pues una muestra de

$$[\theta^{(*)}|\theta^{(k)}] \quad (5.24)$$

que dice simplemente que hacemos la muestra aleatoria de una distribución condicionada a  $\theta^{(k)}$  [74, 95]. Por ejemplo, podríamos hacer una propuesta tomada de una distribución normal con media  $\theta^{(k)}$  y varianza  $\sigma_{\text{tune}}^2$ ,

$$\theta^{(*)} \sim \text{Normal}(\theta^{(k)}, \sigma_{\text{tune}}^2). \quad (5.25)$$

La varianza  $\sigma_{\text{tune}}^2$  es el parámetro de afinación, escogida subjetivamente para fijar la anchura de la distribución propuesta. Este parámetro determina la frecuencia con la que las propuestas son aceptadas para la cadena, una varianza grande lleva a una mayor frecuencia de rechazo [95]. Si el parámetro de afinación no se define, decimos que el muestreo es no supervisado. En este ejemplo la distribución propuesta es simétrica y por lo tanto,

$$[\theta^{(*)}|\theta^{(k)}] = [\theta^{(k)}|\theta^{(*)}] \quad (5.26)$$

lo que significa que la densidad de probabilidad de la variable aleatoria  $\theta^{(*)}$  condicionada a  $\theta^{(k)}$  es igual a la densidad de probabilidad de la variable aleatoria  $\theta^{(k)}$  condicionada a  $\theta^{(*)}$ . Las distribuciones normal y uniformes son distribuciones simétricas. Sin embargo, si el soporte de  $\theta$  no incluye todo los número reales sino sólo el subconjunto de los no negativos, entonces una mejor elección para las distribuciones propuesta serían una lognormal o una gamma:

$$\theta^{(*)} \sim \text{lognormal}(\log(\theta^{(k)}), \sigma_{\text{tune}}^2) \quad (5.27)$$

$$\text{or} \quad (5.28)$$

$$\theta^{(*)} \sim \text{gamma}\left(\frac{(\theta^{(k)})^2}{\sigma_{\text{tune}}^2}, \frac{\theta^{(k)}}{\sigma_{\text{tune}}^2}\right) \quad (5.29)$$

pero en ambos casos, las distribuciones propuesta son no simétricas [95],

$$[\theta^{(*)}|\theta^{(k)}] \neq [\theta^{(k)}|\theta^{(*)}] \quad (5.30)$$

### Actualización de Metropolis

La idea general es que nos gustaría pues incluir valores a la cadena cuya frecuencia sea proporcional a su probabilidad en la distribución posterior. Para hacer esto, actualizamos la cadena utilizando el muestreo de Metropolis, propuesto por Metropolis *et al.* en 1953 [135], como sigue. La densidad de probabilidad del valor actual de la cadena condicionado a los datos es [135]:

$$[\theta^{(k)}|\mathbf{y}] = \frac{[\mathbf{y}|\theta^{(k)}][\theta^{(k)}]}{[\mathbf{y}]} \quad (5.31)$$

la cual, por supuesto, es la posterior de  $\theta^{(k)}$ . De forma análoga, la densidad de probabilidad de la propuesta  $\theta^{(*)}$  condicionada a los datos es [135]:

$$[\theta^{(*)}|\mathbf{y}] = \frac{[\mathbf{y}|\theta^{(*)}][\theta^{(*)}]}{[\mathbf{y}]} \quad (5.32)$$

que es a distribución posterior de  $\theta^{(*)}$ . Resolver esta ecuación es complicado pues se necesita encontrar la integral  $[\mathbf{y}] = \int [\mathbf{y}|\theta][\theta]d\theta$ , lo cual generalmente es imposible ya que el espacio de los valores posibles de  $\theta$  es demasiado grande. El MCMC permite muestrear un subconjunto de este espacio, concentrándose en esa región (relativamente pequeña) donde la posterior es más alcanza sus valores más probables. Así, nos resta sólo preocuparnos por encontrar el siguiente valor de  $\theta$  de la cadena que nos permita converger hacia la mejor posterior posible [95]. El muestreo de Metropolis [135] encuentra dicho valor utilizando la tasa de transición dada por:

$$R = \frac{[\mathbf{y}|\theta^{(*)}][\theta^{(*)}]/[\mathbf{y}]}{[\mathbf{y}|\theta^{(k)}][\theta^{(k)}]/[\mathbf{y}]} \quad (5.33)$$

$$= \frac{[\mathbf{y}|\theta^{(*)}][\theta^{(*)}]}{[\mathbf{y}|\theta^{(k)}][\theta^{(k)}]} \quad (5.34)$$

que  $[\mathbf{y}]$  se cancele nos quita la necesidad de integrar. En el algoritmo de Metropolis, se acepta el valor propuesto siempre que  $R \geq 1$ . Cuando  $R < 1$  aceptamos el valor propuesto con probabilidad  $R$  y lo rechazamos con probabilidad  $1 - R$ , lo cual de manera algorítmica se hace como sigue:

1. Calcular  $R$  como

$$R = \min \left( 1, \frac{[\mathbf{y}|\theta^{(*)}][\theta^{(*)}]}{[\mathbf{y}|\theta^{(k)}][\theta^{(k)}]} \right) \quad (5.35)$$

2. Aceptar  $\theta^{(*)}$  como el siguiente valor de la cadena con probabilidad  $R$  y rechazarlo con probabilidad  $1 - R$

Si ejecutamos estos pasos un número de veces *suficiente*, entonces la cadena converge, lo que significa que la frecuencia de un valor cualquiera en la cadena es proporcional a su valor en la distribución posterior. La convergencia ocurre cuando la adición de nuevas propuestas a la cadena no cambia la representación relativa de los valores en la cadena. Esto está relacionado con el hecho de que la probabilidad de transición de  $\theta^{(k)}$  a  $\theta^{(*)}$  definida por  $R$  satisface el balance detallado, que es una propiedad de una distribución de equilibrio (la posterior) [135]. Para más detalles técnicos sobre la convergencia y eficiencia de este algoritmo se puede consultar [74, 95, 135].

### Actualización de Metropolis-Hastings

La actualización de la cadena utilizando el muestreo de Metropolis requiere que la distribución propuesta sea simétrica, por ejemplo, y las más utilizadas, son la distribución normal o la distribución uniforme. La distribución uniforme como distribución propuesta lleva generalmente a una convergencia mucho más lenta del algoritmo pues todos los valores son igualmente probables y la actualización de la cadena hacia la distribución posterior puede tomar mucho más tiempo [74, 95]. Por otro lado, la distribución normal no resulta apropiada para generar propuestas para parámetros o valores latentes cuyo soporte no es todos los números reales [74, 95].

El muestreo de Metropolis-Hastings, propuesto por Hastings en 1970 [136], relaja el requisito de una distribución propuesta simétrica, permitiéndonos utilizar cualquier distribución para los potenciales nuevos valores de la cadena, multiplicando  $R$  por el cociente  $[\theta^{(k)}|\theta^{(*)}]/[\theta^{(*)}|\theta^{(k)}]$  como sigue:

$$R = \frac{[\mathbf{y}|\theta^{(*)}][\theta^{(*)}]}{[\mathbf{y}|\theta^{(k)}][\theta^{(k)}]} \frac{[\theta^{(k)}|\theta^{(*)}]}{[\theta^{(*)}|\theta^{(k)}]} \quad (5.36)$$

Así, la única diferencia entre Metropolis y Metropolis-Hastings es la manera en la que se calcula  $R$  y el tipo de distribuciones propuesta que podemos utilizar. En Metropolis-Hastings la tasa de transición  $R$  satisface el caso más general en la condición del balance detallado [136].

Una vez entendido de manera más o menos general cómo es que el algoritmo MCMC funciona y el objetivo que tiene, éste puede aplicarse tanto a modelos a nivel individual como a nivel jerárquico. Como en este escrito estamos interesados en aplicar modelos bayesianos jerárquicos a datos observados, nos enfocamos en la siguiente sección en exponer y explicar un algoritmo de dos pasos propuesto en [73] que eficientiza el tiempo de cómputo para resolver modelos a nivel jerárquico y que hace mucho más sencillo el cálculo de las posteriores.

## 5.5. Algoritmo de dos pasos

Como hemos mencionado ya en varias ocasiones, los modelos jerárquicos son de gran utilidad y han liderado muchos de los avances científicos en diversas disciplinas. Sin embargo, con el aumento de la complejidad en los modelos, viene también un aumento en la carga computacional. Incluso el modelo sencillo que tomamos como ejemplo característico de los modelos jerárquicos en la sección 5.3 puede ser trivial de implementar [73] pero sigue requiriendo muchos más pasos analíticos y computacionales que muchos otros modelos, incluso, no jerárquicos (ver [73] para ver la implementación del modelo (5.17)-(5.21)).

Se pueden añadir más niveles en un modelo jerárquico, incluir modelos de datos más complicados que pueden no ser necesariamente independientes, y especificar modelos sofisticados con estructura espacial o temporal, en fin... hacerse tan complicado y complejo como uno lo desee. Entonces, incluso si el algoritmo MCMC puede implementarse a modelos más complejos, el resultado puede llevar a cadenas de Markov con combinaciones pobres e inestabilidades debido a las diversas interdependencias entre los parámetros del modelo y los

procesos [71, 73]. Además el tiempo de ejecución, incluso para los algoritmos ya implementados en diversos softwares, puede llegar a ser muy largo y casi imposible de optimizar.

Existen muchos enfoques para facilitar y eficientar el cálculo por medio del uso de algoritmos personalizados que son específicos para ciertos tipos de modelo. Dentro de toda esta variada gama, hay uno que es particularmente útil para ciertas clases de modelos jerárquicos y es ese el que veremos en esta sección. Esta estrategia de computación es el “método Lunn”, que lleva el nombre de David Lunn, pionero en algunos de los primeros esfuerzos para desarrollar formas de adaptar los modelos bayesianos a los datos mediante el uso de software automático [137, 138]. El método Lunn implica aproximar modelos jerárquicos a partir de muchos modelos no jerárquicos en dos etapas [139]. La primera etapa consiste en, por medio del MCMC, encontrar las posteriores respectivas al conjunto de modelos no jerárquicos de forma independiente para cada individuo, utilizando previas no informativas para los parámetros. La segunda etapa implica un algoritmo MCMC sustancialmente simplificado para encontrar las posteriores del modelo jerárquico completo utilizando la salida de la primera etapa [73].

El método de Lunn produce una inferencia exacta para los parámetros del modelo (en el mismo sentido que MCMC es exacto<sup>1</sup>) y la segunda etapa del algoritmo no se basa en los datos de manera directa, sólo indirectamente a través de la salida de la primera etapa [73]. En [71] se mostró cómo implementar el método de Lunn para modelos jerárquicos de movimiento de animales para hacer inferencia a nivel poblacional en parámetros o procesos latentes comunes. Revisamos este enfoque aquí, pero en el contexto del modelo jerárquico gaussiano de la sección 5.3, que es el mismo modelo jerárquico que implementamos para obtener los resultados que expondremos luego en este mismo capítulo.

El principal truco del método de Lunn está en la segunda etapa del algoritmo donde las muestras del MCMC obtenidas en la primera etapa para los parámetros son usadas como distribuciones propuesta. El uso de las muestras del MCMC del primer paso como distribuciones propuesta simplifica de manera dramática el cálculo de la actualización de Metropolis-Hastings (M-H) en la segunda etapa del algoritmo para cada uno de los parámetros dependientes de los datos a nivel jerárquico. Para ver cómo es que funciona este novedoso método, consideremos el modelo jerárquico gaussiano definido en (5.17)-(5.21). Primero, separaremos el modelo jerárquico para formar un modelo separado para el conjunto de datos  $\mathbf{y}_j$  que vienen de cada individuo  $j = 1, \dots, J$

$$y_{ij} \sim N(\mu_j, \sigma_j^2). \quad (5.37)$$

Cada uno de estos modelos separados se ajusta (se usa MCMC para encontrar las posteriores correspondientes) utilizando previas no informativas para los parámetros del modelo. En este caso se puede asumir que las previas son:  $[\mu_j] \equiv N(\mu, \sigma^2)$  y  $[\sigma_j^2] \equiv \text{IG}(q, r)$ , con  $\mu, \sigma^2, q$  y  $r$  dados. Ajustando cada uno de estos modelos a nivel individual obtenemos, de

---

<sup>1</sup>En este sentido, “exacto” se refiere a las propiedades del MCMC que garantizan que podemos caracterizar, de forma exacta, la distribución posterior con un tiempo de cómputo infinito (es decir, un tamaño de muestra infinito). Por supuesto, en la práctica nuestra inferencia final basada en la salida del MCMC es aproximada debido a la longitud finita de la cadena de Markov, pero podemos acercarla lo más que queramos al extender la longitud de la cadena.

forma separada e independiente, las distribuciones posteriores para cada uno de los  $\mu_j$  y  $\sigma_j^2$  [73],

$$[\mu_j, \sigma_j^2 | \{y_{ij}, \forall i, j\}] \propto \prod_{i=1}^{n_j} [y_{ij} | \mu_j, \sigma_j^2] [\mu_j] [\sigma_j^2], \quad (5.38)$$

con  $n_j$  las observaciones totales para el individuo  $j$ .

Entonces, el primer paso del algoritmo se trata pues de ajustar  $J$  modelos no jerárquicos por separado. Más aún, ya que en esta primera parte los modelos se tratan como independientes, pueden ajustarse todos en paralelo [73].

Para delinear la segunda parte del algoritmo cabe mencionar que el algoritmo MCMC para ajustar el modelo jerárquico completo nos obliga a muestrear cada distribución condicional completa (para los  $J$  individuos) para los parámetros de forma secuencial. De hecho, en [73] muestran que, particularmente, para la actualización de  $\mu_j$  en el algoritmo del modelo jerárquico completo se puede encontrar la distribución condicional para  $\mu_j$  de forma analítica y se puede usar una actualización de Gibbs [134] para muestrearla. Sin embargo, también es posible usar la actualización M-H para muestrear  $\mu_j$  basándose en los valores propuestos  $\mu_j^{(*)}$ . Centrándose en las actualizaciones M-H, Hooten *et al.* proponen en [73] valores propuestos  $\mu_j^{(*)}$  y  $\sigma_j^{2(*)}$  tomados de la distribución propuesta conjunta  $[\mu_j, \sigma_j^2]^*$  y luego se actualizan por bloques en el algoritmo MCMC. La tasa de transición M-H que resulta de este escenario propuesto es:

$$R = \frac{\prod_{i=1}^{n_j} [y_{ij} | \mu_j^{(*)}, \sigma_j^{2(*)}] [\mu_j^{(*)} | \mu^{(k)}, \sigma^{2(k)}] [\sigma_j^{2(*)}] [\mu_j^{(k-1)}, \sigma_j^{2(k-1)}]^*}{\prod_{i=1}^{n_j} [y_{ij} | \mu_j^{(k-1)}, \sigma_j^{2(k-1)}] [\mu_j^{(k-1)} | \mu^{(k)}, \sigma^{2(k)}] [\sigma_j^{2(k-1)}] [\mu_j^{(*)}, \sigma_j^{2(*)}]^*}. \quad (5.39)$$

Analicemos a detalle esta ecuación. Primero, y como hemos ya mencionado antes, esta tasa de transición nos dará las probabilidades de aceptación-rechazo para la actualización de los parámetros individuales  $\mu_j$  y  $\sigma_j^2$  de manera conjunta. Así pues,  $[\mu_j^{(*)}, \sigma_j^{2(*)}]^*$  no es más que la distribución conjunta propuesta evaluada en  $\mu_j^{(*)}$  y  $\sigma_j^{2(*)}$ . Luego, Hooten *et al.* hacen un supuesto importante y es que la previa de  $\mu_j$  depende de los valores actuales  $k$  de los hiper-parámetros  $\mu$  y  $\sigma$  que se toman directamente de sus respectivas previas no informativas dadas por la distribución normal y la gamma inversa definidas en (5.17)-(5.21) para  $\mu_0$ ,  $\sigma_0^2$ ,  $q$  y  $r$  definidos en [73]. Así,  $[\mu_j^{(*)} | \mu^{(k)}, \sigma^{2(k)}]$  es la previa de  $\mu_j$  condicionada a  $\mu^{(k)}$  y  $\sigma^{2(k)}$  evaluada en el valor propuesto  $\mu_j^{(*)}$  y, de forma análoga,  $[\sigma_j^{2(*)}]$  es la previa no informativa para  $\sigma_j^2$  evaluada en el valor propuesto  $\sigma_j^{2(*)}$ .

En lugar de utilizar distribuciones propuestas aleatorias o previas como propuestas, utilizamos la distribución conjunta posterior obtenida del primer paso del algoritmo (Ec.(5.38)) como la distribución propuesta  $[\mu_j, \sigma_j^2]^*$  para esta segunda fase. Así [73],

$$[\mu_j, \sigma_j^2]^* \propto \prod_{i=1}^{n_j} [y_{ij} | \mu_j, \sigma_j^2] [\mu_j] [\sigma_j^2]. \quad (5.40)$$

Sustituyendo (5.40) en (5.39) tenemos que la tasa de transición M-H queda como [71, 73]:

$$\begin{aligned} R &= \frac{\prod_{i=1}^{n_j} [y_{ij} | \mu_j^{(*)}, \sigma_j^{2(*)}] [\mu_j^{(*)} | \mu^{(k)}, \sigma^{2(k)}] [\sigma_j^{2(*)}]}{\prod_{i=1}^{n_j} [y_{ij} | \mu_j^{(k-1)}, \sigma_j^{2(k-1)}] [\mu_j^{(k-1)} | \mu^{(k)}, \sigma^{2(k)}] [\sigma_j^{2(k-1)}]} \frac{\prod_{i=1}^{n_j} [y_{ij} | \mu_j^{(k-1)}, \sigma_j^{2(k-1)}] [\mu_j^{(k-1)}] [\sigma_j^{2(k-1)}]}{\prod_{i=1}^{n_j} [y_{ij} | \mu_j^{(*)}, \sigma_j^{2(*)}] [\mu_j^{(*)}] [\sigma_j^{2(*)}]} \\ &= \frac{[\mu_j^{(*)} | \mu^{(k)}, \sigma^{2(k)}] [\mu_j^{(k-1)}]}{[\mu_j^{(k-1)} | \mu^{(k)}, \sigma^{2(k)}] [\mu_j^{(*)}]} \end{aligned} \quad (5.41)$$

La tasa de transición M-H de (5.41) es mucho más simple que la versión original para el modelo jerárquico completo definido en (5.39), esto se debe a que los componentes del likelihood y las previas para  $\sigma_j^2$  se cancelan <sup>2</sup>. De hecho, si las previas del primer paso para  $\mu_j$  son lo suficientemente difusas, también se cancelarán en el cálculo de  $R$  en el segundo paso, quedando la tasa de transición sólo en función de los modelos del proceso. Más aún, ya que la tasa resultante para el segundo paso del algoritmo para  $\mu_j$  no depende de  $\sigma_j^2$  ni tampoco ninguna de las restantes distribuciones condicionales completas para los parámetros restantes  $\mu$  o  $\sigma^2$ , se puede ignorar el muestreo de  $\sigma_j^2$  en la primera etapa del algoritmo.

Así, para implementar el método de Lunn para ajustar modelos jerárquicos como el definido en (5.17)-(5.21), se ajusta primero, por separado y de manera independiente, cada uno de los modelos a nivel individual para cada uno de los sujetos de los que se poseen datos. Luego se utiliza la salida del MCMC que resulta del ajuste individual como distribución propuesta en la segunda fase del algoritmo con actualizaciones M-H para  $\mu_j$  definida en la Ec.(5.41), las actualizaciones de los hiper-parámetros  $\mu$  y  $\sigma^2$  se hacen de manera exacta a partir de las previas y se encuentran expuestas en [73]. Esta jerarquización nos dará las posteriores que definen las variaciones de los parámetros individuales a nivel jerárquico y también las posteriores de los hiper-parámetros que definen el comportamiento a nivel poblacional. Para el caso en el que el modelo considere múltiples parámetros, cada uno de ellos puede ajustarse de manera independiente usando el algoritmo de dos pasos anteriormente expuesto.

En este escrito utilizaremos en las secciones posteriores de este capítulo este algoritmo de dos pasos para jerarquizar nuestros modelos de movimiento entre parches para los datos de 21 ciervos canadienses. Nuestros modelos poseen más de un parámetro, por lo que la segunda fase del algoritmo se implementará de manera independiente para cada uno de ellos obteniendo como resultado las variaciones (media y desviación) de dichos parámetros a nivel jerárquico para cada individuo y también para un individuo “promedio” que será representativo de toda la población estudiada. Más adelante daremos los detalles técnicos de los modelos y cómo se ajustaron éstos al conjunto de datos que poseemos.

---

<sup>2</sup>La cancelación sólo ocurre cuando no existe autocorrelación en la muestra del primer paso del algoritmo.

Ahora bien, ya que poseemos todas las herramientas heurísticas para aplicar e implementar modelos bayesianos, queda sólo responder a la pregunta: dentro de todos los posibles modelos que se pueden sugerir para describir y predecir un fenómeno, ¿cómo identificamos y escogemos al mejor de ellos? Para esto existen métodos cualitativos y cuantitativos que nos permiten identificar, dentro de un conjunto de modelos, el mejor de ellos. En las siguientes dos secciones presentaremos dos de estas herramientas (una cualitativa y otra cuantitativa) que son relativamente sencillas de implementar y bastante efectivas. Nos enfocamos en estas dos, pues son las que utilizaremos en la sección de resultados de este capítulo cinco.

## 5.6. Comprobación Predictiva Posterior

La Comprobación Predictiva Posterior (PPC por sus siglas en inglés) es una aproximación relativamente reciente [74, 95, 140] que ha sido desarrollada para comprobar si un modelo propuesto debe o no de ser desechado porque dicho modelo falla en proporcionar un buen ajuste para los datos que se deseaban explicar con él. Esta herramienta no sólo nos ayuda a verificar si un modelo resulta o no adecuado para reproducir los datos observados sino también nos ayuda, de manera cualitativa también, a concluir si un modelo propuesto nos ayuda o no a “predecir” datos no observados utilizando como referencia los datos ya observados. El PPC ha recibido una cantidad de atención razonable pero aún no está del todo bien entendido, por lo menos no en el contexto de la estimación del likelihood máximo, es por eso que se le considera una herramienta cualitativa de comprobación y comparación [140].

El PPC fue desarrollado originalmente para comprobar modelos bayesianos. Como ya hemos mencionado anteriormente en este mismo capítulo, los modelos bayesianos utilizan distribuciones previas para sus parámetros en conjunto con los datos observados para encontrar la distribución posterior que define el comportamiento de dichos parámetros de tal suerte que el modelo se ajuste lo mejor posible a los datos reales. Si los datos son relativamente no informativos del fenómeno que se desea modelar, entonces la posterior y la previa serán muy similares y no obtendremos demasiada información de nuestro modelo. Cuando las previas son relativamente no informativas, entonces la posterior para los parámetros del modelo podrían dar un ajuste pobre sin lograr reproducir, de manera confiable, los datos observados, i.e., sin ajustarse de manera satisfactoria a ellos. Idealmente nos gustaría que las posteriores obtenidas fueran lo suficientemente buenas para reproducir nuestros datos observados pero ¿cómo medir cuán buenas o malas resultan las posteriores? [140]. De acuerdo al PPC, el camino a seguir es: dada alguna característica fundamental (llamémosla *estadístico*), calculada de las observaciones, está deberá de ser comparada con el mismo estadístico calculado desde la distribución posterior. Si esta comparación falla en cumplir ciertos criterios preespecificados, entonces el modelo debe rechazarse [74, 140].

La aplicación del PPC se compone de dos pasos principales [74, 140]:

*Estimación:* Al realizar la estimación de los parámetros usando el algoritmo MCMC para ajustar nuestro modelo bayesiano a nivel individual, obtenemos una matriz de muestras  $M_j$  para cada individuo  $j = 1, \dots, J$  cuyas filas definen las variaciones (la posterior) para cada

uno de los parámetros a considerar, i.e., esta matriz  $M_j \in \mathbb{M}(p, s)$  donde  $p$  es el número total de parámetros estimados y  $s$  es la cantidad de muestras que arroja el MCMC. Por lo tanto la entrada  $M_j(l, k)$  define la muestra  $k$  del parámetro  $l$ . Más aún, cada una de las columnas representa una muestra de la posterior conjunta de los parámetros. Ahora bien, si la estimación se hace a nivel jerárquico, nuestro algoritmo de dos pasos nos regresará dos arreglos para cada uno de los parámetros  $h = 1, \dots, p$  del modelo; el vector  $\boldsymbol{\theta}_h$  de longitud  $s$  y la matriz  $Q_h \in \mathbb{M}(J, s)$ . Las entradas del vector  $\boldsymbol{\theta}_h$  nos dará las variaciones del parámetro  $h$  a nivel poblacional (esto podría interpretarse como las variaciones del parámetro  $h$  para un individuo “promedio” representativo de la población), y cada una de las filas de la matriz  $Q_h$  nos darán las variaciones del parámetro  $h$  para cada individuo  $j = 1 : J$  a nivel poblacional. Claramente esta estimación se hizo considerando los datos observados  $y_{ij}$  con  $i = 1, \dots, n_j$  y  $n_j$  el número de observaciones para el individuo  $j$ .

*Cálculo del estadístico:* Sea  $\Psi$  el estadístico con el cuál compararemos las distribuciones posteriores obtenidas de la estimación con los datos reales. Como resulta obvio, hay que aplicar este estadístico a cada vector  $y_{ij}$ ,  $\Psi(y_{ij})$  fijando  $j$ , i.e., el estadístico se aplicará de manera independiente a cada individuo. Ahora bien, para comparar las estimaciones con los valores observados, se hace necesario crear réplicas de  $y_{ij}$ . Estas replicas se obtendrán a partir de la simulación del modelo propuesto, dándole a los parámetros valores específicos que vendrán de muestras de sus respectivas posteriores. Como ya hemos mencionado antes, estas muestras se encuentran en las entradas de las matrices antes definidas que nos regresa el MCMC. Así pues la réplica  $k$  de  $y_{ij}$  ( $y_{ij}^{\text{rep}k}$ ) para el un individuo  $j$  fijo, será el resultado de la  $k$ -ésima simulación del modelo tomando la  $k$ -ésima muestra conjunta con repetición de la matriz  $M_j$  (el vector columna  $k$  definido por  $M_j[:, k]$ ), a nivel individual, o bien la  $k$ -ésima entrada de la  $j$ -ésima fila de cada una de las matrices  $Q_h$ ,  $h = 1 : p$  para el nivel jerárquico. Es decir,

$$\begin{aligned} & y_{ij}^{\text{rep}k}(M_j[:, k]), \text{ a nivel individual} \\ & y_{ij}^{\text{rep}k}(Q_h[j, k]), \text{ para cada individuo a nivel jerárquico y } \forall h = 1 : p \end{aligned}$$

con  $k \leq s$ . Si con los datos que poseemos podemos obtener el comportamiento de un individuo promedio observado  $\bar{y}_i$  entonces la  $k$ -ésima repetición para dicho individuo representativo de la operación estará dada por la  $k$ -ésima simulación de nuestro modelo tomando como valores de los parámetros la  $k$ -ésima entrada con repetición de cada uno de los vectores  $\boldsymbol{\theta}_h$ ,  $h = 1 : p$ , i.e.,

$$\bar{y}_i^{\text{rep}k}(\boldsymbol{\theta}_h[k]) \quad \forall h = 1 : p$$

Una vez que tenemos suficientes repeticiones, entiéndase por suficientes un número tal que el que el estadístico sea significativo, aplicamos dicho estadístico a cada una de las repeticiones  $\Psi(y_{ij}^{\text{rep}k})$  con  $j = 1 : J$  fija y comparamos con el estadístico de los datos originales. Esta comparación se hace de diversas maneras y muchas veces es específica de los datos o del fenómeno que se desee modelar. Si dado cierto criterio, la comparación falla, entonces el

modelo se rechaza.

Como ya mencionamos, este es un método cualitativo para descartar modelos y también para compararlos, aunque dicha comparación resulta subjetiva y muchas veces hecha, como coloquialmente se dice, “al ojo” cuando dos o más modelos resultan aceptables para describir los datos observados y predecir futuros datos que no han sido observados aún. En la sección de resultados de este capítulo explicaremos a detalle el estadístico que utilizamos y la forma de comparación. Por lo pronto, en la siguiente sección expondremos un par de métodos de comparación cuantitativa que han sido ampliamente estudiados a nivel teórico.

## 5.7. Criterio de Información de Akaike y de Watanabe-Akaike

Los modelos bayesianos pueden ser evaluados y comparados de diversas maneras. Es de interés evaluarlos y compararlos pues en muchas ocasiones, en el afán de mejorar nuestro modelo de ajuste (hacerlo más apegado a la realidad), aumentamos el número de parámetros a considerar y logramos con ello (a veces) una mejor reproducibilidad de los datos y una mejor predictibilidad, por lo que nos es de interés comprobar si estas mejoras son lo suficientemente buenas para justificar un modelo más complicado. En la sección anterior mostramos una herramienta de comprobación de modelos, el PPC, que también puede ser útil para comparar el nivel de ajuste y predicción entre varios modelos posibles. Como mencionamos ya, esta herramienta es solamente cualitativa, pues el resultado de dicha evaluación y comparación depende completamente del observador pero no responde de manera exacta a si el aumento en la complejidad del modelo nos lleva directamente a un mejor modelo, no sólo en su nivel de ajuste sino también en la capacidad de predicción de datos no observados. Así, cuando existen varios modelos que resultan ser aceptables para describir los datos observados nos interesa contar con una herramienta más precisa de comparación para poder elegir el mejor de ellos. La forma más intuitiva de evaluar y comparar un modelo es a través de la exactitud de sus predicciones. Existen muchas herramientas para cuantificar esta característica en un modelo, entre ellas destacan varios criterios de información como el criterio de información de devianza o el criterio de información bayesiana [141], en este escrito nos enfocaremos en definir el Criterio de Información de Akaike (AIC por sus siglas en inglés) y, en su generalización, el Criterio de Información de Watanabe-Akaike (WAIC por sus siglas en inglés) pues este segundo es el que utilizaremos en las posteriores secciones de resultados.

Consideremos los datos observados  $\mathbf{y} = (y_1, \dots, y_n)$  como independientes dados los parámetros  $\boldsymbol{\theta}$ , así el likelihood puede descomponerse como  $[\mathbf{y}|\boldsymbol{\theta}] = \prod_{i=1}^n [y_i|\boldsymbol{\theta}]$ . Para medir la exactitud de las predicciones de un modelo existen las medidas de precisión predictiva para predicciones probabilísticas que son llamadas reglas de puntuación. La puntuación logarítmica es la regla más utilizada en la selección y comparación de modelos [142]. El resumen más general del ajuste predictivo es la densidad predictiva logarítmica,  $\log[\mathbf{y}|\boldsymbol{\theta}]$ , que es proporcional al error cuadrático medio si el modelo es normal con varianza constante. La densidad predictiva

logarítmica es también llamada log-likelihood o simplemente la puntuación logarítmica [141].

Como hemos ya mencionado, estamos interesados en medir y comparar el desempeño predictivo entre modelos cuando se les dan datos fuera de la muestra observada, es decir, para datos nuevos que fueron producidos con el mismo proceso *real* de generación de los datos observados  $\mathbf{y}$ . Sea  $f$  dicho modelo *real*,  $\mathbf{y}$  el conjunto de datos observados (entonces, una sola realización del conjunto de datos  $\mathbf{y}$  viene de la distribución  $f(\mathbf{y})$ ) y sea  $\tilde{\mathbf{y}}$  el conjunto de datos no observados. El ajuste para la predicción fuera de la muestra (fuera del conjunto de datos sobre el cual se calcularon las posteriores de los parámetros del modelo de ajuste) para un nuevo dato  $\tilde{y}_i \in \tilde{\mathbf{y}}$  está dado por:

$$\begin{aligned} \log[\tilde{y}_i]_{\text{post}} &= \log E_{\text{post}}([\tilde{y}_i | \boldsymbol{\theta}]) \\ &= \log \int [\tilde{y}_i | \boldsymbol{\theta}] [\boldsymbol{\theta}]_{\text{post}} d\boldsymbol{\theta} \end{aligned}$$

donde  $[\tilde{y}_i]_{\text{post}}$  es la densidad predictiva para  $\tilde{y}_i$  inducida por la distribución posterior  $[\boldsymbol{\theta}]_{\text{post}}$  obtenida a partir de  $y_i \in \mathbf{y}$ . Introducimos la notación  $[\ ]_{\text{post}}$  para representar la distribución posterior pues más adelante las expresiones se harán más complicadas y escribir las condicionales completas podría generar confusión. Usaremos también  $E_{\text{post}}$  para referirnos al promedios sobre la distribución posterior de  $\boldsymbol{\theta}$  [141].

Por otro lado, interesa calcular la esperanza de la densidad predictiva logarítmica fuera de la muestra, la cual viene dada por:

$$\begin{aligned} \text{elpd} &= \text{esperanza de la densidad predictiva logarítmica para un nuevo dato puntual} \\ &= E_f(\log[\tilde{y}_i]_{\text{post}}) = \int (\log[\tilde{y}_i]_{\text{post}}) f(\tilde{y}_i) d\tilde{y}, \end{aligned} \quad (5.42)$$

con  $E_f$  el promedio con la función  $f$ . Para mantener comparabilidad con el conjunto de datos observados  $\mathbf{y}$ , asumimos que  $\tilde{\mathbf{y}} = (\tilde{y}_1, \dots, \tilde{y}_n)$ , con lo que se puede definir una medida de la precisión predictiva para los  $n$  datos del conjunto de datos no observados, tomados uno a la vez como:

$$\begin{aligned} \text{elppd} &= \text{esperanza de la densidad predictiva puntual logarítmica} \\ &\quad \text{para un nuevo conjunto de datos} \\ &= \sum_{i=1}^n E_f(\log[\tilde{y}_i]_{\text{post}}). \end{aligned} \quad (5.43)$$

En la práctica, el parámetro  $\boldsymbol{\theta}$  no es conocido, y la densidad predictiva logarítmica no puede calcularse de manera exacta, es por eso que se trabaja con la densidad posterior  $[\boldsymbol{\theta}]_{\text{post}}$  para calcular la precisión predictiva del modelo ajustado a los datos que se define como:

$$\begin{aligned}
\text{lppd} &= \text{densidad predictiva puntual logarítmica} \\
&= \log \prod_{i=1}^n [y_i]_{\text{post}} \\
&= \sum_{i=1}^n \log \int [y_i | \boldsymbol{\theta}] [\boldsymbol{\theta}]_{\text{post}} d\boldsymbol{\theta}.
\end{aligned} \tag{5.44}$$

Ahora bien, para calcular la densidad predictiva en la práctica se hace necesario evaluar la esperanza en muestras aleatorias tomadas de  $[\boldsymbol{\theta}]_{\text{post}}$ , las cuales denotaremos por  $\boldsymbol{\theta}^{(s)}$ ,  $s = 1, \dots, S$  y  $S$  un número lo suficientemente grande. Entonces (5.44) queda como:

$$\begin{aligned}
\text{clppd} &= \text{densidad predictiva puntual logarítmica calculada} \\
&= \sum_{i=1}^n \log \left( \frac{1}{S} \sum_{s=1}^S [y_i | \boldsymbol{\theta}^{(s)}] \right).
\end{aligned} \tag{5.45}$$

Fórmulas como (5.43) no pueden ser calculadas de manera directa porque no conocemos  $f$  y la estimación de ésta dada por (5.45) resulta muchas veces inadecuada [141]. Pero existen varias aproximaciones que intentan corregir la estimación del elppd y de las cuales podemos echar mano a la hora de querer comparar entre varios modelos de ajuste para elegir el mejor de ellos. En particular, podemos considerar la aproximación dada por Akaike en 1973 [143], cuya idea general es la siguiente: Dado que el clppd definido en la Ec.(5.45) es una estimación sesgada del elppd definido en la Ec.(5.43), el siguiente paso lógico es corregir ese sesgo. En la fórmula de Akaike dicha corrección se da al restar el número de parámetros del modelo de ajuste de la estimación clppd y multiplicar todo por un factor de -2. La ecuación queda pues como sigue:

$$\text{AIC} = -2\text{clppd} + 2k, \tag{5.46}$$

con  $k$  el número de parámetros ajustados del modelo. Por todo lo mencionado anteriormente, resulta intuitivo hacer esta corrección agregando sólo, y de manera directa, el número de parámetros considerados en el modelo de ajuste, pero cuando el modelo es no lineal o considera previas altamente no informativas, la corrección ya no resulta adecuada agregando simplemente  $k$  a la ecuación pues se genera un sobreajuste aún mayor que al sólo considerar el clppd [141]. Es por eso que en 2010 Watanabe propone en [144] una aproximación bayesiana más completa para estimar la esperanza fuera de la muestra definida en (5.43) añadiendo una corrección al sobreajuste por medio del número efectivo de parámetros del modelo. Este número efectivo está definido en términos del likelihood como sigue [141]:

$$p_{\text{WAIC}} = 2 \sum_{i=1}^n \left( \log \left( \frac{1}{S} \sum_{s=1}^S [y_i | \boldsymbol{\theta}^{(s)}] \right) - \frac{1}{S} \sum_{s=1}^S \log [y_i | \boldsymbol{\theta}^{(s)}] \right) \tag{5.47}$$

Así, el ajuste de la precisión predictiva está dado por [141]:

$$\text{WAIC} = -2\text{clppd} + 2p_{\text{WAIC}} \quad (5.48)$$

Por lo tanto, el modelo con el menor WAIC (AIC)- y por lo tanto, con la mayor esperanza de la densidad predictiva- es también el modelo con la mejor precisión predictiva. En particular, el WAIC funciona bien con modelos a nivel individual pero resulta extraordinariamente útil también en modelos con estructura jerárquica en los cuales el número de parámetros crece con el tamaño de la muestra [141].

En el contexto del trabajo presentado en este capítulo nos centraremos en el cálculo del WAIC, éste nos resulta de interés pues presentaremos cuatro modelos bayesianos que escalan en complejidad en el afán de ajustar mejor los datos de movimiento de 21 ciervos canadienses pero, generalmente, modelos más apegados a la realidad (si bien tienen la ventaja de ser más descriptivos) tienen la gran desventaja de ser mucho más difíciles de comprender y de calcular. Así, la pregunta clave en la comparación de modelos, ¿la mejora en el ajuste es lo suficientemente buena para justificar un modelo más complejo?, queda resuelta por el cálculo del WAIC.

En las siguientes secciones nos enfocaremos en presentar los detalles y resultados de los cuatro modelos bayesianos que acabamos de mencionar, al igual que las especificaciones de los datos que se utilizaron para hacer el ajuste de dichos modelos. Con estos datos nos interesa estudiar el efecto que el uso de la memoria tiene sobre el uso del espacio, el desplazamiento y la emergencia de ámbitos hogareños en los ciervos canadienses. Nos es también de interés entender, en la medida de lo posible, cuáles son los mecanismos bajo los cuales el uso de la memoria se traduce en aprendizaje (afinidad por sitios), es decir, si los ciervos crean afinidad por los sitios visitados en un pasado reciente (decaimiento de memoria) o bien son capaces de recordar y volver a sitios visitados en el pasado lejano (memoria de largo alcance). Los datos cuentan con una particularidad que resulta de gran interés y beneficio para este estudio. Mencionamos las especificaciones a continuación.

## 5.8. Los datos

Los datos utilizados en este estudio son los mismos que los colectados por Frair *et al.* y presentados en [145, 146]. Las especificaciones son las siguientes:

El área de estudio consistió de 15 800 km<sup>2</sup> que se extienden a lo largo de las pendientes del este de las Montañas Rocosas en el centro de Alberta, Canadá. Aproximadamente 2000 ciervos habitaban el área durante el período de estudio que abarcó de Diciembre del 2000 a Septiembre de 2002 [145]. La elevación del terreno era de 500-1500 m y el área era en gran parte boscosa (68.7% del área total). Las especies dominantes de árboles incluían: pino lodgepole (*Pinus Contorta*), picea blanca (*Picea Glauca*), y álamo (*Populus Tremuloides* y *P. Balsamea*). Intercaladas en la matriz boscosa se encontraron: praderas tanto secas como húmedas (7.1%), corte forestal que precede a la tala de madera (4.3%), suelo desnudo/afloramientos rocosos (12.3%), ríos y lagos (2.1%), y áreas regeneradas de incendios

forestales (<1 %) [145, 146].

Durante el período de estudio 21 ciervos hembras fueron reubicadas al área de estudio desde tres sitios dentro de Alberta: 1) Los Parques Nacionales Banff y Jasper, 2) El Parque Nacional de la Isla del Ciervo, y 3) El área de Conservación Cruzada (20 km al suroeste de Calgary, Canadá) (ver [145] para más detalles de estos tres sitios de origen). Los animales incluían: Seis hembras de la ciudad de Banff liberadas en Febrero del 2001. Nueve hembras liberadas desde el área Cruzada, seis durante Diciembre del 2000 y tres más en Diciembre del 2001. Seis hembras fueron liberadas desde la Isla del Ciervo entre Enero y Febrero del 2002. Los animales fueron capturados utilizando, principalmente, trampas de corral cebadas con heno. Estos animales fueron transportados hacia las áreas de liberación en remolques de ganado que tenían entre nueve y dieciséis animales dependiendo de la composición de sexo y clase de edad. Los ciervos fueron liberados directamente del remolque al área de estudio. Los animales fueron liberados en un número separado de ubicaciones para aumentar la independencia entre los resultados de diferentes individuos [145, 146].

Previo a la liberación los ciervos reubicados fueron equipados con collares de radio y GPS (LMRT4 and GPS2200, Lotek Wireless, ON, Canadá) que rastreaban y recolectaban las ubicaciones de los animales cada 2 h durante los 11 meses de estudio. Se utilizaron todas las ubicaciones de cada uno de los collares de radio durante una temporada o hasta que el contacto digital del radio collar se perdía, el animal moría o el collar GPS se recuperaba (luego de 11 meses de la liberación). Todos los collares fueron equipados con sensores de mortalidad que se activaban luego de 7 h de inmovilidad. Las pruebas de collar en el rango de cobertura y condiciones del terreno encontradas dentro del área de estudio indicaron una alta tasa de fijación y precisión posicional de  $\leq 50$  m el 80 % de las veces [145, 146].

Se identificaron parches de forraje para ciervos de una cuadrícula de cobertura terrestre desarrollada para esta región (ver [147]). La cuadrícula tiene un tamaño de celda de 28.5 m y una precisión de clasificación general de 82.7 %. Utilizando ArcGIS<sup>3</sup> [148], se reclasificaron los prados secos/mésicos y húmedos, los matorrales, los claros, se recuperaron las clases herbáceas en una sola clase de hábitat de alimentación, y se convirtió la cuadrícula en una capa poligonal. Se eliminaron los polígonos de tamaño menor a 0.27 ha (< 3 píxeles contiguos) y se conservaron 16,782 parches para el análisis. Los parches conservados promedian  $6.93 \pm 29.4$  ha en tamaño. Para cada ubicación GPS, para cada uno de los ciervos, que ocurre dentro de un parche, se registra el *número único* para ese parche (cada parche se encuentra identificado de manera única por una etiqueta o *número único*), lo que permite obtener información sobre el tiempo dedicado a moverse entre parches, el tiempo de residencia dentro de los parches y el tiempo de regreso a los parches visitados previamente. Por lo tanto, se transforman las trayectorias GPS originales en una serie temporal de visitas de parche a parche que incluye el tiempo dedicado a cada parche y el tiempo de viaje entre parches. En los datos analizados en esta sección se asume omite, además, las visitas consecutivas a cada

---

<sup>3</sup>Es el nombre de un conjunto de productos de software en el campo de los Sistemas de Información Geográfica o SIG. Producido y comercializado por Instituto de Investigaciones en Sistemas Ambientales (California), bajo el nombre genérico ArcGIS se agrupan varias aplicaciones para la captura, edición, análisis, tratamiento, diseño, publicación e impresión de información geográfica.

parche.

## 5.9. Modelos

Para cada modelo considerado más adelante se consideran las siguientes suposiciones:

- Los animales se mueven en un ambiente estacionario de dos dimensiones que consiste de un conjunto de  $N$  parches (sitios de recursos) disponibles.  $N$  es un número fijo y se obtiene de mapas de recursos recolectados, como se ha especificado ya en la sección anterior. Los parches están caracterizados por su abundancia de recurso (área)  $a_n$  con  $n \in \{1, 2, \dots, N\}$ , el área está medida en ha. La distancia entre los parches  $n$  y  $m$  será la distancia euclideana entre sus respectivos centroides y la denotaremos por  $d_{n,m}$ , la unidad de distancia considerada aquí es en km (por simplicidad de cómputo convertimos los datos recolectados en m a km).
- Asumiremos que el tiempo es discreto y que los animales sólo gastan una unidad de tiempo en cada parche y que a cada paso  $t \rightarrow t + 1$  el individuo decide moverse a otro parche (el movimiento sólo se da entre parches) siguiendo un conjunto de reglas que se explicarán más adelante.
- Dado que un individuo ha decidido abandonar el parche  $n$ , irá al parche  $m$  con probabilidad  $P_{n,m}$ . Esta probabilidad se calculará de manera diferente en cada modelo.
- Todos los parámetros a estimar  $\in \mathbf{R}^+$ .

### 5.9.1. Modelo I

En este modelo supondremos que cada ciervo escoge visitar alguno de los parches disponibles en el ambiente considerando la distancia a su actual posición  $n$  y la abundancia de recurso. Asumiremos que este tipo de movimiento es meramente aleatorio ya que el individuo no sigue ninguna de las reglas de revisita preferencial expuestas en los capítulos anteriores, es decir, no posee memoria y sus decisiones sólo se basan en la distancia entre parches y los tamaños de éstos (áreas).

Considerando lo anterior definimos un vector de probabilidad  $\mathbf{k} = (k_1, \dots, k_N)$  cuya entrada  $m$  nos dará la probabilidad de que el ciervo elija ir al parche  $m$  desde el parche  $n$  que ocupa. Cada entrada del vector  $\mathbf{k}$  se define como:

$$k_m = d_m * c_m / \sum_r d_r * c_r, \quad r = 1, \dots, N \quad (5.49)$$

con  $d_m = \exp(-(d_{n,m}/\alpha)^\beta)$  y  $c_m = \kappa a_m^\lambda / (1 + \kappa a_m^\lambda)$  es decir, suponemos que la probabilidad de visitar el parche  $m$  dado que el animal se encuentra en el parche  $n$  decae exponencialmente

con la distancia y crece de forma logística<sup>45</sup> con el área. El principal objetivo será obtener una estimación jerárquica para los parámetros  $\alpha, \beta, \lambda$  y  $\kappa$ .

### 5.9.2. Modelo II

Considerando el hecho, antes ya mencionado aquí, de que reciente evidencia experimental ha mostrado que muchas especies animales tienen la capacidad de utilizar la memoria para regresar a sitios previamente visitados y habiendo mostrado ya la importancia que la memoria tiene en el aprendizaje individual y colectivo, incorporamos el parámetro  $q \in (0, 1)$  que define la frecuencia con la cual el individuo decide usar su memoria para escoger un parche del ambiente visitado en el pasado, este parámetro  $q$  es similar al ya utilizado en los capítulos 3 y 4 de este mismo escrito. En este Modelo II asumiremos además que el animal posee memoria infinita (así como el modelo expuesto en el capítulo 3), es decir, es capaz de recordar todos y cada uno de los sitios de recurso que ha visitado y el tiempo que ha gastado en ellos (cantidad de visitas). La regla de revisita preferencial será lineal. Por lo tanto, este modelo tiene, intrínsecamente, dos tipos de movimiento:

- Con probabilidad  $q$  el animal se moverá del parche  $n$  al parche  $m$  considerando, además de la distancia y el área de éste, la memoria, es decir, si el sitio  $m$  ha sido o no visitado en el pasado y la cantidad de visitas que ha recibido en el caso de que el individuo ya haya estado en él. Entonces la entrada  $m$  del vector de probabilidad  $\mathbf{k}$  se modificará como sigue:

$$\mathbf{k}_m = d_m * c_m * m_m / \sum_r d_r * c_r * m_r, \quad r = 1, \dots, N \quad (5.50)$$

con  $d_m$  y  $c_m$  definidas en (5.49). Ahora bien,  $m_m = 0$  si el animal nunca ha estado en el parche  $m$  antes y  $m_m = n_m$  con  $n_m$  el número de visitas al sitio  $m$  hasta el tiempo actual  $t$ .

- Con probabilidad  $1 - q$  el individuo no usará su memoria y escogerá el parche  $m$  usando el vector de probabilidad  $\mathbf{k}$  definido en (5.49). Podemos decir entonces que el animal eligió hacer un movimiento aleatorio.

Por lo tanto, ahora tenemos un nuevo parámetro  $q$  a estimar, haciendo un total de cinco parámetros a ajustar.

### 5.9.3. Modelo III

Si tomamos en cuenta que las trayectorias en el conjunto de datos considerado aquí pertenecen a animales que fueron puestos en el ambiente sin tener ningún tipo de información previa de él, entonces es razonable pensar que al inicio sus movimientos son gobernados por elecciones aleatorias en lugar del uso de memoria. Esta última puede ser clave a tiempos grandes para los pasos finales en la trayectoria una vez que el individuo ha explorado el ambiente (o parte de él) y ha establecido una zona preferencial de forrajeo. Así, sería razonable

<sup>4</sup>No confundir con el mapa logístico.

<sup>5</sup>En el cálculo computacional consideramos el  $\log(a_m)$ , así la ecuación logística  $f(x) = e^x / (1 + e^x)$  toma la forma dada por  $c_m$  cuando  $x = \lambda \log(a_m) + \tilde{\kappa}$ , con  $\kappa = \exp(\tilde{\kappa})$ .

pensar que el parámetro que dicta la frecuencia del uso de memoria se modifica con el tiempo.

Ahora para estimar el parámetro de memoria consideraremos el número de sitios únicos visitados (uvs) por el individuo al tiempo  $t$ . Entonces definiremos  $\mathbf{u} = (u_1, \dots, u_T)$  un vector de longitud  $T$ , donde  $T$  es la longitud de la trayectoria. Así, cada entrada de este vector  $\mathbf{u}$  nos dirá la cantidad de sitios únicos (sitios distintos) que ha visitado el caminante al tiempo  $t$  con  $u_1 = 1$ . Este vector es un dato observado y  $q$  dependerá de él cómo sigue:

$$q(\mathbf{u}_t) = 1 - \exp(-(\mathbf{u}_t/\rho)^\epsilon) \quad (5.51)$$

Así, la frecuencia del uso de memoria crecerá de manera exponencial con el número de sitios únicos visitados.

Para este modelo el número total de parámetros a estimar será seis, los cuatro ya considerados en el Modelo I más los dos parámetros para el incremento del uso de memoria con el número de sitios únicos visitados ( $\rho$  y  $\epsilon$ ).

#### 5.9.4. Modelo IV

Hasta ahora hemos considerado, en los Modelos II y III, que los animales poseen memoria infinita, pero como hemos ya estudiado en el capítulo anterior, un decaimiento de memoria podría significar una mejora en la búsqueda y explotación de los mejores sitios de recurso. Así, en este Modelo IV cambiaremos la regla de revisita preferencial lineal a una que decae exponencialmente con el tiempo. Ahora, el individuo olvidará aquellos parches visitados en un pasado lejano y recordará con mayor fuerza los sitios visitados recientemente, además de considerar la cantidad de tiempo que ha pasado en cada uno de ellos. Por lo tanto, los parches visitados en el pasado reciente tendrán una mayor probabilidad de ser visitados nuevamente. Aquí cabe hacer una nota especial; en el capítulo anterior consideramos un decaimiento de memoria en ley de potencia, mientras que en este capítulo consideraremos un decaimiento exponencial o de exponencial estirada, la razón para hacer este cambio es que las trayectorias observadas no son lo suficientemente largas para lograr ver una mejora en el ajuste tomando en cuenta sólo un cambio en la función de decaimiento. Así que por simplicidad (sobre todo computacional) hemos elegido una función exponencial para el decaimiento de memoria.

Para considerar este nuevo supuesto en la definición del modelo tomaremos en cuenta la siguiente expresión:

$$m_m(t) = \sum_{i=1}^{n_m} \exp\{-[(t - t_i)/\nu]^\theta\} \quad (5.52)$$

con  $n_m$  el número de visitas al parche  $m$  hasta el tiempo  $t$  y  $t_i$  los tiempos en los que dicho parche recibió cada una de estas visitas. Por lo tanto el vector de probabilidad  $\mathbf{k}$  estará definido por la ya conocida Ec.(5.50) con  $m$  dada en la ecuación anterior. Cabe mencionar que  $m$  de la Ec.(5.52) marcará un decaimiento exponencial cuando  $\theta = 1$ , uno de exponencial estirada si  $\theta < 1$ , y un decaimiento súper exponencial si  $\theta > 1$ . Así pues tenemos ahora ocho parámetros a estimar: los seis parámetros ya definidos en el Modelo III y dos más para el

decaimiento exponencial de la memoria ( $\nu$  y  $\theta$ ). Aquí es importante recalcar que una cosa es la frecuencia con la que el animal decide usar sus memoria ( $q$ ) y otra muy distinta la capacidad de memoria que posee (si es ésta infinita o decae con el tiempo).

Así, el objetivo principal en este capítulo es ajustar cada uno de estos modelos al conjunto de datos expuesto antes y compararlos entre ellos para encontrar cuál es el “mejor” no sólo para explicar el mecanismo subyacente con el cual los animales escogen los parches que visitan si no también para predecir algunas características esenciales de nuevas trayectorias aún no observadas, como el número de sitios únicos visitados que se traduce directamente en uso del espacio. Para probar el nivel de ajuste de cada uno de los modelos utilizaremos el PPC (definido ya en la secc. 5.6 de este mismo escrito) y para comparar el nivel predictivo entre modelos calcularemos el WIAC (secc. 5.7) para cada uno de ellos. Las especificaciones a considerar para la estimación de los parámetros se explican a continuación.

## 5.10. Distribuciones Previas

En medida de lo posible utilizaremos previas poco informativas para los parámetros a estimar. Las previas y la interpretación de cada uno de los parámetros se especifican en la Tabla 5.1).

Los modelos se ajustaron implementado el algoritmo a dos pasos propuesto en [73]. El primer paso del algoritmo consiste en aproximar el conjunto de datos a un nivel individual (modelos no jerárquicos) y de manera independiente utilizando las previas mostradas en la Tabla 5.1 para el ajuste de los parámetros. Este primer paso fue hecho utilizando el MCMC implementado desde el software Stan [149] disponible para R [150]<sup>6</sup>. Para todos los modelos se ejecutaron, para cada uno de los individuos, tres cadenas independientes con 5 000 iteraciones cada una para así obtener las distribuciones posteriores para cada uno de los parámetros. La segunda fase del algoritmo se trata de implementar una dinámica simplificada del Markov Chain Monte Carlo (MCMC) para aproximar el modelo jerárquico gaussiano utilizando como entrada las posteriores obtenidas en la fase anterior [73].

---

<sup>6</sup>El algoritmo MCMC disponible en Stan es una extensión del algoritmo explicado en secciones anteriores, esta extensión es conocida como Monte Carlo Hamiltoniano (HMC por sus siglas en inglés). De manera general, esta nueva extensión actualiza la cadena reduciendo la correlación entre estados muestreados sucesivos mediante el uso de una evolución hamiltoniana entre estados. Como consecuencia inmediata de esto último, la cadena converge más rápidamente (que el MCMC común) a la distribución de probabilidad absoluta. En [151] se hace un estudio detallado de esta extensión.

Parámetro	Distribución previa	Interpretación
$\alpha$	normal(0,10)	Parámetro de escala para la función exponencial que define la disminución de la probabilidad con la distancia.
$\beta$	normal(0,1)	Parámetro de forma para la función exponencial que define la disminución de la probabilidad con la distancia.
$\lambda$	normal(0,1)	Parámetro para la pendiente de la función logística que define el incremento de la probabilidad como función del área de los parches.
$\kappa$	normal(0,1)	Parámetro para la ordenada al origen de la función logística que define el incremento de la probabilidad como función del área de los parches.
$q$	beta(1,1)	Parámetro que define la frecuencia del uso de memoria.
$\rho$	normal(0,10)	Parámetro de escala para la función exponencial que define el aumento del uso de memoria como función de los sitios únicos visitados.
$\epsilon$	normal(0,1)	Parámetro de forma para la función exponencial que define el aumento del uso de memoria como función de los sitios únicos visitados.
$\nu$	normal(0,10)	Parámetro de escala para la función exponencial que define el decaimiento de la memoria como función del tiempo desde la última visita a cada parche.
$\theta$	normal(0,10)	Parámetro de forma para la función exponencial que define el decaimiento de la memoria como función del tiempo desde la última visita a cada parche.

Cuadro 5.1: Distribuciones previas

Los resultados obtenidos al ajustar cada uno de los modelos, con sus respectivas especificaciones, se muestran en la siguiente sección.

## 5.11. Parámetros determinados

### Modelo I

La Fig.5.2 muestra las medias de la posterior marginal (puntos) e intervalos de credibilidad al 95 % (lineas verticales) para cada uno de los individuos (identificados por su ID en el eje X), y la población (denotada por “pop”) para cada uno de los parámetros del Modelo I. En negro (a la izquierda de cada grupo) los resultados de ajuste que corresponden a la primera fase del algoritmo de jerarquización, es decir el ajuste de cada parámetro individual  $p_j$  para cada uno de los  $j = 1 : 21$  animales a nivel individual. En gris claro (a la derecha de cada grupo) los resultados de la segunda fase del algoritmo de jerarquización para cada individuo, es decir, los ajustes de cada parámetro individual  $p_j$  a nivel jerárquico. En gris oscuro se muestra el resultado de la segunda fase del algoritmo para el parámetro poblacional  $p$ .

Las medias de la posterior marginal para el parámetro individual  $\alpha_j$  (Fig.5.2-Arriba Izquierda) que determina la escala para la función exponencial que define la disminución de la probabilidad de que el animal escoja un determinado parche en función de la distancia de éste a su posición actual tiene ligeras fluctuaciones entre cero y cinco ( $0 < \alpha_j < 5$ ) para casi todos los individuos a nivel jerárquico (lineas gris claro), en complemento, el parámetro  $\beta$  (parámetro de forma de la función exponencial de la distancia, Fig.5.2)- Arriba Derecha) tiene fluctuaciones entre cero y dos, i.e.  $0 < \beta < 2$  para los 21 individuos a nivel jerárquico. El valor medio de ambos parámetros a nivel poblacional (“pop”) es de  $\alpha = 1.02$  y  $\beta = 0.93$ . Estos valores medios nos dicen que, a nivel jerárquico, la distancia juega un papel muy importante a la hora de elegir un nuevo parche a visitar. La probabilidad de visitar un sitio como función de la distancia de éste a la posición actual decae muy rápido (aproximadamente de forma exponencial ya que  $\beta \approx 1$ ), es decir, entre más lejos esté el nuevo sitio a visitar, menor probabilidad hay de en verdad escogerlo. Este decaimiento rápido con la distancia puede interpretarse, también, como un movimiento similar al de una caminata aleatoria a primeros vecinos, ya que la mayor probabilidad de salto se encuentra en aquellos parches cercanos al parche de ubicación actual. La Fig.5.3-Izquierda muestra el cambio del vector  $d$  como función de la distancia  $d$  a nivel poblacional. Nos interesa el nivel poblacional pues éste define el comportamiento de un individuo promedio, entendiéndolo por individuo promedio: o bien un individuo representativo de toda la población o bien un individuo aún no observado. Para obtener esta gráfica calculamos varias curvas de  $d$  a partir de distintas muestras de las densidades marginales de los parámetros  $\alpha$  y  $\beta$  a nivel poblacional, en gris claro se observan estas curvas y en gris oscuro el intervalo de confianza al 95 % definido por éstas. Es fácil observar que a nivel poblacional no se escogen parches que estén más allá de uno o dos kilómetros de la posición actual. Aquí es importante recalcar que el parámetro  $\alpha$  tiene una marcada diferencia entre el ajuste individual y jerárquico para los individuos 19 y 20. Esto puede ser un efecto de la poca cantidad de datos que tenemos disponibles para esos dos animales.

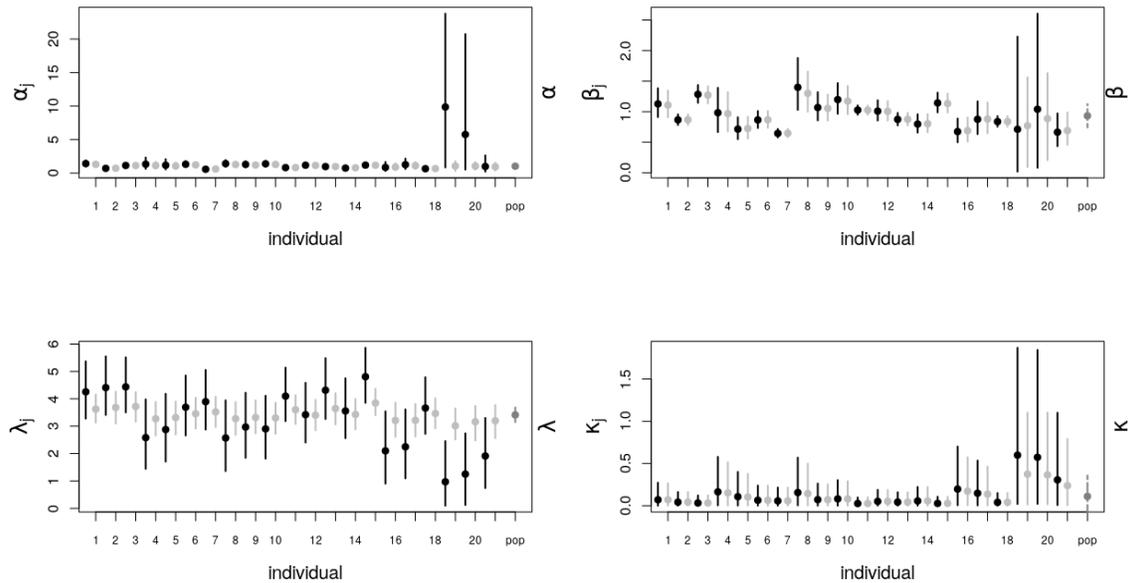


Figura 5.2: Medias de la posterior marginal (puntos) e intervalos de credibilidad al 95% (lineas verticales) para cada uno de los individuos (identificados por su ID en el eje X), y la población (denotada por “pop”) para cada uno de los parámetros del Modelo I. En negro (a la izquierda de cada grupo) los resultados que corresponden a la primera fase del algoritmo de jerarquización. En gris claro (a la derecha de cada grupo) los resultados para la segunda fase del algoritmo de jerarquización para cada uno de los individuos. En gris oscuro se muestra el resultado de la segunda fase del algoritmo para la población.

La Figura 5.2-Arriba Derecha, muestra las medias de las marginales posteriores para el parámetro  $\lambda$  que define la pendiente de la función logística que determina el aumento de la probabilidad de visita a otro parche como función del área (ver nota al pie de la pag. 109). Este parámetro tiene fluctuaciones entre tres y cuatro ( $3 < \lambda < 4$ ) a nivel jerárquico, en complemento, el parámetro  $\kappa$  (la ordenada al origen de la función logística para la probabilidad) fluctúa entre cero y uno ( $0 < \kappa < 1$ ) también a nivel jerárquico. Estos valores nos indican que el área de los parches tiene un impacto importante en el aumento de la probabilidad para elegir un nuevo sitio de visita. El valor medio de ambos parámetros para la población (“pop”) es de  $\lambda = 3.40$  y  $\kappa = 0.11$ . En la Fig.5.3-Derecha podemos observar como, a escala logarítmica en el eje X (ver nota al pie de la pag. 109) y a nivel poblacional, la probabilidad aumenta de manera brusca cuando el área del parche rebasa la ha, alcanzando rápidamente la saturación de la probabilidad para parches con área aproximada de 2 ha (a escala logarítmica). Al igual que pasa con el parámetro  $\alpha$ , el parámetro  $\lambda$  tiene una diferencia significativa en los ajustes para los individuos 19 y 20, algo similar (y por las mismas razones) ocurre para el individuo 21.

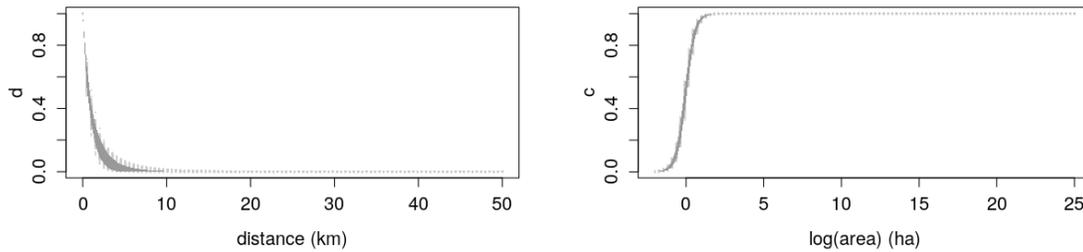


Figura 5.3: Modelo I: Cambio en la probabilidad para escoger un parche como función de la distancia (Izquierda) y como función del área (Derecha) a nivel poblacional.

Podemos decir, de manera tentativa pues este modelo aún no ha sido probado con el PPC, que la probabilidad de escoger un parche para visitar decrecerá rápido con la distancia, es decir, no importa lo bueno que sea un parche, si se encuentra lo suficientemente lejos, el animal lo elegirá con una probabilidad pequeña y decidirá quedarse en los parches que se encuentren mucho más cerca aunque no sean los mejores.

Al ser el Modelo I una caminata aleatoria es pues muy similar a la dinámica de movimiento que el animal realiza cuando se encuentra en una fase de exploración de su ambiente. Así, es razonable pensar que este modelo ajustará (si el PPC da resultados satisfactorios) mucho mejor los inicios de las trayectorias o, bien, aquellas trayectorias que sean básicamente exploratorias. En la sección 5.12 explicaremos esto con más detalle.

En las Tablas 7.1 (a nivel individual) y 7.6 (a nivel jerárquico) de las secciones 7.1 y 7.2 del Apéndice, respectivamente, se pueden encontrar los estadísticos de interés para cada uno de los cuatro parámetros del Modelo I, para cada uno de los 21 individuos. Estos datos nos permitieron obtener los resultados ya expuestos en las Figs.5.2-5.3.

## Modelo II

La Figura 5.4 muestra las medias de la marginal posterior para todos los parámetros del Modelo II. Los resultados para aquellos parámetros que se comparten con el Modelo I no tienen variaciones importantes y, por lo tanto, la interpretación del aumento de la probabilidad con el área y su disminución con la distancia son los mismos ya discutidos en el Modelo I (Fig.5.5).

En este modelo agregamos un nuevo parámetro  $q$  que define la frecuencia del uso de memoria para cada uno de los individuos. Con probabilidad  $q$  el animal tomará en cuenta, además de la distancia y el área, los sitios visitados previamente y el número de visitas a cada uno de ellos, y elegirá, sólo dentro de este conjunto, uno de ellos para visitar. La Fig.5.4-Abajo Izquierda muestra las medias de las marginales posteriores del parámetro  $q$

para cada uno de los individuos. Una de las primeras cosas a observar es que la diferencia entre el nivel individual (en negro) y el nivel jerárquico (en gris claro), es prácticamente nula, lo que significa que no se perdió información que podría resultar valiosa al momento de efectuar el segundo paso del algoritmo de jerarquización y que la corrección hecha por la información del resto de los individuos al pasar del nivel individual al jerárquico, no tuvo mayor impacto en el ajuste (algo similar ocurre para el resto de los parámetros, excepto para los individuos 19 y 20, como se hizo notar con anterioridad).

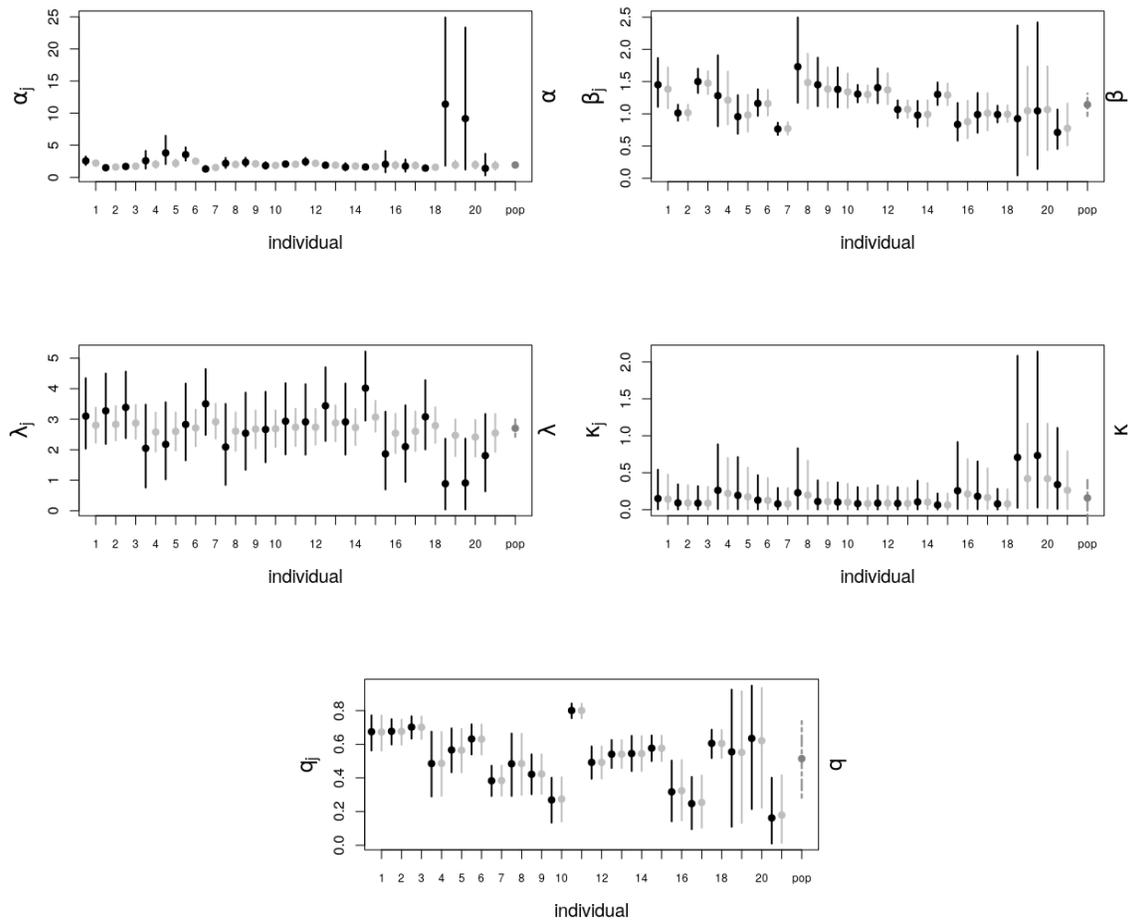


Figura 5.4: Medias de la posterior marginal (puntos) e intervalos de credibilidad al 95% (líneas verticales) para cada uno de los individuos (identificados por su ID en el eje X) y la población (denotada por “pop”) para cada uno de los parámetros del Modelo II. En negro (a la izquierda de cada grupo) los resultados que corresponden a la primera fase del algoritmo de jerarquización. En gris claro (a la derecha de cada grupo) los resultados para la segunda fase del algoritmo de jerarquización. En gris oscuro se muestra el resultado de la segunda fase del algoritmo para el nivel poblacional.

Por otro lado, la mayoría de los individuos tienen un parámetro de uso de memoria alrededor de 0.5, pero existen algunos casos particulares que se salen de esta regla; como el individuo 11 que tiene una frecuencia de uso de memoria bastante grande ( $q_{11} \approx 0.8$ ) comparada con la del resto de sus compañeros y los animales con ID 10,17 y 21 que tienen un valor medio de  $q$  muy pequeño comparado con el resto de los individuos. Esto resulta interesante de recalcar pues dos de estos animales serán tomados como ejemplos característicos para la sección 5.12. El valor medio de  $q$  para el individuo promedio (a nivel poblacional) es  $q = 0.51$ .

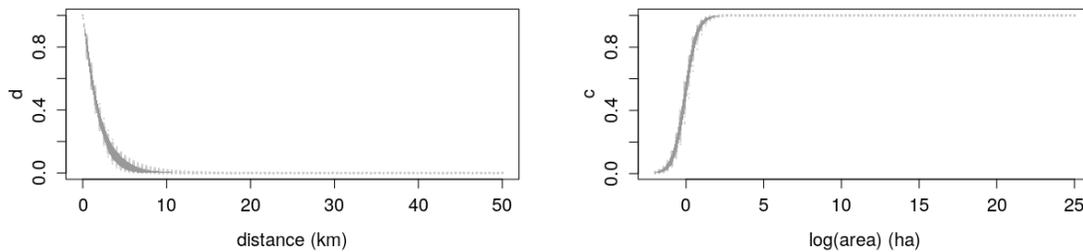


Figura 5.5: Modelo II: Cambio en la probabilidad para escoger un parche como función de la distancia (Izquierda) y como función del área (Derecha) a nivel poblacional.

Para recordar y poner en contexto, un valor muy grande de  $q$  se traduce en una trayectoria gobernada por el uso de memoria, es decir, muchas de las elecciones del animal (aproximadamente la mitad) para la visita a parches se basan en la experiencia pues, con probabilidad alta, el parche a visitar será escogido de entre el conjunto de parches ya visitados. Un valor pequeño de  $q$  se traduce en decisiones frecuentemente aleatorias que se basan solamente en la distancia y área de los parches disponibles. La mayoría de los animales toma elecciones a partes casi iguales, la mitad del tiempo eligen usar la memoria y la otra mitad eligen realizar movimientos aleatorios. Cabe mencionar que para este caso de estudio el uso de memoria es bastante alto en comparación con el reportado para otras especies, por ejemplo en [85] se reporta un  $q \approx 0.2$  para el ajuste de trayectorias reales de monos capuchinos.

En las Tablas 7.2 (nivel individual) y 7.7 (nivel jerárquico) de las secciones 7.1 y 7.2 del Apéndice, respectivamente, se pueden encontrar los estadísticos de interés para cada uno de los cuatro parámetros del Modelo II, para cada uno de los 21 individuos y el individuo promedio. Estos datos nos permitieron obtener los resultados ya expuestos en las Figs.5.4-5.5.

### Modelo III

La Figura 5.6 muestra las medias de las marginales posteriores para los parámetros del Modelo III. Los resultados para los parámetros compartidos con el Modelo I no tienen variaciones importantes a las ya discutidas anteriormente.

En este modelo asumimos que el parámetro de uso de memoria  $q$  crece de manera exponencial con el número de sitios únicos visitados (uvs). Para esto agregamos dos nuevos parámetros,  $\rho$  y  $\epsilon$ . La Fig.5.6-Abajo Izquierda muestra las medias de las marginales posteriores para el parámetro individual  $\rho_j$  y el parámetro poblacional  $\rho$  que determina la escala de la función exponencial que define el aumento de la frecuencia del uso de memoria como función del uvs  $\mathbf{u}$ , el parámetro individual  $\rho_j$  varía (a nivel jerárquico, líneas gris claro) entre seis y doce ( $6 < \rho < 12$ ), en complemento, el parámetro  $\epsilon$  (parámetro de forma de la función exponencial para el aumento de memoria, Fig.5.6-Abajo Derecha) tiene fluctuaciones, entre individuos ( $\epsilon_j$ ) a nivel jerárquico, entre cero y uno ( $0 < \epsilon < 1$ ). Aquí se hace importante recalcar que la diferencia entre las estimaciones a nivel individual y jerárquico para el parámetro individual  $\rho_j$  son significativas en el sentido de que la corrección que proporciona la información estadística poblacional llega a disminuir de manera notable (en algunos casos a más de la mitad) la estimación del parámetro  $\rho_j$ . Veamos esto con calma, en el panel para  $\rho$  de la Fig.5.6 podemos ver que la mayoría de los individuos tienen un valor de  $\rho_j$  cercano a doce ( $\rho_j \approx 12$ ) y que para casi la mayoría de ellos la corrección del ajuste a nivel jerárquico no es mayor pero, en particular, para el individuo 10 la corrección en la estimación hace que su valor a nivel individual descienda a más de la mitad (de  $\rho_{10} \approx 40$  a  $\rho_{10} \approx 12$ ). Para este caso en particular recordemos que el modelo anterior nos dio que a nivel jerárquico la frecuencia del uso de memoria  $q$  del animal 10 era baja, lo cual se traduce en una  $\rho$  muy alta pues el crecimiento del uso de memoria será muy lento como función del uvs. Por otro lado, el individuo 11 tiene una  $\rho$  muy pequeña en comparación a los otros animales, lo cual se traduce un crecimiento rápido del uso de memoria (lo cual coincide con el resultado del modelo anterior) pero aquí la corrección a nivel jerárquico no resulta tan marcada como la del individuo 10.

Algo similar a lo del individuo 10 le ocurre al individuo 15 y al individuo 27. Todos estos resultados son consistentes con lo mostrado en el modelo anterior para  $q$  constante. Es importante mencionar que estas diferencias entre el ajuste a nivel individual y jerárquico pueden deberse tanto a la gran influencia de la información estadística obtenida de los otros individuos, o bien a una pérdida de la misma al ejecutar el segundo paso del algoritmo de jerarquización. En ambos casos, la gran variabilidad que existe en el ajuste de este parámetro  $\rho$  entre individuos a nivel individual lleva a que las estimaciones de este mismo  $\rho$  pero a nivel jerárquico tengan un gran nivel de autocorrelación (Tabla 7.8, columna n\_eff), lo que puede complicar la predicción de datos no observados.

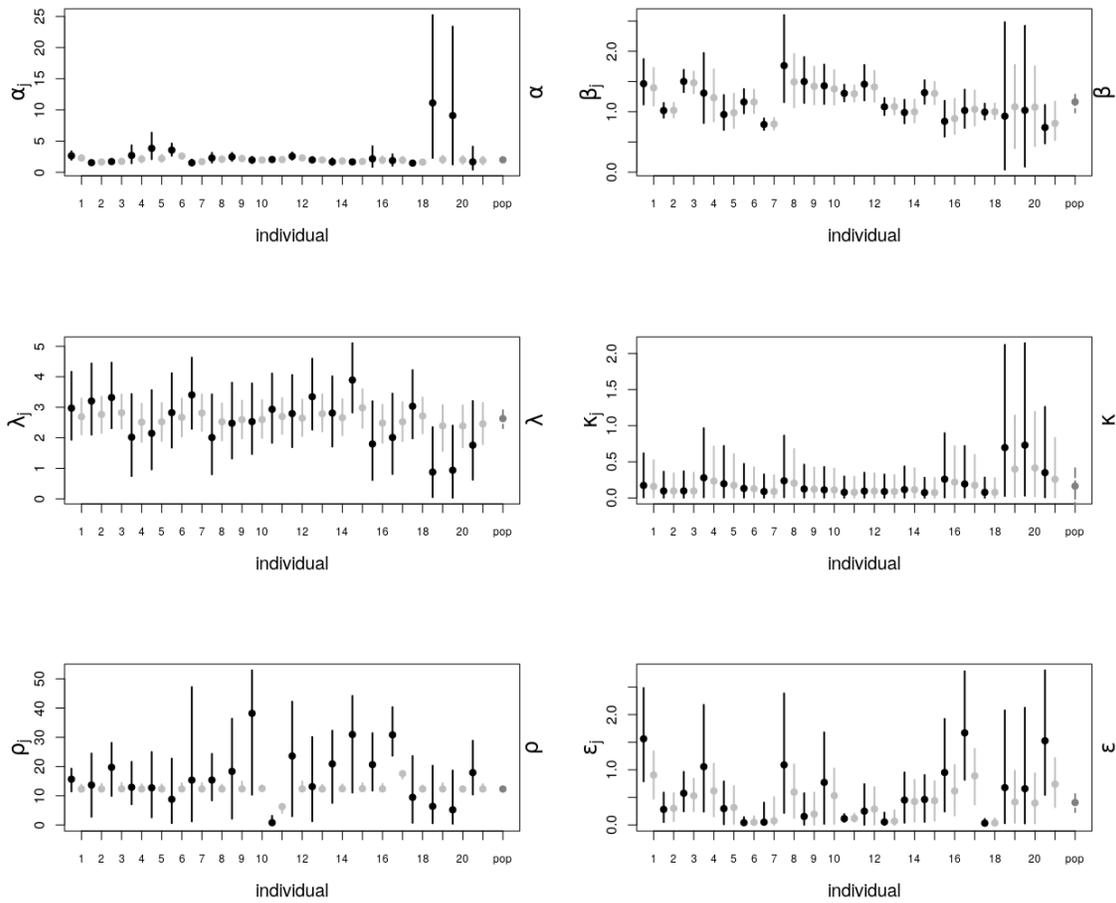


Figura 5.6: Medias de la posterior marginal (puntos) e intervalos de credibilidad al 95% (lineas verticales) para cada uno de los individuos (identificados por su ID en el eje X) y la población (denotada por ‘pop’) para cada uno de los parámetros del Modelo III. En negro (a la izquierda de cada grupo) los resultados que corresponden a la primera fase del algoritmo de jerarquización. En gris claro (a la derecha de cada grupo) los resultados para la segunda fase del algoritmo de jerarquización. En gris oscuro se muestra el resultado de la segunda fase del algoritmo para el nivel poblacional.

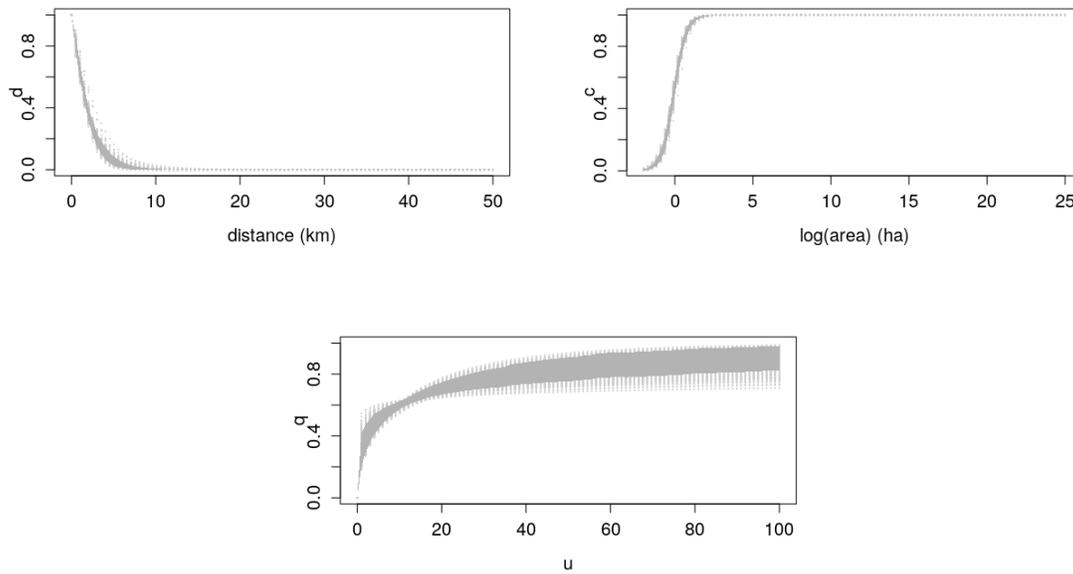


Figura 5.7: Modelo III: Cambio en la probabilidad para escoger un parche como función de la distancia (Arriba-Izquierda) y como función del área (Arriba-Derecha) a nivel poblacional. Aumento de la frecuencia del uso de memoria como función del número de sitios únicos visitados (Abajo). Todas las gráficas corresponden a ajuste a nivel poblacional.

Como mencionamos al inicio de esta subsección, los resultados para el ajuste de los parámetros que se comparten con el Modelo I no tienen variaciones importantes y por lo tanto tampoco la tienen las interpretaciones del aumento y disminución de la probabilidad como función del área y la distancia, respectivamente, como se muestra en la Fig.5.7-Arriba. La Fig.5.7-Abajo, muestra el crecimiento del parámetro de memoria  $q$  como función del número de sitios únicos visitados  $u$  para el individuo promedio (nivel poblacional). Las curvas mostradas aquí se obtuvieron de la misma manera que las curvas de área y distancia: tomando muestras aleatorias de la densidad marginal de los parámetros poblacionales  $\rho$  y  $\epsilon$  (curvas gris claro) y obteniendo el intervalo de confianza al 95 % (gris oscuro) generado por dichas curvas. Podemos notar que el uso de memoria, para este individuo promedio, crece relativamente rápido hacia 0.5 cuando los  $u$ s están entre cero y diez, luego de este “umbral”,  $q$  crece asintóticamente a uno hasta alcanzar la saturación. Esto tiene sentido hipotético (aún no hemos mostrado que este modelo sea adecuado para ajustar los datos observados) pues, a mayor número de sitios visitados, mayor cantidad de espacio explorado y es mucho más probable que el animal haya ya encontrado sus sitios de forrajeo preferidos, volviendo de manera muy frecuente a ellos.

En las Tablas 7.3 (nivel individual) y 7.8 (nivel jerárquico) de las secciones 7.1 y 7.2 del Apéndice, respectivamente, se pueden encontrar los estadísticos de interés para cada uno de los cuatro parámetros del Modelo III, para cada uno de los 21 individuos y el individuo promedio. Estos datos nos permitieron obtener los resultados ya expuestos en las Figs.5.6-5.7.

## Modelo IV

La Figura 5.8 muestra las medias de las marginales posteriores para los parámetros del Modelo IV. Los resultados para los parámetros compartidos con el Modelo I y II no tienen variaciones importantes a las ya discutidas y por lo tanto, la interpretación permanece tanto para el aumento de probabilidad con el área y la disminución de probabilidad con la distancia.(Fig.5.9-Arriba). Aquí se hace importante recalcar que los parámetros compartidos con el Modelo III que dictan el aumento de la frecuencia del uso de memoria  $q$  con el número de sitios únicos visitados, sí muestran una variación considerable. En particular, los parámetros  $\rho_j$  y  $\rho$ . Veamos, para el modelo anterior teníamos que el parámetro individual  $\rho_j$  a nivel jerárquico variaba entre seis y doce, pero ahora, para este Modelo IV tenemos que  $\rho_j$  varía entre veinticinco y veintisiete ( $25 < \rho < 27$ ). Esto se traduce directamente en un aumento de uso memoria mucho más lento, es decir, los animales necesitan un mayor número de sitios únicos visitados para alcanzar la saturación en  $q$ . Dicho de otra manera, mientras que para el Modelo III eran suficientes 10 uvs para que  $q$  aumentara hasta 0.5, en este Modelo IV se hacen necesarios alrededor de 30 uvs para que  $q$  alcance el mismo valor (Fig.5.9-Abajo). Este cambio puede deberse al nuevo ingrediente de decaimiento de memoria que implementamos en la dinámica de este último modelo.

Como acabamos de mencionar, en este Modelo IV, además del supuesto de que la frecuencia del uso de memoria crece y se satura de manera exponencial con el uvs, asumimos que los animales tienen un decaimiento de memoria, es decir, en la probabilidad de visitar un parche dado, el peso de cada visita pasada al parche decae exponencialmente con el tiempo transcurrido desde dicha visita. Así, aquellos parches que han sido visitados en un pasado lejano, tienen menos probabilidad de ser elegidos para ser visitados de nuevo que aquellos que han sido visitados en el pasado reciente.

La Fig.5.8-Abajo Izquierda, muestra las medias de las marginales posteriores para el parámetro  $\nu$  (tanto el parámetro individual  $\nu_j$  como el parámetro poblacional  $\nu$ ) que define la escala de la función exponencial que determina el decaimiento de memoria. El parámetro individual  $\nu_j$  casi no tiene fluctuaciones a nivel jerárquico (tampoco a nivel individual), manteniendo un valor cercano a once ( $\nu_j \approx 11$ ). En este caso para casi todos los individuos la diferencia entre el ajuste a nivel individual y jerárquico no es significativa pero cabe mencionar que la variabilidad de cada  $\nu_j$  a nivel individual (líneas negras) es importante, lo que podría hablarnos de una dificultad al momento de ajustar el valor del parámetro en la primera fase del algoritmo de jerarquización. Por otro lado, el parámetro individual  $\theta_j$  fluctúa (tanto a nivel individual como jerárquico) por debajo de 1, que corresponde al decaimiento exponencial simple. Entonces, la memoria decae mas lento que una exponencial, es decir como una exponencial estirada.. La diferencia entre los ajustes entre los niveles individual y jerárquico no es tan significativa y marcada como lo es para el parámetro  $\nu$ . Los valores medios de estos parámetros para el individuo promedio (nivel poblacional ‘pop ‘) son:  $\nu = 11.02$  y  $\theta = 0.30$ .

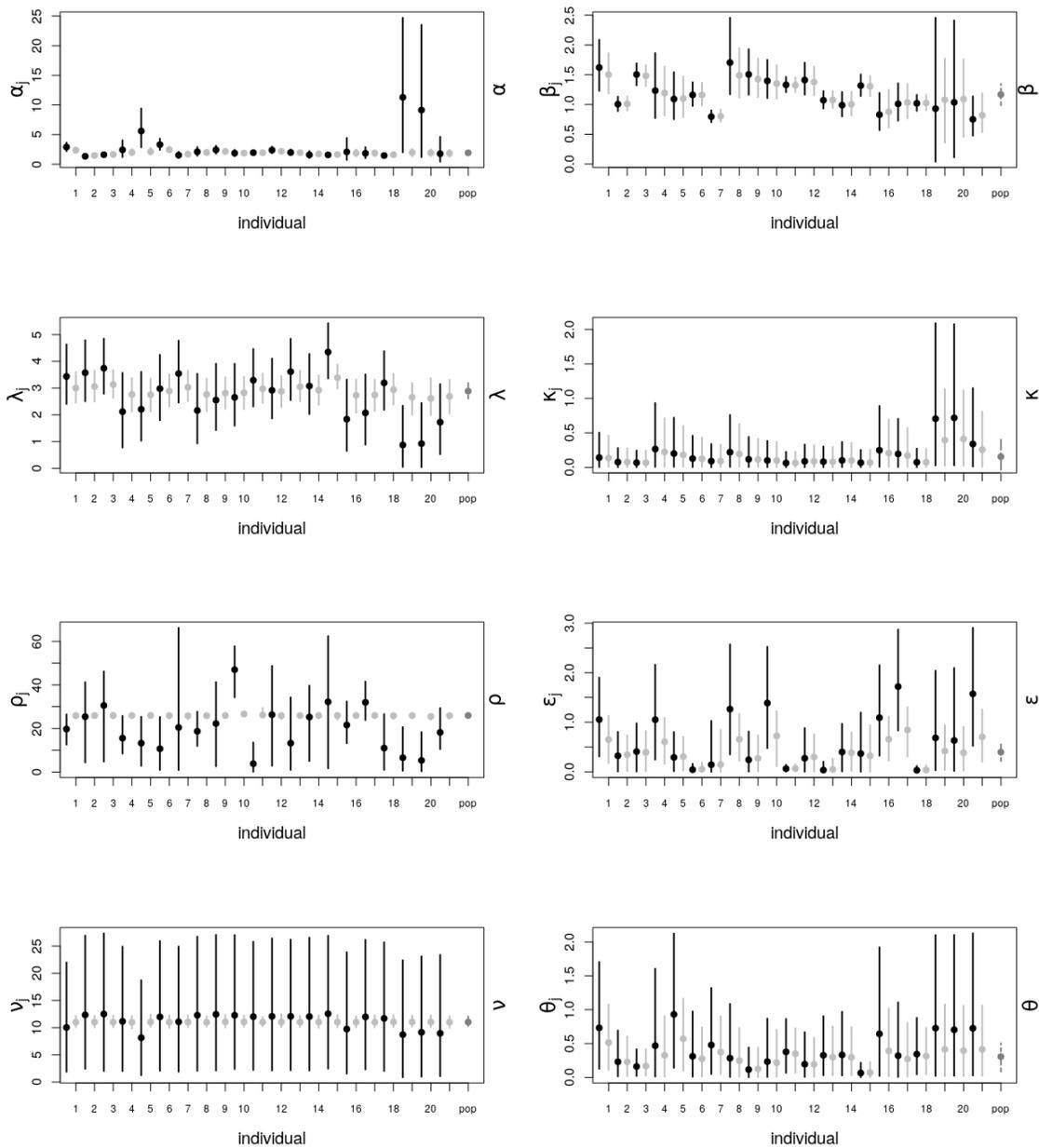


Figura 5.8: Medias de la posterior marginal (puntos) e intervalos de credibilidad al 95% (lineas verticales) para cada uno de los individuos (identificados por su ID en el eje X) y la población (denotada por 'pop') para cada uno de los parámetros del Modelo IV. En negro (a la izquierda de cada grupo) los resultados que corresponden a la primera fase del algoritmo de jerarquización. En gris claro (a la derecha de cada grupo) los resultados para la segunda fase del algoritmo de jerarquización. En gris oscuro se muestra el resultado de la segunda fase del algoritmo para el nivel poblacional.

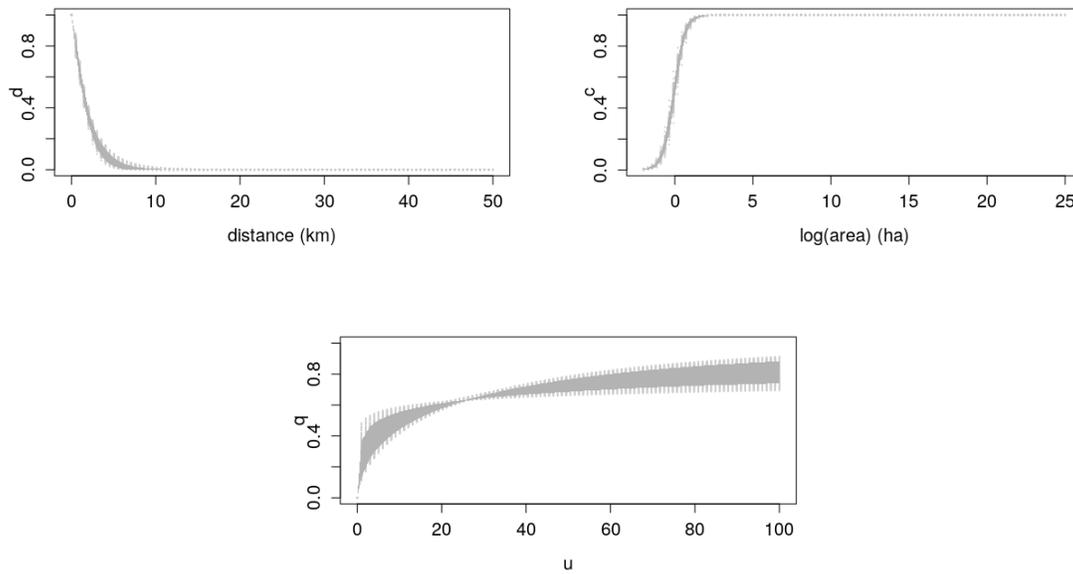


Figura 5.9: Modelo IV: Cambio en la probabilidad para escoger un parche como función de la distancia (Arriba-Izquierda) y el área (Arriba-Derecha). Aumento de la frecuencia del uso de memoria como función del número de sitios únicos visitados  $u$  (Abajo). Todas las gráficas corresponden al ajuste a nivel poblacional.

Así, a nivel poblacional hay un decaimiento lento de la memoria, lo cual se traduce en una capacidad del individuo promedio de recordar sitios visitados en un pasado no tan reciente. Recalcando, un valor de  $\nu$  por debajo de este promedio, lleva a un decaimiento de memoria mucho más rápido, mientras que un valor por encima de él lleva a un decaimiento más lento. También es importante notar que el valor del parámetro  $\theta$  está significativamente por debajo de 1, lo que nos lleva directamente a que el decaimiento de la memoria será mucho más lento, traduciéndose así en una gran capacidad de memoria de estos animales. Recordemos que el caso  $\theta = 0$  nos lleva al Modelo III, donde suponemos que el animal tiene una capacidad de memoria infinita. Estas interpretaciones hipotéticas (aún no hemos verificado que este modelo sea el mejor de entre los cuatro posibles) son válidas no sólo para el individuo promedio, sino también para cada uno de los individuos observados  $j$  a nivel jerárquico, pues para este caso en particular los valores de los parámetros individuales  $\nu_j$  y  $\theta_j$  son muy similares a los parámetros poblacionales  $\nu$  y  $\theta$  para todo  $j = 1 : 21$ .

En las Tablas 7.4 (nivel individual) y 7.9 (nivel jerárquico) de las secciones 7.1 y 7.2 del Apéndice, respectivamente, se pueden encontrar los estadísticos de interés para cada uno de los cuatro parámetros del Modelo IV, para cada uno de los 21 individuos y el individuo promedio. Estos datos nos permitieron obtener los resultados ya expuestos en las Figs.5.8-5.9. En estas tablas puede verse reflejado el ajuste dramático de los parámetros  $\rho$  y  $\nu$  a nivel jerárquico en comparación a los mismos a nivel individual. Como lo mencionamos antes, esto puede deberse, o bien, a la gran influencia de la información estadística obtenida de los otros

individuos, o a una pérdida de la misma al ejecutar el segundo paso del algoritmo de jerarquización. En ambos casos, la gran variabilidad que existe del ajuste de estos parámetros entre individuos a nivel individual lleva a que las estimaciones de  $\rho$  y  $\nu$  a nivel jerárquico tengan un gran nivel de autocorrelación (Tabla 7.9), lo que puede complicar la predicción para datos no observados.

Ahora bien, nos interesa ver cuál de todos estos cuatro modelos, a nivel jerárquico, describe mejor los datos observados y cuál de ellos tiene el mejor nivel de predicción. Para esto utilizamos las herramientas PPC y WAIC ya expuestas en secciones previas. Averiguar cuál de estos modelos es el mejor (tanto en términos descriptivos como predictivos), nos ayudará también a encontrar los mecanismos mínimos necesarios para caracterizar y explicar las trayectorias de estos 21 individuos. Los resultados se muestran a continuación.

## 5.12. Comparación entre modelos

En esta sección nos enfocaremos en probar el nivel descriptivo y predictivo de nuestros modelos por medio de las dos herramientas expuestas en las secciones anteriores; la Comprobación Predictiva Posterior (PPC), que nos servirá para aceptar o rechazar un modelo, y el Criterio de Información de Watanabe Akaike, que nos permitirá cuantificar el nivel de predictibilidad entre modelos y compararlos entre sí. Nos centraremos en el nivel jerárquico para la aplicación de estas dos medidas enfocándonos en el comportamiento individual de cada uno de los animales. Esto lo haremos pues para el PPC probaremos la capacidad de ajuste de los modelos utilizando el uvs, este estadístico no puede obtenerse de manera poblacional (para un individuo promedio) pues el uvs entre individuos es muy variable y es imposible obtener una curva promedio. Para ser consistentes, también calcularemos el WAIC de cada modelo para cada individuo a nivel jerárquico. De manera ilustrativa presentaremos los resultados para cuatro de los veintiún individuos que, a nuestro parecer, caracterizan los diversos comportamientos que se pueden encontrar en las trayectorias. Como nota al pie, en la sección 7.2 del Apéndice (en las subsecciones PPC y WAIC) se pueden encontrar los resultados a nivel jerárquico para los 21 animales y en la sección 7.1 de la misma Apéndice se exponen los resultados a nivel individual.

La Figura 5.10 muestra los resultados del PPC para los individuos 1,7,11 y 17. Las trayectorias de los animales fueron caracterizadas por el número de sitios únicos visitados (uvs) al tiempo  $t$  ( $\mathbf{u}$ ). Esto nos permite hacer mucho más sencilla la comparación entre las réplicas y las trayectorias, pues si consideramos sólo las posiciones visitadas por el animal reportadas en los datos, es muy poco probable que las réplicas sean parecidas a éstas y, por lo tanto, incomparables. Pero con esta caracterización de los sitios únicos visitados, tanto en las réplicas como en los datos reales, podemos tener comportamientos comparables. Las réplicas se calcularon a partir de correr el modelo con mil muestras de las marginales posteriores de los parámetros individuales  $p_j, j = 1 : 21$  a nivel jerárquico (cada muestra de la conjunta da una réplica). Una vez obtenidas las réplicas, calculamos el intervalo de credibilidad al 95% generado por dichas réplicas, y si los datos están contenidos, significativamente, dentro de éste intervalo diremos pues que aceptamos el modelo como un buen modelo descriptivo y

potencial modelo predictivo.

En la Fig.5.10 se muestran: en rojo, el número de sitios únicos visitados de la trayectoria observada, en gris claro las mil curvas réplica obtenidas al hacer mil muestras de la marginal posterior, y en gris oscuro el intervalo de credibilidad al 95 % generado por estas réplicas. Las gráficas se organizan como sigue: cada fila representa el PPC de los cuatro modelos para un sólo individuo y cada columna representa el PPC de los cuatro individuos para un sólo modelo. El PPC para el resto de los animales se puede ver en la Fig.7.2 de la sección 7.2 de este mismo escrito.

En la Fig.5.10 podemos ver que para los individuos 1,7 y 11 el Modelo I resulta adecuado sólo para aproximar los inicios de las trayectorias cuando el uvs crece relativamente rápido, es decir, cuando los animales se encuentran en fase de exploración. Algo contrario pasa para el animal 17; el Modelo I resulta adecuado para ajustar de manera satisfactoria casi toda la trayectoria. Esto nos dice que la trayectoria de este animal esta compuesta, en general, de pasos aleatorios y que, mientras los datos se recolectaron, éste se encontraba en fase de exploración.

El Modelo II muestra un mejor ajuste para todos los individuos, incluso para el individuo 17 donde ahora el PPC logra contener toda la trayectoria observada dentro de los intervalos de confianza generados por las réplicas. Recordemos que para este animal el valor medio del parámetro de uso de memoria era  $q \approx 0.25$ , el cual es bajo comparado con el resto de los individuos y refuerza nuestra hipótesis de que este particular individuo estuvo, mayoritariamente, explorando el espacio a su alrededor. Particularmente, en el caso del animal 1 este Modelo II resulta adecuado para el ajuste sólo de los últimos tiempos de la trayectoria cuando, podemos suponer que, el animal está ya en una fase de explotación de los sitios conocidos. Para los individuos 7 y 11 pasa algo muy similar a lo ya explicado para el animal 17, los valores medios de  $q$  para los animales 7 y 11 son 0.38 y 0.80, respectivamente. Así, para el animal 11 podemos decir, de forma cualitativa, que el animal encontró “rápidamente” un conjunto de buenos parches, se asentó en ellos y los explotó de manera regular, o bien, que adoptó desde un inicio una estrategia muy poco exploratoria.

El Modelo III parece ser más adecuado para el ajuste de la trayectoria del individuo 1, conteniéndola casi en su totalidad dentro del intervalo de credibilidad generado por las réplicas. Entonces, el individuo aumenta la frecuencia del uso de memoria conforme explora el ambiente. Para el animal 5, el Modelo IV tiene un ajuste tan bueno como el que proporciona el Modelo II pero este Modelo IV reduce la ventana de confiabilidad, lo que podría sugerir (cualitativamente, remarco) un mejor nivel de predicción. Para los animales 7 y 11 el Modelo III parece no ser adecuado en lo absoluto. De hecho, el ajuste proporcionado por ambos es peor que el que nos da el Modelo II. En el caso del animal 7, la trayectoria observada está por debajo de la ventana de credibilidad del PPC y la del animal 11 está por encima (lo mismo pasa para este individuo con el ajuste proporcionado por el Modelo IV). Esto es consistente con nuestras hipótesis de que la trayectoria del animal 7 es una trayectoria de exploración y la del individuo 11 de explotación de sitios conocidos. Por último, para el individuo 17 el Modelo IV ajusta bastante bien la trayectoria observada dentro de los rangos de credibilidad, lo cual puede significar que el decaimiento de memoria puede jugar un rol importante para

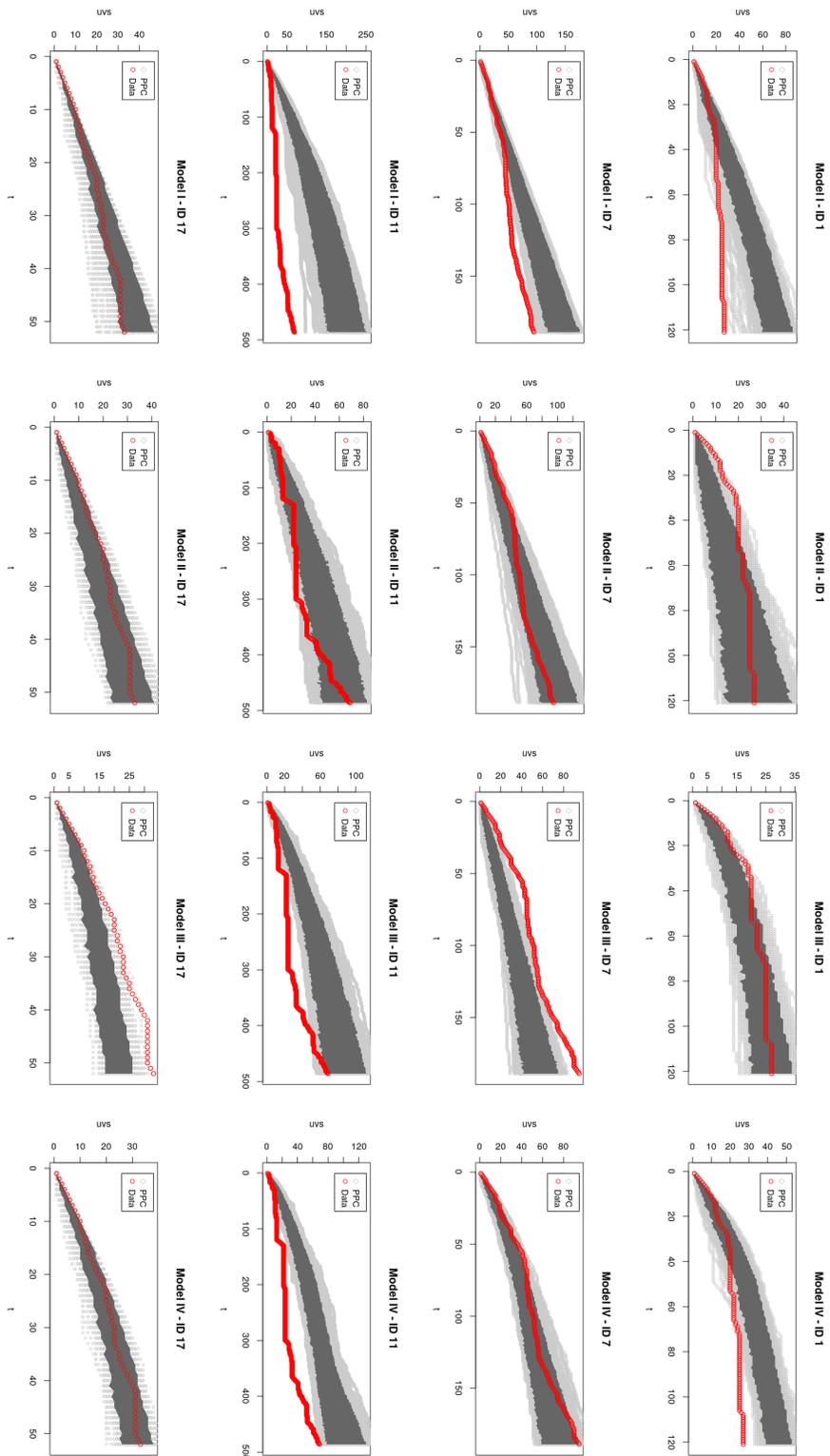


Figura 5.10: PPC de los modelos I-IV para los ciervos con ID 1,7,11 y 17. El número de sitios únicos visitados (*uvs*) como función del tiempo *t*. En gris claro las curvas provenientes de la réplica del PPC, en gris oscuro el intervalo de confianza al 95 % generado por dichas réplicas y en rojo los datos observados.

este animal en particular a la hora de escoger nuevos sitios para visitar. Contando el resto de los individuos (ver Tabla 7.2 de la sección 7.2 del Apéndice) podemos decir que el Modelo I no resulta adecuado para ninguno de los individuos, mientras que el Modelo II resulta adecuado para un total de 12 animales (los animales con ID 3,7,9-16,18 y 21) mientras que para los 9 animales restantes resultan adecuados los Modelos III y IV; 8 individuos para el Modelo III y sólo 1 (el animal 17) para el Modelo IV.

Ahora bien, como hemos ya mencionado antes, el PPC nos permite deducir de manera cualitativa como rechazar o aceptar un modelo, y, si varios resultan adecuados, cuál de ellos tiene el mejor desempeño predictivo. Pero sigue siendo sólo una herramienta cualitativa. En afán de reafirmar lo ya señalado por el PPC, calculamos el Criterio de Información de Watanabe Akaike para todos los individuos y todos los modelos a nivel jerárquico. Para esto utilizamos la herramienta `loo.waic` disponible en R para una muestra de likelihoods calculados desde la marginal posterior obtenida en la segunda etapa del algoritmo de jerarquización.

La Tabla 5.2 muestra las estimaciones puntuales para el Criterio de Información de Watanabe Akaike (WAIC) para todos los modelos y para cuatro individuos representativos de la población, la diferencia entre el menor WAIC y el resto de los WAIC que corresponden a un mismo individuo, se denota por  $\delta$ , se muestra también el número efectivo de parámetros  $P_W$ . En negritas resaltamos el mejor modelo para cada uno de los individuos, es decir, aquél modelo con el menor WAIC para ese animal. Para el individuo 1 el mejor resultado se obtiene para el Modelo III con un WAIC estimado de 769.74. Aquí es importante resaltar que el  $\delta$  con el resto de los modelos para este individuo es mayor a 2, por lo que podemos decir que hay una diferencia significativa entre el Modelo III y el resto <sup>7</sup>, con lo que dicho Modelo III resulta ser el mejor para este individuo, siendo el que mayor capacidad de predicción tiene y el que resulta más adecuados para el animal 1. Luego, para los individuos 7 y 11, es el Modelo II es el que reporta el menor WAIC y tiene con el resto de los modelos  $\delta$ 's significativas (es decir  $\delta > 2$ ), por lo que podemos decir que, a nivel jerárquico, este modelo describe de manera adecuada la trayectoria de estos individuo y tiene la mejor predictibilidad para datos no observados. Finalmente, para el individuo 17, el Modelo IV resulta el mejor de entre los cuatro modelos planteados, pero aquí la  $\delta$  entre el Modelo II y el Modelo IV es menor a 2, por lo que podemos decir que, con estos resultados, no podemos discernir entre ambos, los dos resultan igualmente adecuados para describir la trayectoria de este animal y ambos poseen el mismo nivel de predicción. Esto es consistente con el hecho de que el decaimiento de memoria para el Modelo IV es muy lento, dándole al animal una gran capacidad de memoria, una similar a la que se describe en el Modelo II. Estos resultados cuantitativos coinciden y complementan lo ya discutido e interpretado por los resultados del PPC. La Tabla 7.10 de la sección 7.2 del Apéndice muestra los resultados para el resto de los individuos. Esta tabla nos dice que el Modelo I no resulta adecuado para ninguno de los individuos a nivel jerárquico. El Modelo II es el mejor modelo para 12 de los 21 individuos de los que se poseen datos, mientras que el Modelo III lo es para 8 animales, y el Modelo IV sólo lo es para 1 individuo. Esto nos dice que para la mayoría de los individuos es suficiente considerar una frecuencia de uso de memoria

---

<sup>7</sup>Esta convención es generalmente aceptada por los estadísticos y ecólogos que utilizan modelos bayesianos [141].

constante (y alta) y con reforzamiento lineal a aquellos parches ya visitados. Estos simples ingredientes nos proporcionan un buen modelo descriptivo y con una buena capacidad de predicción. Ahondaremos más en estas conclusiones en el capítulo siguiente, discutiendo las implicaciones ecológicas de que sea justamente el Modelo II el mejor para la mayoría de los individuos observados.

ID	1			7		
Modelo	WAIC		$P_w$	WAIC		$P_w$
	Est	$\delta$	Est	Est	$\delta$	Est
Modelo I	916.51	146.77	12.06	1670.40	70.08	8.93
Modelo II	788.51	18.77	7.27	<b>1600.32</b>	<b>0.00</b>	<b>11.42</b>
Modelo III	<b>769.74</b>	<b>0.00</b>	<b>7.56</b>	1654.47	54.15	18.00
Modelo IV	800.53	30.79	10.56	1678.14	77.82	19.79
ID	11			17		
Modelo	WAIC		$P_w$	WAIC		$P_w$
	Est	$\delta$	Est	Est	$\delta$	Est
Modelo I	3406.23	677.88	6.71	499.80	35.39	27.65
Modelo II	<b>2728.35</b>	<b>0.00</b>	<b>7.34</b>	465.15	0.74	10.51
Modelo III	2754.21	25.86	6.56	473.24	8.83	10.20
Modelo IV	2867.84	139.49	17.22	<b>464.41</b>	<b>0.00</b>	<b>10.49</b>

Cuadro 5.2: WAIC para el log-likelihood puntual (considerando sólo un valor de la muestra completa del log-likelihood calculada por stan) para cada Modelo y cada animal a nivel poblacional. La tabla muestra las estimaciones puntuales (Est) para el criterio de información WAIC, el número efectivo de parámetros ( $P_w$ ).  $\delta$  será la diferencia entre el menor WAIC (en negritas) y los restantes WAIC para cada individuo. Por convención tomaremos que si el  $\delta$  entre dos (o más) modelos es menor que 2, no podemos decidir cuál de ellos es mejor

En el siguiente capítulo expondremos las conclusiones generales y particulares a las cuales hemos llegado con el trabajo presentado en los tres capítulos anteriores de este mismo escrito. Hasta aquí es importante recalcar que nuestro modelo mínimo (el modelo simple con reforzamiento lineal, memoria infinita y tasa de uso de memoria constante) expuesto en el capítulo 3 es capaz de capturar los mecanismos principales de movimiento para datos de animales reales, mecanismos tales como un uso heterogéneo del espacio (distinguiendo fases de movimiento de explotación y exploración) y creación de afinidad por sitios. Esta simplicidad nos proporciona una gran ventaja para implementar este modelo en otros conjuntos de datos.

# Capítulo 6

## Conclusiones

En este capítulo mostraremos las principales conclusiones de este trabajo. Para una mejor lectura, hemos decidido organizar las conclusiones en tres secciones cortas: en la primera sección expondremos las conclusiones del capítulo 3, en la segunda discutiremos el capítulo 4, y en la tercera sección extenderemos las conclusiones del capítulo 5.

### Un modelo mínimo para el movimiento animal

Los procesos de caminatas aleatorias sin memoria proporcionan, en ciertos casos, estrategias ventajosas para la búsqueda de recursos en ambientes no conocidos pero generalmente no retratan de manera adecuada muchas características esenciales del movimiento animal como lo son: el uso heterogéneo del espacio y la tendencia a visitar de manera preferencial ciertos sitios del ambiente como consecuencia de la adaptación por medio del aprendizaje. Varios estudios experimentales han mostrado que muchas especies animales poseen habilidades cognitivas sofisticadas, como la memoria espacial que les permite volver a sitios de interés que se encuentran fuera de su rango de visión y así adaptarse a su ambiente. Este nuevo descubrimiento ha llevado a un gran interés por incluir el uso de memoria en los modelos teóricos que se utilizan para caracterizar el movimiento animal y, en específico, la actividad de forrajeo. La mayoría de estos estudios teóricos que incluyen el nuevo ingrediente de la memoria en sus dinámicas son, en general, difíciles de resolver analíticamente y por lo tanto se hace complicado entender los efectos específicos que tiene el uso de memoria sobre la dinámica de movimiento y búsqueda.

En el capítulo 3 hemos presentado un modelo mínimo (inspirado en otros previamente estudiados) que presenta dos tipos de dinámicas: una gobernada por pasos aleatorios y otra que se define por medio de reubicaciones preferenciales a sitios previamente visitados, esta dinámica de reubicación ocurre con frecuencia  $q$  (uso de memoria). La regla de revisita preferencial es lineal, de tal suerte que aquellos sitios con más número de visitas tienen mayor probabilidad de ser visitados nuevamente. Este modelo considera un espacio heterogéneo que posee un sólo sitio de recurso permanente en el origen caracterizado por un peso  $\gamma$ . La gran ventaja de nuestro modelo mínimo es que es analíticamente soluble lo cual nos permitió encontrar de manera cuantitativa la tasa de uso de memoria que se hace necesaria para que

el caminante modelo se adapte a su ambiente por medio del aprendizaje reforzado. Dicha adaptación se revela por medio de la localización que emerge al rededor del sitio de recurso. Este fenómeno de localización se presenta de forma diferente dependiendo si el proceso Markoviano subyacente a la dinámica es o no transiente. Así, para una distribución de pasos que define un proceso recurrente, la localización (y por lo tanto la adaptación) se da a cualquier frecuencia de uso de memoria  $q$  que sea mayor que cero. Algo contrario ocurre cuando la distribución de pasos lleva a un proceso transiente, para este caso surge una transición de fase para el fenómeno de localización en el cual existe un valor crítico de uso de memoria  $q_c$  de tal suerte que para todos los valores anteriores a él, el caminante modelo es incapaz de localizarse al rededor del sitio de recurso (aún habiéndolo encontrado) y por tanto, es incapaz de aprender y adaptarse a su ambiente. En cambio, para los valores superiores a ese  $q_c$  la localización ocurre, es decir, existe un mínimo de uso de memoria para el cual el individuo es capaz de aprender y adaptarse. Para ambos casos, el caminante localizado se comporta, asintóticamente, como el del modelo de Evans y Majumdar que se reubica al origen a tasa constante  $r$ , sólo que la reubicación de nuestro caminante es una propiedad emergente de la dinámica, pues si bien el sitio de interés posee un peso superior al resto de los sitios en la rejilla, el aprendizaje y adaptación que emergen de las reubicaciones preferenciales no está explícita en nuestras ecuaciones, como sí es el caso del caminante de Evans y Majumdar. Por otro lado, probamos la robustez del fenómeno de localización agregando un nuevo parámetro  $\tau$  que define la probabilidad de que el sitio de recurso, luego de ser consumido, vuelva a su peso original  $\tau$ . Se asume pues que luego de cada visita, la heterogeneidad se vacía (peso igual a cero) y con tasa  $\tau$  se regenera. El fenómeno de localización tiene lugar a valores críticos de uso de memoria similares a la transición original. Aún y con recursos efímeros el caminante modelo sigue adaptándose a su ambiente.

De este modelo mínimo expusimos otro con una leve pero significativa modificación que da lugar a nueva interpretación ecológica, pues en vez de que el sitio de interés promueva la estadía en él, promueve los saltos aleatorios a sus primeros vecinos. Lo cual podría interpretarse como un sitio de atracción que merece una búsqueda local. Resulta interesante que al modificar las reglas de movimiento que el caminante sigue al estar en el origen, igual lleven a un fenómeno de localización/delocalización al rededor del mismo pero no de la misma manera que en el modelo mínimo pues, cabe recordar que, la localización se refiere a que las trayectorias no sean libres difusivas, sino confinadas a una región alrededor del sitio de interés. En el caso del modelo mínimo, la homogeneidad concentraba la mayor probabilidad de ocupación decayendo de forma exponencial (o logarítmica) hacia el resto de los sitios, para el caso de esta modificación son los sitios vecinos al origen los que abarcan la mayor cantidad de probabilidad de ocupación, lo cual tiene sentido, pues éste tiende a promover la visita aleatoria a primeros vecinos más que la relocalización, siendo la estadía en el origen completamente nula.

En otro contexto, los procesos presentados aquí, definidos por caminatas aleatorias, poseen ciertas analogías con la transición de localización de Anderson. Las similitudes entre las propiedades críticas de las caminatas aleatorias reforzadas en espacios discretos y la transición de Anderson se han estudiado ya en el pasado [26, 31], aunque de una forma bastante cualitativa. En el capítulo tres de este trabajo, hemos expuesto, a nuestro conocimiento, la primera conexión precisa entre estas dos clases de problemas, a través de un estudio analítico de los

diagramas de fase, puntos críticos y dimensiones críticas. La presencia de un sitio de impureza, un ingrediente clave responsable de la transición de fase, generalmente no se considera en la literatura sobre caminatas aleatorias reforzadas.

Con relación a las diversas líneas de investigación que este estudio puede abrir, un trabajo futuro que nos es de interés, y del cual presentamos ya un breve avance en la sección de resultados del modelo mínimo, es considerar un sitio de recurso que no está siempre presente. Esto puede ocurrir ya sea porque fue consumido por el caminante (caso ya considerado) o bien por cuestiones ecológicas específicas como temporada de frutos. Con esto nos interesaría averiguar si las condiciones de uso de memoria para que la localización se presente siguen siendo las mismas que las ya encontradas o si bien se modifica la tasa de memoria necesaria para la adaptación. Otro trabajo futuro de interés es considerar cómo es que el caminante comparte su tiempo de consumo entre dos sitios que se encuentran o bien activos al mismo tiempo o bien a periodos intercalados. Para el caso de un sólo caminante explorando el espacio queda también pendiente analizar las consecuencias que el un decaimiento en la memoria pueden tener sobre el fenómeno del aprendizaje individual y el éxito en la adaptación.

En el contexto del ámbito ecológico, los dos modelos expuesto en el capítulo tres pueden contribuir de manera importante para entender los mecanismos envueltos en el proceso de aprendizaje espacial y adaptación en el forrajeo animal. Más aún, nuestros resultados proveen soporte y pruebas matemáticas a la hipótesis presentada tiempo atrás en la que la memoria juega un rol de suma importancia durante la formación de ámbitos hogareños [18, 83, 152]. Esto ya que, de forma análoga al desarrollo del uso de memoria, existe un un peso crítico  $\gamma_c$  en el que el caminante modelo construye un NESS, es decir, una distribución estacionara del uso del espacio comparable con los ámbitos hogareños en el contexto del forrajeo animal. El comportamiento difusivo que tiene lugar debajo de dicho peso crítico puede tener ventajas de exploración hacia otras regiones del espacio donde muchos otros sitios de recursos pueden ser encontrados. Así pues, interesa estudiar los sistemas con múltiples sitios de recursos.

## Un enjambre con aprendizaje colectivo

En el capítulo cuatro presentamos un modelo basado en agentes que están conectados por una red de transferencia de información y siguen reglas de movimiento similares a las presentadas en el capítulo tres de este mismo escrito. La dinámica general es: En un espacio discreto cuyos sitios pueden o no ser sitios de recurso con pesos variables, los agentes además de poseer la habilidad del uso de memoria para reubicarse de manera preferencial a sitios previamente visitados, también tiene la posibilidad de “comunicarse” con aquellos otros agentes con los que estén conectados dentro de la red de transferencia de información, de tal suerte que cuando un individuo elige reubicarse, puede elegir entre usar su propia experiencia o usar la experiencia de alguno de sus vecinos en la red. Uno de los principales resultados es que, siguiendo una dinámica de movimiento sencilla, una fracción grande de los miembros del grupo es capaz de agregarse al rededor del mejor sitio de recursos en un ambiente desordenado con múltiples sitios de interés. En el contexto ya discutido en el capítulo cuatro, este fenómeno

representa una manifestación de aprendizaje colectivo.

Una propiedad característica del aprendizaje colectivo revelada por las simulaciones del capítulo cuatro para una dinámica a primeros vecinos, es que los individuos que intercambian información tiene una gran probabilidad de ocupar el mejor sitio de recurso dentro de un conjunto grande de sitios disponibles a visitar. En contraste, cuando los agentes no se comunican, dicha localización selectiva no tiene lugar y cada uno de los individuos aprende sólo de su ambiente local. Así, tanto la memoria individual como la transferencia de información son necesarios para un forrajeo grupal exitoso vía la agregación al rededor de los mejores sitios de recurso. Algo similar sucede cuando la dinámica pasa de ser a primeros vecinos a vuelos de Lévy, aquí la comunicación entre los miembros del enjambre mejora de manera significativa (a casi más del doble) la localización selectiva, es decir, muchos más individuos son capaces de localizarse al rededor del mejor sitio de recursos. Es importante recordar que para el caso de vuelos Lévy el enjambre sin comunicación presenta también una leve localización selectiva pero no tan marcada como cuando los agentes pueden intercambiar información de posible interés. Esto tiene sentido con lo ya mostrado por varios estudios teóricos y experimentales en los que se muestra que los vuelos de Lévy son la estrategia de forrajeo óptima en un ambiente desordenado. Así, es preferible siempre un enjambre que tenga ambas habilidades en cada uno de sus individuos para considerar que el forrajeo grupal ha sido exitoso. En el capítulo cuatro hemos presentado uno de los primeros estudios (a nuestro saber) de caminatas aleatorias reforzadas en un espacio heterogéneo con múltiples heterogeneidades que presenta un fenómeno de agregación como consecuencia del aprendizaje colectivo.

Nuestro estudio remarca la importancia de la topología de la red de transferencia de información en la habilidad del grupo para aprender de manera colectiva. Cuando la red de interacción es una red del tipo Erdős-Renyi (ER), un número promedio relativamente modesto de vecinos por individuo (seis o siete) es suficiente para desencadenar la agregación colectiva al rededor del mejor sitio de recursos. Luego de este umbral, si la conectividad promedia crece, esto no se traduce en una mayor agregación/cohesión. Este número característico de número de conexiones es, sin embargo, mucho mayor al umbral de percolación de las redes ER que se alcanza a una conectividad promedia de uno [117]. Esto significa que las redes que son eficientes para el aprendizaje colectivo no sólo poseen un componente gigante sino que también se encuentran densamente conectadas, tal como se ha observado en muchos grupos de primates [153]. En las redes sociales de los monos araña, en particular, la suma de los índices de asociación de un nodo (peso de un nodo) varía entre cero y siete dependiendo del individuo [116]. Valores al rededor de cinco se han reportado en otros animales como los delfines [120] y las jirafas [154]. También encontramos que la presencia de individuos altamente conectados en una red aleatoria libre de escala resulta importante para el aprendizaje colectivo. Esto es consistente con los resultados empíricos encontrados para los monos araña [68] en los que se expone que los individuos más centrales en la red lideran el movimiento colectivo y son seguidos por el grupo hacia los árboles con recursos disponibles que conocen.

En este mismo capítulo cuatro realizamos una leve modificación a la regla de visita preferencial pues, hasta ahora, ésta se traducía en una capacidad de memoria infinita, lo cual está bastante alejado de la realidad. Supusimos pues que la memoria de los individuos del

enjambre decae conforme una ley de potencia con exponente  $\beta$  de tal suerte que los sitios visitados en un pasado lejano tienden a ser olvidados y tienen mucho menos probabilidad de volver a ser visitados que aquellos sitios visitados en el pasado cercano. Resulta que esta capacidad de olvidar, le da al enjambre una mucho mejor capacidad de localización selectiva y aumenta el éxito del forrajeo grupal sólo cuando el decaimiento en la memoria no es tan rápido ( $\beta \approx 2$ ) como para que los sitios visitados en el pasado queden renegados al total olvido, ni tampoco tan lento ( $\beta \approx 0$ ) como para que sitios malos visitados en el pasado sigan teniendo un impacto importante en las decisiones del presente.

Nuestros resultados nos permiten hacer algunas predicciones para futuros estudios sobre aprendizaje colectivo en animales. Primero, los grupos de animales con la habilidad de recordar las posiciones de los sitios de recursos forrajearán de manera más eficiente cuando pueden intercambiar información entre ellos sobre dichos sitios. Esto puede probarse manipulando u obteniendo las posiciones de nuevos sitios y monitoreando la transferencia de información entre individuos como en [68]. Segundo, el número de vecinos para interactuar, así como la configuración de la red de transferencia de información pueden ser importantes para que el aprendizaje colectivo emerja. Esto podría probarse experimentalmente manipulando el número de compañeros (vecinos) a los que cada individuo tiene acceso y evaluando si el mínimo número de conexiones mencionado arriba es relevante en las redes reales. Dando un pequeño paso en este sentido fue que nos adentramos en los modelos bayesianos para estudiar las trayectorias de 21 ciervos canadienses, si bien los modelos presentados no consideran el intercambio de información real entre individuos, la aproximación jerárquica proporciona un camino para el estudio poblacional.

Por otro lado, este estudio abre un parteaguas importante hacia trabajos futuros y nuevas perspectivas en las aplicaciones. Por ejemplo, nos interesa estudiar los mecanismos específicos de uso de memoria y comunicación para que el aprendizaje colectivo emerja en un ambiente de parches consumibles y dinámicos. Luego, aplicar este algoritmo simple de búsqueda para redes de transferencia de información observadas en campo, como redes de interacción social o redes de jerarquía en sociedades animales. Nos interesa igualmente la posibilidad de estudiar fenómenos de competencia cuando dos especies distintas luchan por consumir los recursos de un mismo ambiente, o bien, relaciones de mutualismo en las que dos especies diferentes cooperan para explotar recursos. Este algoritmo no sólo puede usarse con fines a aplicaciones ecológicas sino que puede adaptarse también como un algoritmo de búsqueda computacional para aproximar problemas de optimización en altas dimensiones. En este último sentido realizamos algunos trabajos de prueba obteniendo resultados prometedores.

## Modelos bayesianos para el forrajeo de los ciervos canadienses

Como hemos mencionado ya a lo largo, el uso de memoria ha sido muy bien documentado y reportado en muchas especies animales y es crítico para entender el movimiento animal y, en particular, el fenómeno de forrajeo. El desarrollo teórico al rededor de estos dos fenómenos es profundo y extensivo, aquí mismo hemos ya discutido tres aproximaciones que, esperamos,

aporten al entendimiento de ambos. Recientemente el aumento en la disponibilidad de fuentes de datos de telemetría ha promovido el desarrollo de nuevas técnicas para entender los mecanismos subyacentes en fenómenos ecológicos tales como uso del espacio o formación de ámbitos hogareños.

En el capítulo cinco hemos presentado cuatro modelos bayesianos simples para ajustar un conjunto de datos observados de 21 ciervos canadienses recolectados en 2000-2002 en Alberta, Canadá. Los datos aquí analizados son de gran interés pues estos ciervos fueron reintroducidos en un ambiente desconocido, por lo tanto constituyen un muy buen sistema para comparar con modelos de caminatas aleatorias que a  $t = 0$  no tienen un pasado. El ajuste de los cuatro distintos modelos fue hecho, a un primer paso y a nivel individual, utilizando el MCMC disponible en RStan. A nivel jerárquico implementamos el algoritmo de dos pasos propuesto en [73]. Los principales resultados reportados aquí se centran en el nivel jerárquico, el cuál nos proporciona la información necesaria para hacer inferencia a nivel poblacional, dándonos la oportunidad de entender cómo el grupo de ciervos responde a ciertas características ambientales y los mecanismos que utilizaron para trazar la trayectoria observada. Para comparar y evaluar estos cuatro modelos utilizamos el PPC como herramienta cualitativa y el WAIC como herramienta cuantitativa, los resultados dados por ambas son consistentes.

Recordando un poco y de manera rápida: el primer modelo, aleatorio en esencia, no proporciona una buena descripción para ninguno de los individuos a nivel jerárquico. Esto nos dice que, incluso cuando estos animales fueron reinsertados en un nuevo ambiente, sus trayectorias están lejos de ser descritas y predichas por un modelo basado en un proceso de caminata aleatoria. El segundo modelo, que considera una capacidad de memoria infinita y un reforzamiento lineal para las visitas, ajusta y predice (un buen PPC y el menor WAIC) de manera satisfactoria la mayoría de las trayectorias (12 de 21). Este segundo modelo es el más parecido a nuestro modelo mínimo presentado en el capítulo 3 de este mismo escrito. Este pues es un modelo simple que puede ser fácilmente implementado a nuevos datos. Por último, el tercer y cuarto modelos, que consideran un uso de memoria dinámico y un decaimiento en la memoria (respectivamente), si bien resultan adecuados (un buen PPC) parecen ser demasiado sofisticados (un aumento en el WAIC) ajustando sólo de manera adecuada 9 de los 21 animales disponibles (ocho para el tercer modelo y sólo uno para el cuarto).

Como acabamos de mencionar, el Modelo II resulta ser el mejor para la mayoría de los individuos a nivel jerárquico, como ya se probó en el capítulo 5 y como se refleja en los anexos de la sección 7.2. Este modelo tiene, intrínsecamente, dos tipos de movimiento: Uno que es gobernado, básicamente, por decisiones aleatorias en las cuales el animal escoge un nuevo parche de visita solamente considerando la distancia entre éste y su posición actual y el área del mismo. El otro tipo de movimiento, que está inspirado en la dinámica que trabajamos en los capítulos 3 y 4, y que ocurre con probabilidad  $q$  (frecuencia del uso de memoria), es aquél en el que el animal considera su experiencia previa de tal suerte que el animal elige uno de los sitios previamente visitados para reubicarse con probabilidad proporcional al tiempo que ha gastado en cada uno de ellos, en este tipo de movimiento igual se considera la distancia y el área para la elección de este parche de visita. Los resultados muestran que estas dos simples suposiciones son suficientes para describir y predecir las trayectorias de doce ciervos. Aquí se

hace importante notar que estos doce animales (y los otros ocho que ajustan el Modelo III y IV) tienen (o alcanzan) un valor muy alto de  $q$ , al rededor de 0.5. Esto significa que estos ciervos, en particular, tienen un intensivo uso de memoria y la exploración del ambiente resulta del mismo orden que la explotación que hacen del mismo. En un principio esto resulta sorprendente y un poco atípico considerando los resultados que han sido reportados para otras especies como los monos capuchinos, en [85] el valor  $q$  encontrado para esta especie de monos y que describe mejor las trayectorias de dichos animales, sobre todo el rango temporal, es de apenas 0.12. En este modelo propuesto por Boyer y Solis-Salas, el ambiente fue modelado como una rejilla regular discreta en la cual cada uno de los puntos es un sitio a visitar. El alto uso de memoria reportado por nosotros puede ser también interpretado como un uso limitado del espacio que puede resultar riesgoso, es decir, nuestros ciervos prefieren regresar a parches seguros y conocidos en los cuales saben que el riesgo de depredación es bajo o el ambiente es seguro en otros aspectos (menor riesgo de caída o ahogamiento). Esto resulta consistente con los resultados principales expuestos en [147] por Frair *et al.* en los cuales se reporta que los ciervos generan afinidad rápidamente por ciertos parches y no exploran más allá de éstos, reduciendo así su tasa de mortalidad y su riesgo de depredación. Resulta importante recalcar que al no resultar adecuado el modelo que considera un decaimiento en la capacidad de memoria nos dice que, la memoria prácticamente no decae durante el período de recolección de los datos (un año). Es decir, esto parece indicar que estos animales recuerdan sitios visitados por un año a lo menos, lo cual podría traducirse (con un poco de abuso) en “buena memoria”.

Nuestros modelos, incluso siendo simples y relativamente sencillos, son capaces de capturar los mecanismos principales del movimiento de los animales observados, ayudándonos a entender las causas subyacentes para las mismas (como un alto uso de memoria, una rápida afinidad por ciertos parches y una excelente habilidad de memoria de largo alcance). Fuimos capaces de lograr esto con un número mínimo de parámetros y con una dinámica muy simple. Dicha simplicidad resulta tremendamente ventajosa para aplicar estos modelos a otros conjuntos de datos para otras especies animales.

Así pues en estos cuatro años de investigación hemos logrado proponer y desarrollar una serie de modelos simples en espacios heterogéneos que incluyen la memoria de una forma intuitiva y sencilla, aportando así evidencia matemática, computacional y estadística de la importancia que tiene el uso de memoria en ciertas especies animales, con implicaciones en fenómenos ecológicos tales como el uso heterogéneo del espacio, formación de ámbitos hogareños, aprendizaje, adaptación y técnicas de forrajeo exitosas.



# Capítulo 7

## Apéndices

En este capítulo exponemos todos los resultados de las estimaciones de los parámetros de cada uno de los modelos para cada uno de los 21 animales (a nivel individual [7.1] y nivel poblacional [7.2]). Se consideran estimaciones tales como la media, la desviación estándar (SD), cuantiles al 2.5 % y 97.5 %, el número efectivo de muestras ( $n_{\text{eff}}$ )<sup>1</sup> y el estadístico de orden Rhat<sup>2</sup>. Se muestra también el resultado del cálculo del WAIC para todos los modelos y cada uno de los individuos. Finalmente se exponen las gráficas del PPC, también, para cada modelo y cada uno de los animales. Estos resultados ayudan a entender de manera mucho más completa lo mostrado en el capítulo 5 de esta tesis.

### 7.1. Nivel individual

Las estimaciones y cálculos mostrados a continuación corresponden al ajuste para cada uno de los individuos hecho de manera independiente, es decir, los resultados obtenidos al ejecutar el primer paso del algoritmo de jerarquización explicado en el capítulo 5.

---

<sup>1</sup>Para una serie de tiempo  $x$  de longitud  $N$ , el error estándar de la media es la raíz cuadrada de  $\text{var}(x)/n$  donde  $n$  es el tamaño efectivo de la muestra.  $n = N$  sólo cuando no hay autocorrelación.

<sup>2</sup>El grado de convergencia de un MCMC puede ser estimado utilizando el estadístico de convergencia de Gelman-Rubin, Rhat, basado en la estabilidad de los resultados entre y dentro de  $m$  cadenas de la misma longitud  $n$ . Valores cercanos a uno indican convergencia hacia la distribución posterior. Mientras que valores mayores a 1.1 indican una convergencia inadecuada o pobre.

## Modelo I

ID	Parámetro	Media	SD	2.5 %	97.5 %	n_eff	Rhat
1	$\alpha$	1.42	0.23	1.00	1.88	3126	1.00
	$\beta$	1.13	0.12	0.91	1.38	3184	1.00
	$\lambda$	4.25	0.54	3.27	5.37	5076	1.00
	$\kappa$	0.07	0.07	0.00	0.26	6234	1.00
2	$\alpha$	0.72	0.08	0.56	0.86	3703	1.00
	$\beta$	0.87	0.05	0.78	0.96	3773	1.00
	$\lambda$	4.41	0.55	3.41	5.55	4850	1.00
	$\kappa$	0.04	0.04	0.00	0.16	6174	1.00
3	$\alpha$	1.13	0.09	0.97	1.30	4190	1.00
	$\beta$	1.28	0.07	1.15	1.43	4317	1.00
	$\lambda$	4.43	0.52	3.50	5.50	5274	1.00
	$\kappa$	0.03	0.03	0.00	0.12	6896	1.00
4	$\alpha$	1.32	0.42	0.62	2.28	4185	1.00
	$\beta$	0.98	0.18	0.67	1.38	3976	1.00
	$\lambda$	2.58	0.65	1.41	4.00	5322	1.00
	$\kappa$	0.17	0.16	0.00	0.57	5572	1.00
5	$\alpha$	1.17	0.39	0.54	2.06	4025	1.00
	$\beta$	0.71	0.09	0.55	0.91	4044	1.00
	$\lambda$	2.88	0.63	1.70	4.16	5636	1.00
	$\kappa$	0.11	0.11	0.00	0.40	6228	1.00
6	$\alpha$	1.31	0.23	0.89	1.80	4159	1.00
	$\beta$	0.86	0.07	0.73	1.01	4096	1.00
	$\lambda$	3.70	0.57	2.66	4.87	5198	1.00
	$\kappa$	0.06	0.06	0.00	0.22	6679	1.00
7	$\alpha$	0.57	0.10	0.40	0.78	3580	1.00
	$\beta$	0.64	0.03	0.58	0.72	3593	1.00
	$\lambda$	3.89	0.55	2.89	5.01	5697	1.00
	$\kappa$	0.06	0.06	0.00	0.21	6320	1.00
8	$\alpha$	1.41	0.25	0.93	1.93	3958	1.00
	$\beta$	1.40	0.21	1.03	1.86	3925	1.00
	$\lambda$	2.57	0.64	1.37	3.91	4790	1.00
	$\kappa$	0.16	0.16	0.00	0.58	5285	1.00
9	$\alpha$	1.29	0.24	0.87	1.79	3544	1.00
	$\beta$	1.07	0.12	0.86	1.32	3539	1.00
	$\lambda$	2.97	0.61	1.84	4.23	5064	1.00
	$\kappa$	0.07	0.07	0.00	0.27	5528	1.00
10	$\alpha$	1.38	0.22	0.98	1.82	4109	1.00
	$\beta$	1.20	0.13	0.96	1.47	4141	1.00
	$\lambda$	2.90	0.59	1.83	4.10	5302	1.00
	$\kappa$	0.08	0.08	0.00	0.30	6192	1.00

ID	Parámetro	Media	SD	2.5 %	97.5 %	n_eff	Rhat
11	$\alpha$	0.82	0.05	0.71	0.92	3916	1.00
	$\beta$	1.03	0.04	0.95	1.10	3962	1.00
	$\lambda$	4.09	0.50	3.20	5.15	4653	1.00
	$\kappa$	0.03	0.03	0.00	0.10	6565	1.00
12	$\alpha$	1.17	0.18	0.86	1.55	4037	1.00
	$\beta$	1.01	0.08	0.86	1.18	4041	1.00
	$\lambda$	3.41	0.55	2.41	4.57	5695	1.00
	$\kappa$	0.05	0.05	0.00	0.19	6553	1.00
13	$\alpha$	0.97	0.12	0.88	1.21	3658	1.00
	$\beta$	0.87	0.05	0.84	0.98	3615	1.00
	$\lambda$	4.31	0.56	3.92	5.48	5450	1.00
	$\kappa$	0.04	0.04	0.01	0.16	5753	1.00
14	$\alpha$	0.75	0.16	0.47	1.09	2905	1.00
	$\beta$	0.80	0.07	0.66	0.96	2786	1.00
	$\lambda$	0.54	0.56	2.54	4.73	4602	1.00
	$\kappa$	0.06	0.06	0.00	0.22	5066	1.00
15	$\alpha$	1.18	0.13	0.94	1.44	3173	1.00
	$\beta$	1.14	0.08	0.99	1.32	3210	1.00
	$\lambda$	4.81	0.52	3.86	5.88	5195	1.00
	$\kappa$	0.03	0.03	0.00	0.11	6492	1.00
16	$\alpha$	0.83	0.36	0.32	1.66	4077	1.00
	$\beta$	0.67	0.10	0.50	0.88	3916	1.00
	$\lambda$	2.10	0.67	0.89	3.54	3614	1.00
	$\kappa$	0.20	0.19	0.01	0.70	6206	1.00
17	$\alpha$	1.26	0.41	0.60	2.17	3642	1.00
	$\beta$	0.88	0.14	0.63	1.18	3638	1.00
	$\lambda$	2.25	0.64	1.12	3.59	4704	1.00
	$\kappa$	0.15	0.15	0.00	0.54	5938	1.00
18	$\alpha$	0.65	0.08	0.51	0.81	4422	1.00
	$\beta$	0.84	0.04	0.75	0.93	4577	1.00
	$\lambda$	3.66	0.52	2.73	4.78	5168	1.00
	$\kappa$	0.04	0.04	0.00	0.06	6759	1.00
19	$\alpha$	9.92	6.17	0.83	23.95	3747	1.00
	$\beta$	0.72	0.62	0.02	2.23	3627	1.00
	$\lambda$	0.97	0.63	0.09	2.46	4309	1.00
	$\kappa$	0.59	0.51	0.02	1.87	5222	1.00

ID	Parámetro	Media	SD	2.5 %	97.5 %	n_eff	Rhat
20	$\alpha$	5.72	5.59	0.54	20.37	2618	1.00
	$\beta$	1.04	0.67	0.08	2.57	2826	1.00
	$\lambda$	1.25	0.78	0.13	2.74	3722	1.00
	$\kappa$	0.57	0.49	0.02	1.83	4943	1.00
21	$\alpha$	0.99	0.61	0.23	2.62	3536	1.00
	$\beta$	0.67	0.14	0.44	0.98	3399	1.00
	$\lambda$	1.90	0.66	0.73	3.28	4516	1.00
	$\kappa$	0.31	0.29	0.01	1.09	5756	1.00

Cuadro 7.1: Estimaciones para todos los parámetros del Modelo I para cada uno de los animales. Estos resultados corresponden a aquellos para el primer paso del algoritmo de jerarquización. Nivel individual.

## Modelo II

ID	Parámetro	Media	SD	2.5 %	97.5 %	n_eff	Rhat
1	$\alpha$	2.56	0.35	1.92	3.27	4193	1.00
	$\beta$	1.45	0.19	1.11	1.87	4228	1.00
	$\lambda$	3.10	0.58	2.05	4.32	6145	1.00
	$\kappa$	0.15	0.15	0.00	0.55	7500	1.00
	$q$	0.67	0.05	0.56	0.77	5902	1.00
2	$\alpha$	1.50	0.18	1.18	1.87	4177	1.00
	$\beta$	1.01	0.06	0.89	1.14	4301	1.00
	$\lambda$	3.27	0.61	2.18	4.50	7123	1.00
	$\kappa$	0.09	0.09	0.00	0.34	7500	1.00
	$q$	0.68	0.04	0.60	0.75	6213	1.00
3	$\alpha$	1.69	0.12	1.46	1.94	3977	1.00
	$\beta$	1.50	0.10	1.32	1.70	3981	1.00
	$\lambda$	3.39	0.56	2.38	4.56	6999	1.00
	$\kappa$	0.09	0.09	0.00	0.33	7500	1.00
4	$\alpha$	2.58	0.71	1.36	4.14	4319	1.00
	$\beta$	1.28	0.29	0.81	1.92	4300	1.00
	$\lambda$	2.06	0.70	0.78	3.50	4954	1.00
	$\kappa$	0.26	0.25	0.01	0.89	7500	1.00
	$q$	0.48	0.10	0.29	0.67	5592	1.00
5	$\alpha$	3.78	1.09	2.06	6.32	4049	1.00
	$\beta$	0.95	0.15	0.69	1.28	3619	1.00
	$\lambda$	2.18	0.65	1.04	3.54	6943	1.00
	$\kappa$	0.20	0.19	0.00	0.71	7500	1.00
	$q$	0.57	0.07	0.43	0.69	6516	1.00
6	$\alpha$	3.54	0.52	2.60	4.67	3821	1.00
	$\beta$	1.16	0.11	0.97	1.38	3894	1.00
	$\lambda$	2.83	0.63	1.68	4.15	6933	1.00
	$\kappa$	0.13	0.13	0.00	0.47	7500	1.00
	$q$	0.63	0.05	0.54	0.66	6211	1.00
7	$\alpha$	1.38	0.23	0.89	1.79	4482	1.00
	$\beta$	0.76	0.05	0.67	0.86	4781	1.00
	$\lambda$	3.51	0.56	2.49	4.65	6664	1.00
	$\kappa$	0.08	0.08	0.00	0.30	7500	1.00
	$q$	0.38	0.05	0.30	0.47	6346	1.00
8	$\alpha$	2.18	0.40	1.44	3.01	4229	1.00
	$\beta$	1.72	0.34	1.16	2.49	4152	1.00
	$\lambda$	2.09	0.69	0.83	3.46	4699	1.00
	$\kappa$	0.23	0.22	0.01	0.83	7500	1.00
	$q$	0.48	0.10	0.29	0.66	4938	1.00

ID	Parámetro	Media	SD	2.5 %	97.5 %	n_eff	Rhat
9	$\alpha$	2.34	0.34	1.69	3.05	4502	1.00
	$\beta$	1.45	0.19	1.12	1.87	4489	1.00
	$\lambda$	2.54	0.64	1.34	3.84	5820	1.00
	$\kappa$	0.11	0.11	0.00	0.40	7500	1.00
	$q$	0.42	0.06	0.31	0.54	6405	1.00
10	$\alpha$	1.82	0.26	1.35	2.35	4119	1.00
	$\beta$	1.38	0.16	1.10	1.72	4026	1.00
	$\lambda$	2.66	0.60	1.61	3.91	5645	1.00
	$\kappa$	0.10	0.10	0.00	0.37	7500	1.00
	$q$	0.27	0.07	0.14	0.40	5914	1.00
11	$\alpha$	2.07	0.16	1.78	2.40	4473	1.00
	$\beta$	1.31	0.07	1.18	1.45	4565	1.00
	$\lambda$	2.93	0.59	1.85	4.17	6577	1.00
	$\kappa$	0.08	0.08	0.00	0.31	7500	1.00
	$q$	0.80	0.02	0.76	0.84	6445	1.00
12	$\alpha$	2.42	0.29	1.87	3.04	4414	1.00
	$\beta$	1.41	0.14	1.16	1.70	4424	1.00
	$\lambda$	2.91	0.60	1.83	4.16	6144	1.00
	$\kappa$	0.09	0.09	0.00	0.32	7500	1.00
	$q$	0.49	0.05	0.40	0.59	5862	1.00
13	$\alpha$	1.88	0.21	1.50	2.30	4951	1.00
	$\beta$	1.07	0.07	0.94	1.21	5013	1.00
	$\lambda$	3.44	0.61	2.30	4.70	5987	1.00
	$\kappa$	0.08	0.08	0.00	0.31	7500	1.00
	$q$	0.54	0.04	0.46	0.63	5781	1.00
14	$\alpha$	1.59	0.29	1.09	2.21	4595	1.00
	$\beta$	0.98	0.10	0.80	1.20	4560	1.00
	$\lambda$	2.90	0.59	1.83	4.16	6051	1.00
	$\kappa$	0.10	0.10	0.00	0.39	7500	1.00
	$q$	0.55	0.05	0.44	0.65	6290	1.00
15	$\alpha$	1.61	0.14	1.35	1.90	4688	1.00
	$\beta$	1.30	0.09	1.13	1.49	4729	1.00
	$\lambda$	4.02	0.56	2.98	5.20	5989	1.00
	$\kappa$	0.06	0.06	0.00	0.22	7500	1.00
	$q$	0.58	0.04	0.50	0.65	5758	1.00

ID	Parámetro	Media	SD	2.5 %	97.5 %	n_eff	Rhat
16	$\alpha$	2.05	0.88	0.77	4.13	3998	1.00
	$\beta$	0.83	0.15	0.58	1.17	3800	1.00
	$\lambda$	1.86	0.67	0.70	3.24	5959	1.00
	$\kappa$	0.25	0.22	0.01	0.90	7500	1.00
	$q$	0.32	0.09	0.14	0.51	6488	1.00
17	$\alpha$	1.74	0.47	0.91	2.77	4817	1.00
	$\beta$	0.99	0.16	0.72	1.33	4919	1.00
	$\lambda$	2.09	0.65	0.93	3.44	6230	1.00
	$\kappa$	0.18	0.18	0.01	0.67	7500	1.00
	$q$	0.25	0.08	0.10	0.40	5090	1.00
18	$\alpha$	1.43	0.19	1.09	1.85	4221	1.00
	$\beta$	0.99	0.07	0.86	1.13	4551	1.00
	$\lambda$	3.07	0.57	2.03	4.25	6083	1.00
	$\kappa$	0.08	0.08	0.00	0.28	7500	1.00
	$q$	0.61	0.04	0.52	0.69	5730	1.00
19	$\alpha$	11.24	6.07	1.80	25.05	4883	1.00
	$\beta$	0.92	0.64	0.04	2.39	4169	1.00
	$\lambda$	0.87	0.63	0.04	2.35	5030	1.00
	$\kappa$	0.71	0.56	0.03	2.08	7500	1.00
	$q$	0.56	0.22	0.11	0.93	5489	1.00
20	$\alpha$	9.07	5.97	1.22	23.09	6179	1.00
	$\beta$	1.05	0.60	0.15	2.43	7500	1.00
	$\lambda$	0.91	0.63	0.04	2.37	4219	1.00
	$\kappa$	0.74	0.57	0.03	2.14	6296	1.00
	$q$	0.63	0.20	0.20	0.95	5559	1.00
21	$\alpha$	1.39	0.88	0.30	3.64	3852	1.00
	$\beta$	0.71	0.16	0.46	1.07	3956	1.00
	$\lambda$	1.80	0.65	0.63	3.17	4380	1.00
	$\kappa$	0.33	0.31	0.01	1.10	6667	1.00
	$q$	0.16	0.11	0.01	0.41	5163	1.00

Cuadro 7.2: Estimaciones para todos los parámetros del Modelo II para cada uno de los animales. Estos resultados corresponden a aquellos para el primer paso del algoritmo de jerarquización. Nivel individual.

## Modelo III

ID	Parámetro	Media	SD	2.5 %	97.5 %	n_eff	Rhat
1	$\alpha$	2.65	0.35	2.01	3.38	4594	1.00
	$\beta$	1.47	0.19	1.13	1.88	4593	1.00
	$\lambda$	2.96	0.58	1.92	4.17	5798	1.00
	$\kappa$	0.17	0.16	0.01	0.61	7500	1.00
	$\rho$	15.64	2.01	11.36	19.28	3930	1.00
	$\epsilon$	1.56	0.44	0.77	2.49	4198	1.00
2	$\alpha$	1.53	0.17	1.22	1.90	3898	1.00
	$\beta$	1.02	0.06	0.90	1.15	3889	1.00
	$\lambda$	3.20	0.59	2.11	4.47	5658	1.00
	$\kappa$	0.10	0.10	0.00	0.36	7204	1.00
	$\rho$	13.71	5.57	2.78	24.50	3114	1.00
	$\epsilon$	0.28	0.14	0.05	0.59	3580	1.00
3	$\alpha$	1.71	0.12	1.48	1.96	4397	1.00
	$\beta$	1.50	0.10	1.32	1.70	4452	1.00
	$\lambda$	3.32	0.56	2.31	4.48	6186	1.00
	$\kappa$	0.10	0.10	0.00	0.36	7500	1.00
	$\rho$	19.75	4.54	9.97	28.00	3013	1.00
	$\epsilon$	0.57	0.18	0.24	0.96	3279	1.00
4	$\alpha$	2.68	0.73	1.44	4.27	3375	1.00
	$\beta$	1.31	0.30	0.82	1.99	3778	1.00
	$\lambda$	2.02	0.70	0.75	3.45	5148	1.00
	$\kappa$	0.28	0.26	0.01	0.95	7500	1.00
	$\rho$	12.91	3.62	7.07	21.73	4796	1.00
	$\epsilon$	1.05	0.50	0.22	2.18	4843	1.00
5	$\alpha$	3.85	1.09	2.10	6.33	3874	1.00
	$\beta$	0.96	0.15	0.70	1.29	3884	1.00
	$\lambda$	2.15	0.65	0.97	3.54	5697	1.00
	$\kappa$	0.20	0.20	0.00	0.71	6335	1.00
	$\rho$	12.73	5.70	2.49	24.93	3371	1.00
	$\epsilon$	0.29	0.21	0.01	0.79	3753	1.00
6	$\alpha$	3.54	0.51	2.62	4.62	4033	1.00
	$\beta$	1.16	0.10	0.97	1.37	4305	1.00
	$\lambda$	2.83	0.63	1.67	4.10	5380	1.00
	$\kappa$	0.13	0.13	0.00	0.47	7500	1.00
	$\rho$	8.73	5.95	0.61	22.64	7500	1.00
	$\epsilon$	0.04	0.04	0.00	0.14	7500	1.00

ID	Parámetro	Media	SD	2.5 %	97.5 %	n_eff	Rhat
7	$\alpha$	1.51	0.24	1.09	2.04	3678	1.00
	$\beta$	0.79	0.05	0.70	0.89	3571	1.00
	$\lambda$	3.41	0.59	2.31	4.63	3803	1.00
	$\kappa$	0.09	0.09	0.00	0.33	5334	1.00
	$\rho$	15.34	11.40	1.24	47.21	1415	1.00
	$\epsilon$	0.05	0.11	0.00	0.41	1031	1.01
8	$\alpha$	2.29	0.41	1.54	3.15	3839	1.00
	$\beta$	1.76	0.36	1.16	2.59	4215	1.00
	$\lambda$	2.01	0.68	0.80	3.41	5296	1.00
	$\kappa$	0.24	0.23	0.01	0.86	6717	1.00
	$\rho$	15.39	3.83	8.47	24.15	4210	1.00
	$\epsilon$	1.09	0.55	0.22	2.39	4519	1.00
9	$\alpha$	2.46	0.34	1.81	3.13	2836	1.00
	$\beta$	1.49	0.19	1.14	1.90	2841	1.00
	$\lambda$	2.47	0.64	1.31	3.80	5266	1.00
	$\kappa$	0.12	0.12	0.00	0.46	6923	1.00
	$\rho$	18.28	9.36	2.31	36.50	2440	1.00
	$\epsilon$	0.15	0.16	0.00	0.57	3227	1.00
10	$\alpha$	1.96	0.27	1.46	2.50	2704	1.00
	$\beta$	1.43	0.17	1.13	1.79	2865	1.00
	$\lambda$	2.55	0.60	1.47	3.79	3953	1.00
	$\kappa$	0.11	0.11	0.00	0.43	5698	1.00
	$\rho$	38.28	9.71	10.86	52.38	797	1.01
	$\epsilon$	0.77	0.42	0.02	1.68	1433	1.00
11	$\alpha$	2.04	0.16	1.75	2.37	4534	1.00
	$\beta$	1.30	0.07	1.17	1.45	4980	1.00
	$\lambda$	2.94	0.58	1.85	4.13	6548	1.00
	$\kappa$	0.08	0.08	0.00	0.30	7500	1.00
	$\rho$	0.77	0.85	0.01	3.12	5057	1.00
	$\epsilon$	0.11	0.04	0.06	0.20	3596	1.00
12	$\alpha$	2.56	0.30	2.01	3.17	3261	1.00
	$\beta$	1.45	0.15	1.19	1.77	3277	1.00
	$\lambda$	2.81	0.60	1.70	4.04	5418	1.00
	$\kappa$	0.09	0.09	0.00	0.35	5929	1.00
	$\rho$	23.30	11.01	2.76	42.15	1548	1.00
	$\epsilon$	0.25	0.22	0.00	0.74	2034	1.00

ID	Parámetro	Media	SD	2.5 %	97.5 %	n_eff	Rhat
13	$\alpha$	1.97	0.21	1.58	2.41	4168	1.00
	$\beta$	1.08	0.07	0.95	1.23	4223	1.00
	$\lambda$	3.34	0.60	2.25	4.59	5727	1.00
	$\kappa$	0.09	0.09	0.00	0.32	7500	1.00
	$\rho$	13.11	7.75	1.35	30.12	4285	1.00
	$\epsilon$	0.05	0.06	0.00	0.22	4423	1.00
14	$\alpha$	1.66	0.28	1.16	2.27	3986	1.00
	$\beta$	0.99	0.10	0.81	1.20	3996	1.00
	$\lambda$	2.82	0.60	1.71	4.05	4364	1.00
	$\kappa$	0.11	0.12	0.00	0.42	6552	1.00
	$\rho$	20.94	6.05	7.36	32.19	2416	1.00
	$\epsilon$	0.46	0.24	0.04	0.96	2811	1.00
15	$\alpha$	1.66	0.15	1.37	1.95	3023	1.00
	$\beta$	1.31	0.10	1.13	1.51	2754	1.00
	$\lambda$	3.90	0.57	2.83	5.08	5083	1.00
	$\kappa$	0.07	0.08	0.00	0.28	6702	1.00
	$\rho$	30.90	8.39	10.91	44.49	1760	1.00
	$\epsilon$	0.46	0.22	0.05	0.90	2376	1.00
16	$\alpha$	2.15	0.87	0.84	4.19	4218	1.00
	$\beta$	0.84	0.15	0.59	1.18	4170	1.00
	$\lambda$	1.79	0.68	0.61	3.21	5107	1.00
	$\kappa$	0.26	0.24	0.01	0.89	7500	1.00
	$\rho$	20.57	5.04	11.50	31.67	5525	1.00
	$\epsilon$	0.95	0.43	0.24	1.92	7500	1.00
17	$\alpha$	1.87	0.48	1.03	2.91	5139	1.00
	$\beta$	1.02	0.16	0.73	1.36	5184	1.00
	$\lambda$	2.00	0.68	0.81	3.42	6494	1.00
	$\kappa$	0.19	0.19	0.00	0.71	7500	1.00
	$\rho$	30.74	4.08	23.72	40.02	6558	1.00
	$\epsilon$	1.67	0.50	0.81	2.77	6633	1.00
18	$\alpha$	1.46	0.19	1.12	1.86	4359	1.00
	$\beta$	0.99	0.07	0.87	1.13	4304	1.00
	$\lambda$	3.03	0.58	1.99	4.23	5311	1.00
	$\kappa$	0.08	0.08	0.00	0.29	7500	1.00
	$\rho$	9.39	6.19	0.72	23.59	5972	1.00
	$\epsilon$	0.03	0.03	0.00	0.11	6607	1.00

ID	Parámetro	Media	SD	2.5 %	97.5 %	n_eff	Rhat
19	$\alpha$	11.16	5.92	2.27	25.10	5720	1.00
	$\beta$	0.91	0.65	0.04	2.44	4658	1.00
	$\lambda$	0.87	0.62	0.05	2.34	4399	1.00
	$\kappa$	0.71	0.57	0.03	2.13	5278	1.00
	$\rho$	6.32	5.39	0.54	20.24	4898	1.00
	$\epsilon$	0.67	0.55	0.03	2.05	5387	1.00
20	$\alpha$	9.13	5.94	1.24	23.13	5688	1.00
	$\beta$	1.03	0.62	0.10	2.45	5245	1.00
	$\lambda$	0.93	0.64	0.04	2.35	4818	1.00
	$\kappa$	0.73	0.57	0.03	2.15	7500	1.00
	$\rho$	5.13	4.90	0.42	18.48	4340	1.00
	$\epsilon$	0.66	0.56	0.02	2.09	4199	1.00
21	$\alpha$	1.69	1.06	0.42	4.29	4858	1.00
	$\beta$	0.74	0.16	0.48	1.12	4697	1.00
	$\lambda$	1.76	0.67	0.61	3.20	7500	1.00
	$\kappa$	0.34	0.33	0.01	1.20	7500	1.00
	$\rho$	17.91	4.86	10.47	28.86	5837	1.00
	$\epsilon$	1.53	0.58	0.54	2.81	6188	1.00

Cuadro 7.3: Estimaciones para todos los parámetros del Modelo III para cada uno de los animales. Estos resultados corresponden a aquellos para el primer paso del algoritmo de jerarquización. Nivel individual.

## Modelo IV

ID	Parámetro	Media	SD	2.5 %	97.5 %	n_eff	Rhat
1	$\alpha$	2.90	0.38	2.19	3.69	4854	1.00
	$\beta$	1.62	0.22	1.23	2.10	4932	1.00
	$\lambda$	3.43	0.57	2.40	4.62	7902	1.00
	$\kappa$	0.14	0.14	0.00	0.52	9901	1.00
	$\rho$	19.72	3.33	12.84	26.25	2012	1.00
	$\epsilon$	1.05	0.41	0.30	1.91	2943	1.00
	$\nu$	10.11	5.32	1.91	22.13	4455	1.00
	$\theta$	0.73	0.41	0.13	1.68	4345	1.00
2	$\alpha$	1.34	0.17	1.04	1.70	3700	1.00
	$\beta$	1.00	0.06	0.89	1.13	3771	1.00
	$\lambda$	3.57	0.58	2.53	4.78	5492	1.00
	$\kappa$	0.08	0.08	0.00	0.28	7249	1.00
	$\rho$	25.42	9.81	4.25	41.28	1326	1.00
	$\epsilon$	0.32	0.23	0.01	0.81	1903	1.00
	$\nu$	12.26	6.40	2.39	26.64	4461	1.00
	$\theta$	0.23	0.18	0.02	0.69	4063	1.00
3	$\alpha$	1.60	0.12	1.37	1.84	4036	1.00
	$\beta$	1.50	0.09	1.32	1.69	4055	1.00
	$\lambda$	3.75	0.53	2.79	4.85	5695	1.00
	$\kappa$	0.07	0.07	0.00	0.25	6818	1.00
	$\rho$	30.37	11.21	4.99	46.13	1228	1.00
	$\epsilon$	0.40	0.28	0.01	0.97	1466	1.00
	$\nu$	12.44	6.56	2.07	27.16	3804	1.00
	$\theta$	0.16	0.10	0.02	0.41	4024	1.00
4	$\alpha$	2.44	0.73	1.24	4.07	4126	1.00
	$\beta$	1.24	0.28	0.78	1.90	4345	1.00
	$\lambda$	2.12	0.71	0.78	3.55	5921	1.00
	$\kappa$	0.26	0.25	0.01	0.95	8370	1.00
	$\rho$	15.60	4.35	8.50	25.86	4761	1.00
	$\epsilon$	1.05	0.49	0.24	2.15	5530	1.00
	$\nu$	11.02	6.01	2.06	24.80	5991	1.00
	$\theta$	0.48	0.42	0.01	1.59	6948	1.00
5	$\alpha$	5.58	1.71	2.81	9.43	4241	1.00
	$\beta$	1.09	0.20	0.75	1.55	4231	1.00
	$\lambda$	2.21	0.66	1.04	3.57	5713	1.00
	$\kappa$	0.20	0.19	0.01	0.72	8114	1.00
	$\rho$	13.27	5.82	2.65	25.52	3314	1.00
	$\epsilon$	0.29	0.22	0.01	0.79	3944	1.00
	$\nu$	8.17	4.57	1.32	18.76	3887	1.00
	$\theta$	0.94	0.52	0.15	2.15	3800	1.00

ID	Parámetro	Media	SD	2.5 %	97.5 %	n_eff	Rhat
6	$\alpha$	3.30	0.49	2.43	4.32	4325	1.00
	$\beta$	1.16	0.10	0.97	1.37	4593	1.00
	$\lambda$	2.97	0.62	1.84	4.24	6376	1.00
	$\kappa$	0.12	0.12	0.00	0.46	8538	1.00
	$\rho$	10.65	6.57	0.90	25.38	5724	1.00
	$\epsilon$	0.04	0.04	0.00	0.15	6891	1.00
	$\nu$	12.03	6.29	2.08	26.17	6138	1.00
	$\theta$	0.32	0.26	0.01	0.98	5541	1.00
7	$\alpha$	1.53	0.26	1.08	2.12	1430	1.01
	$\beta$	0.80	0.05	0.70	0.90	1858	1.00
	$\lambda$	3.55	0.60	2.46	4.77	500	1.01
	$\kappa$	0.09	0.09	0.00	0.34	773	1.00
	$\rho$	20.81	18.85	1.11	65.86	146	1.02
	$\epsilon$	0.15	0.30	0.00	1.03	127	1.02
	$\nu$	11.09	6.10	1.90	24.53	900	1.00
	$\theta$	0.48	0.34	0.05	1.31	625	1.01
8	$\alpha$	2.09	0.39	1.39	2.93	4145	1.00
	$\beta$	1.71	0.34	1.17	2.50	4239	1.00
	$\lambda$	2.16	0.67	0.93	3.56	6588	1.00
	$\kappa$	0.22	0.21	0.01	0.78	7737	1.00
	$\rho$	18.75	4.06	11.83	27.98	4185	1.00
	$\epsilon$	1.25	0.57	0.34	2.54	5377	1.00
	$\nu$	12.12	6.41	2.05	26.48	6346	1.00
	$\theta$	0.28	0.28	0.01	1.05	6737	1.00
9	$\alpha$	2.41	0.34	1.76	3.12	3400	1.00
	$\beta$	1.50	0.20	1.16	1.92	3545	1.00
	$\lambda$	2.55	0.64	1.40	3.90	5159	1.00
	$\kappa$	0.12	0.12	0.00	0.44	7196	1.00
	$\rho$	22.53	10.83	2.59	41.55	2177	1.00
	$\epsilon$	0.25	0.24	0.00	0.81	2720	1.00
	$\nu$	12.38	6.54	2.16	27.13	4129	1.00
	$\theta$	0.12	0.12	0.00	0.45	5006	1.00

ID	Parámetro	Media	SD	2.5 %	97.5 %	n_eff	Rhat
10	$\alpha$	1.86	0.27	1.35	2.42	4306	1.00
	$\beta$	1.40	0.16	1.11	1.75	4450	1.00
	$\lambda$	2.66	0.61	1.57	3.93	6356	1.00
	$\kappa$	0.10	0.10	0.00	0.39	7802	1.00
	$\rho$	47.13	6.27	34.54	57.92	1693	1.00
	$\epsilon$	1.39	0.52	0.48	2.55	4280	1.00
	$\nu$	12.19	6.50	2.22	26.74	4996	1.00
	$\theta$	0.24	0.23	0.01	0.86	5673	1.00
11	$\alpha$	1.94	0.16	1.66	2.28	4188	1.00
	$\beta$	1.33	0.07	1.21	1.46	4371	1.00
	$\lambda$	3.29	0.55	2.31	4.45	5897	1.00
	$\kappa$	0.06	0.06	0.00	0.22	7627	1.00
	$\rho$	3.84	3.81	0.11	14.17	5665	1.00
	$\epsilon$	0.06	0.03	0.01	0.14	4167	1.00
	$\nu$	11.98	6.19	2.28	25.67	3797	1.00
	$\theta$	0.37	0.21	0.07	0.86	3758	1.00
12	$\alpha$	2.38	0.28	1.86	2.96	3947	1.00
	$\beta$	1.41	0.14	1.17	1.70	3897	1.00
	$\lambda$	2.92	0.58	1.86	4.13	5338	1.00
	$\kappa$	0.09	0.09	0.00	0.33	5642	1.00
	$\rho$	26.16	13.49	2.94	48.97	1570	1.00
	$\epsilon$	0.27	0.27	0.00	0.88	1850	1.00
	$\nu$	12.08	6.39	2.11	26.16	4737	1.00
	$\theta$	0.20	0.17	0.01	0.65	4682	1.00
13	$\alpha$	1.98	0.25	1.53	2.48	4523	1.00
	$\beta$	1.07	0.07	0.93	1.22	4588	1.00
	$\lambda$	3.61	0.58	2.54	4.81	6647	1.00
	$\kappa$	0.08	0.08	0.00	0.31	8122	1.00
	$\rho$	13.26	8.66	0.97	33.71	4539	1.00
	$\epsilon$	0.03	0.05	0.00	0.19	4525	1.00
	$\nu$	12.00	6.30	2.18	26.07	5333	1.00
	$\theta$	0.33	0.24	0.03	0.91	5259	1.00
14	$\alpha$	1.58	0.30	1.04	2.21	3797	1.00
	$\beta$	0.99	0.10	0.80	1.21	3809	1.00
	$\lambda$	3.07	0.58	2.03	4.28	5853	1.00
	$\kappa$	0.10	0.10	0.00	0.37	6806	1.00
	$\rho$	25.31	9.06	4.70	40.01	1095	1.00
	$\epsilon$	0.40	0.27	0.01	0.97	1775	1.00
	$\nu$	12.06	6.43	2.13	26.72	4249	1.00
	$\theta$	0.34	0.26	0.03	0.99	4615	1.00

ID	Parámetro	Media	SD	2.5 %	97.5 %	n_eff	Rhat
15	$\alpha$	1.57	0.14	1.30	1.84	2924	1.00
	$\beta$	1.32	0.09	1.14	1.51	3272	1.00
	$\lambda$	4.35	0.54	3.35	5.46	4165	1.00
	$\kappa$	0.07	0.07	0.00	0.24	5592	1.00
	$\rho$	32.08	20.60	1.79	62.74	526	1.01
	$\epsilon$	0.36	0.41	0.00	1.22	602	1.01
	$\nu$	12.58	6.54	2.45	27.18	3836	1.00
	$\theta$	0.07	0.06	0.00	0.22	940	1.00
16	$\alpha$	2.06	0.94	0.72	4.40	4540	1.00
	$\beta$	0.83	0.16	0.57	1.20	4730	1.00
	$\lambda$	1.83	0.68	0.64	3.29	7765	1.00
	$\kappa$	0.25	0.24	0.01	0.90	8891	1.00
	$\rho$	21.54	4.94	13.08	32.45	5457	1.00
	$\epsilon$	1.09	0.47	0.31	2.17	6894	1.00
	$\nu$	9.83	5.92	1.46	23.80	7405	1.00
	$\theta$	0.64	0.52	0.02	1.92	7281	1.00
17	$\alpha$	1.85	0.50	0.99	2.91	5436	1.00
	$\beta$	1.01	0.16	0.73	1.36	5584	1.00
	$\lambda$	2.06	0.67	0.88	3.50	7398	1.00
	$\kappa$	0.19	0.19	0.00	0.70	8636	1.00
	$\rho$	31.99	4.37	24.23	41.32	6089	1.00
	$\epsilon$	1.72	0.52	0.83	2.85	6091	1.00
	$\nu$	11.87	6.11	2.18	25.59	6839	1.00
	$\theta$	0.32	0.30	0.01	1.12	6409	1.00
18	$\alpha$	1.45	0.20	1.11	1.88	4195	1.00
	$\beta$	1.02	0.07	0.89	1.16	4305	1.00
	$\lambda$	3.18	0.57	2.15	4.38	7072	1.00
	$\kappa$	0.07	0.08	0.00	0.28	8334	1.00
	$\rho$	10.84	6.77	0.87	26.33	6079	1.00
	$\epsilon$	0.03	0.03	0.00	0.12	6208	1.00
	$\nu$	11.81	6.29	2.02	26.08	4577	1.00
	$\theta$	0.35	0.23	0.05	0.90	4871	1.00

ID	Parámetro	Media	SD	2.5 %	97.5 %	n_eff	Rhat
19	$\alpha$	11.28	5.90	1.98	24.90	5657	1.00
	$\beta$	0.92	0.66	0.04	2.46	5814	1.00
	$\lambda$	0.88	0.63	0.05	2.37	6732	1.00
	$\kappa$	0.70	0.56	0.03	2.06	7501	1.00
	$\rho$	6.51	5.46	0.54	20.34	6590	1.00
	$\epsilon$	0.69	0.55	0.03	2.05	6014	1.00
	$\nu$	8.74	5.89	0.91	22.63	7487	1.00
	$\theta$	0.73	0.58	0.02	2.10	6190	1.00
20	$\alpha$	9.13	5.97	1.23	23.41	7443	1.00
	$\beta$	1.03	0.60	0.13	2.40	6461	1.00
	$\lambda$	0.93	0.65	0.04	2.43	6426	1.00
	$\kappa$	0.72	0.57	0.03	2.12	6858	1.00
	$\rho$	5.31	5.01	0.35	18.37	4549	1.00
	$\epsilon$	0.64	0.56	0.02	2.12	5056	1.00
	$\nu$	9.10	5.88	1.00	23.13	6365	1.00
	$\theta$	0.71	0.56	0.03	2.10	6236	1.00
21	$\alpha$	1.75	1.09	0.40	4.57	4207	1.00
	$\beta$	0.75	0.17	0.47	1.14	4282	1.00
	$\lambda$	1.74	0.68	0.54	3.20	6414	1.00
	$\kappa$	0.34	0.31	0.01	1.17	8859	1.00
	$\rho$	18.25	4.90	10.70	29.59	6398	1.00
	$\epsilon$	1.58	0.60	0.54	2.92	5546	1.00
	$\nu$	8.95	5.87	1.14	23.18	7752	1.00
	$\theta$	0.73	0.57	0.03	2.12	7315	1.00

Cuadro 7.4: Estimaciones para todos los parámetros del Modelo IV para cada uno de los animales. Estos resultados corresponden a aquellos para el primer paso del algoritmo de jerarquización. Nivel individual.

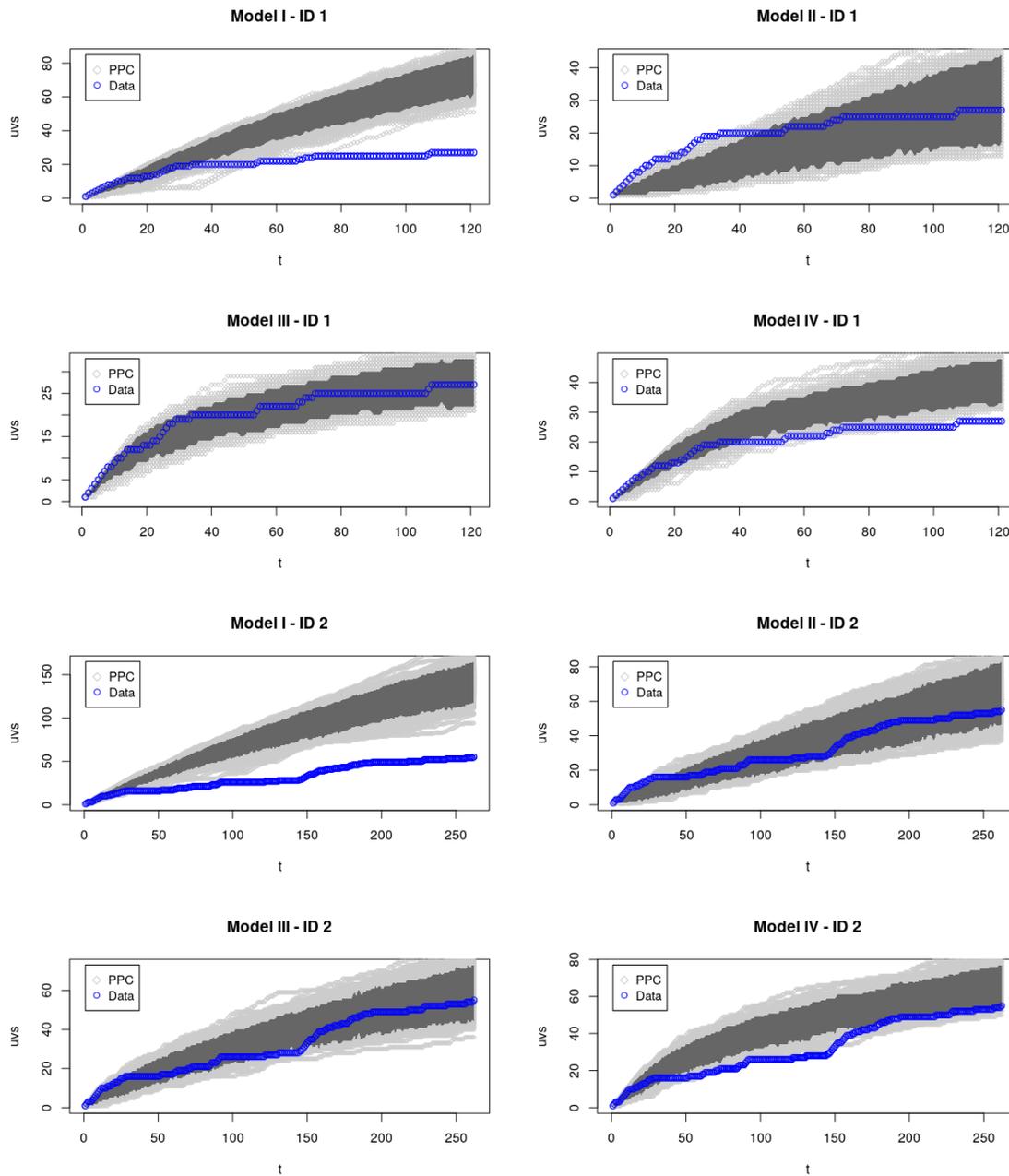
## WAIC

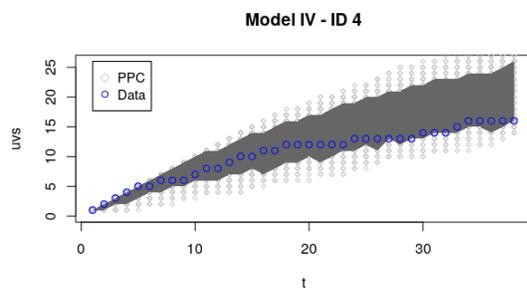
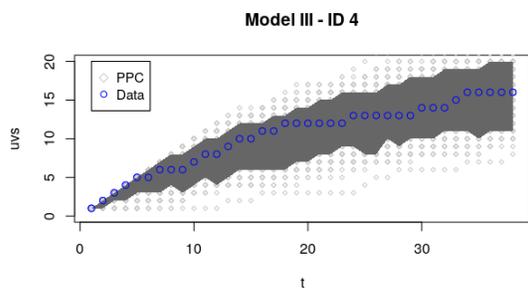
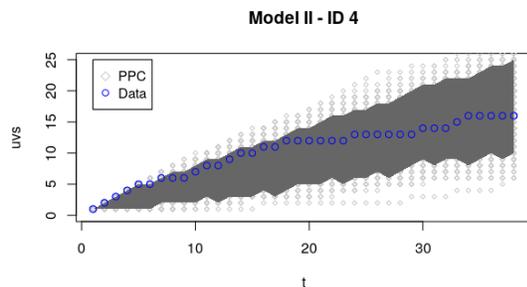
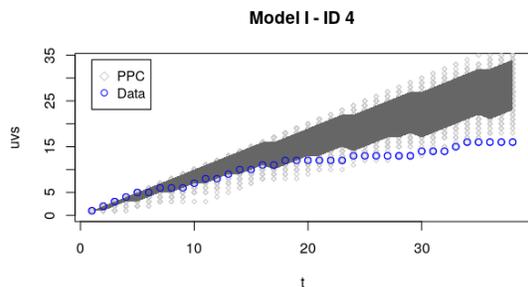
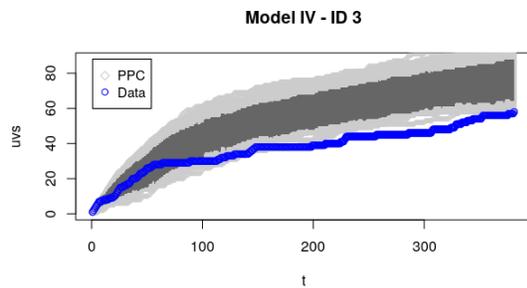
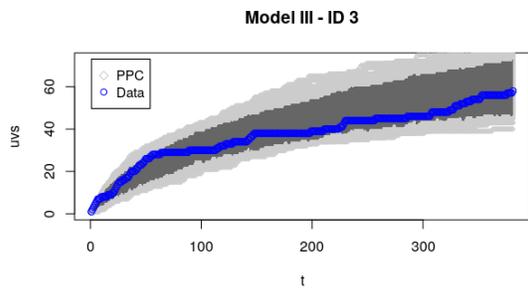
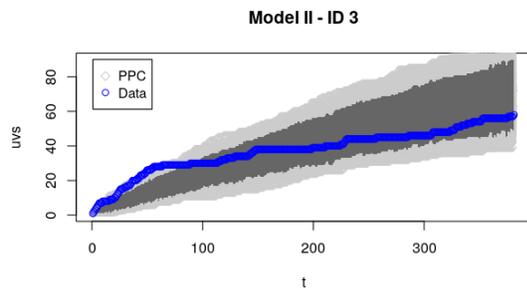
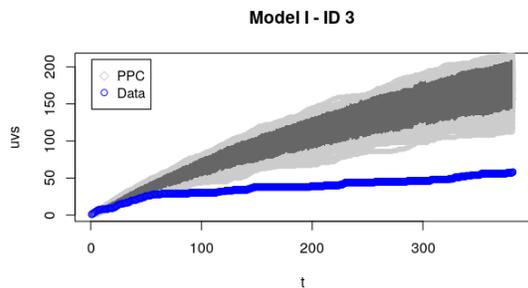
Modelo ID	Modelo I		Modelo II		Modelo III		Modelo IV				
	WAIC	P <sub>w</sub>	WAIC	P <sub>w</sub>	WAIC	P <sub>w</sub>	WAIC	P <sub>w</sub>			
	Est	$\delta$	Est	$\delta$	Est	$\delta$	Est	$\delta$			
1	891.80	137.29	778.55	24.04	3.20	<b>754.51</b>	<b>0.00</b>	3.76	781.87	27.36	5.25
2	1853.59	219.51	1639.01	4.93	5.88	<b>1634.08</b>	<b>0.00</b>	<b>5.99</b>	1705.97	71.89	8.14
3	2739.68	351.57	2404.02	15.91	4.04	<b>2388.11</b>	<b>0.00</b>	<b>4.70</b>	2463.32	75.21	6.98
4	275.76	23.54	257.40	5.18	2.99	<b>252.22</b>	<b>0.00</b>	<b>3.03</b>	263.69	11.47	3.67
5	578.03	76.87	501.76	0.50	3.77	<b>501.16</b>	<b>18.73</b>	<b>3.79</b>	501.17	0.01	4.11
6	1128.17	190.04	938.73	0.60	5.76	<b>938.13</b>	<b>0.00</b>	<b>4.64</b>	974.15	36.02	5.47
7	1658.74	76.84	<b>1581.90</b>	<b>0.00</b>	<b>4.56</b>	1616.15	34.25	3.97	1640.63	58.73	7.37
8	388.13	22.61	371.02	5.50	2.60	<b>365.52</b>	<b>0.00</b>	<b>2.78</b>	372.39	6.87	2.92
9	830.40	48.43	<b>781.97</b>	<b>0.00</b>	<b>3.40</b>	794.68	12.71	3.22	817.39	35.42	4.52
10	780.99	12.51	<b>768.48</b>	<b>0.00</b>	<b>3.62</b>	789.63	21.15	5.79	794.34	25.86	5.12
11	3400.63	676.57	<b>2724.06</b>	<b>0.00</b>	<b>4.97</b>	2731.11	7.05	5.11	2841.65	117.59	7.77
12	1248.45	103.88	<b>1144.57</b>	<b>0.00</b>	<b>3.82</b>	1151.11	6.54	4.37	1192.22	47.65	5.65
13	1722.04	160.84	<b>1561.20</b>	<b>0.00</b>	<b>6.35</b>	1566.47	5.44	5.27	1610.94	49.74	6.39
14	1012.85	102.98	911.82	1.95	4.42	<b>909.87</b>	<b>0.00</b>	<b>4.74</b>	941.96	32.09	6.81
15	2539.97	213.89	2334.23	8.15	3.48	<b>2326.08</b>	<b>0.00</b>	<b>4.69</b>	2395.97	69.89	8.76
16	273.84	6.26	<b>267.58</b>	<b>0.00</b>	<b>3.88</b>	272.58	5.00	4.50	270.82	3.24	3.24
17	454.51	12.17	445.67	3.33	3.83	<b>442.34</b>	<b>0.00</b>	<b>3.82</b>	446.17	3.83	4.29
18	1599.71	121.43	<b>1478.28</b>	<b>0.00</b>	<b>4.19</b>	1478.33	0.05	2.92	1528.78	50.50	4.67
19	20.96	4.54	17.35	0.93	0.88	<b>16.42</b>	<b>0.00</b>	<b>0.57</b>	17.53	1.11	0.57
20	26.16	8.14	19.03	1.01	0.98	<b>18.02</b>	<b>0.00</b>	<b>0.61</b>	19.36	1.34	0.65
21	<b>151.95</b>	<b>0.00</b>	154.23	2.28	3.78	153.32	1.37	3.57	155.31	3.36	3.75

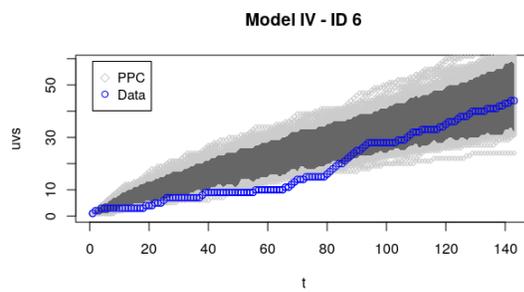
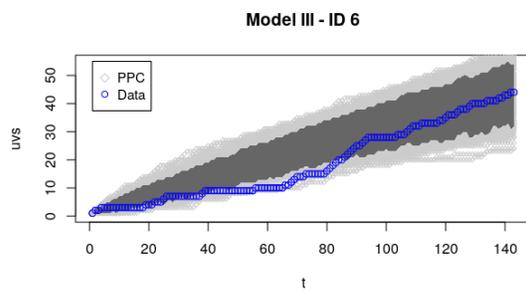
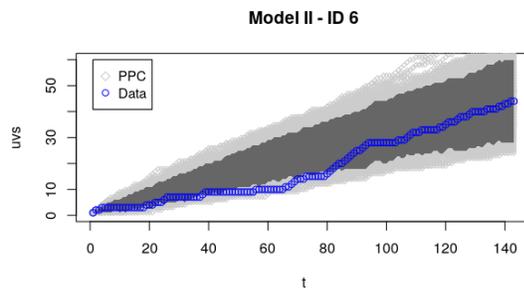
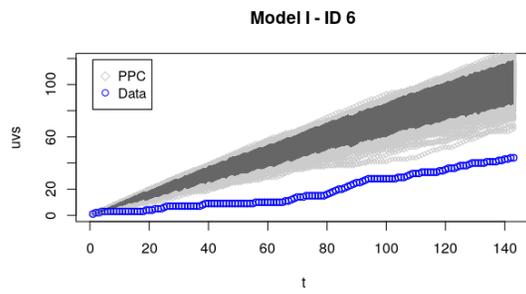
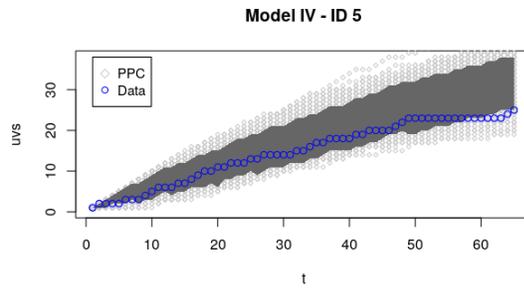
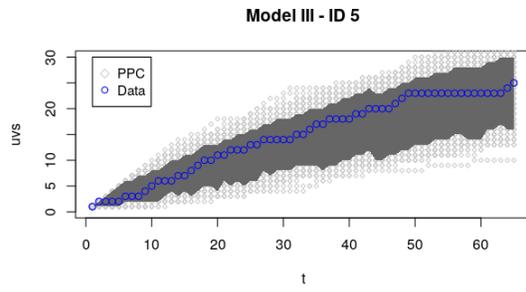
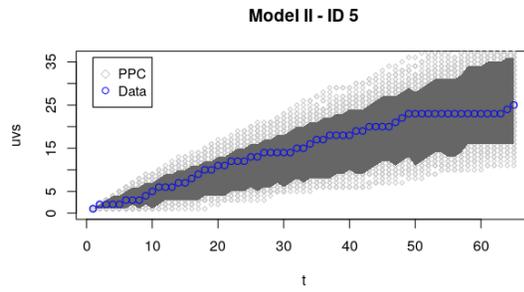
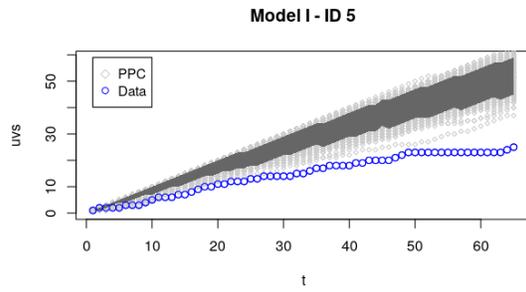
Cuadro 7.5: WAIC para el log-likelihood puntual para cada Modelo y cada animal a nivel individual. La tabla muestra las estimaciones puntuales (Est) para el criterio de información WAIC, el número efectivo de parámetros (P<sub>w</sub>) y la diferencia entre los diversos valores del WAIC entre Modelos como  $\delta$ .

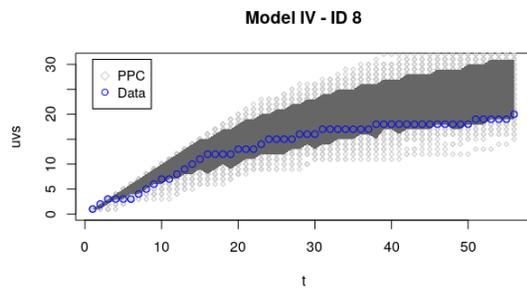
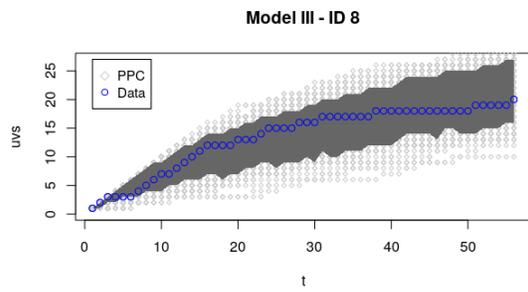
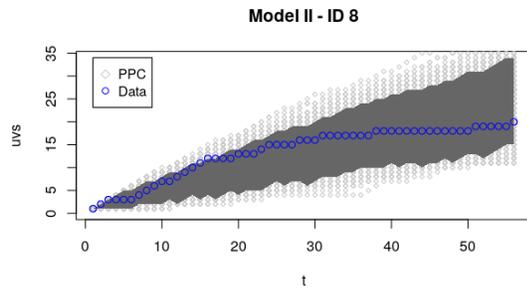
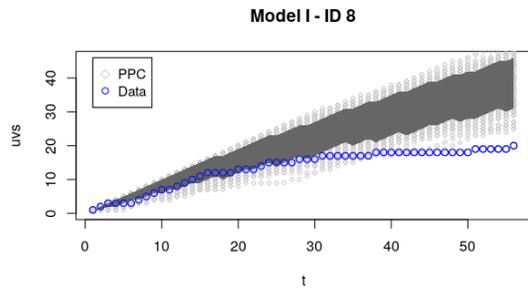
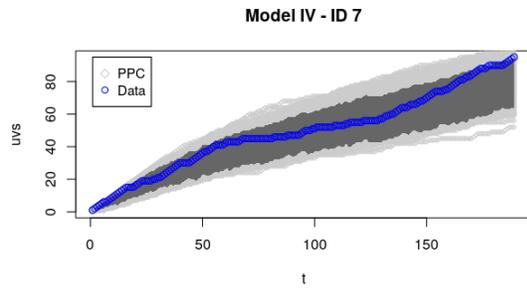
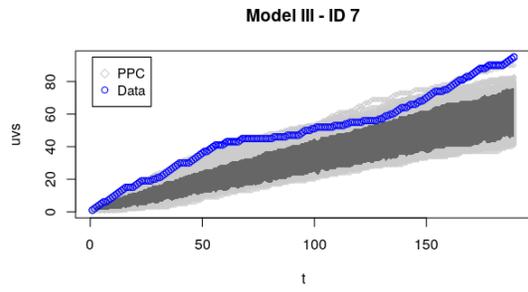
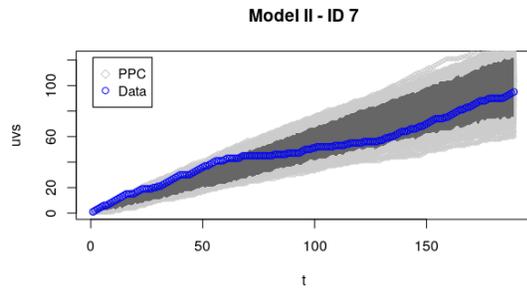
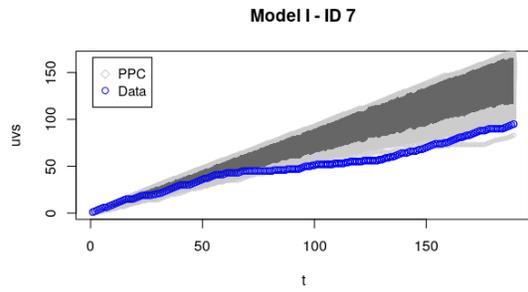
## PPC

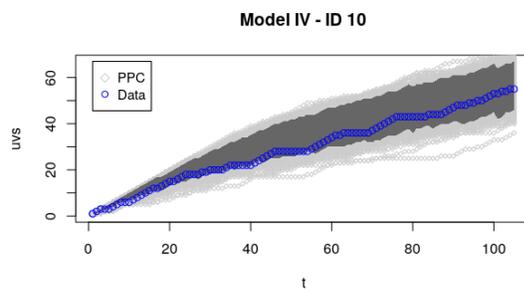
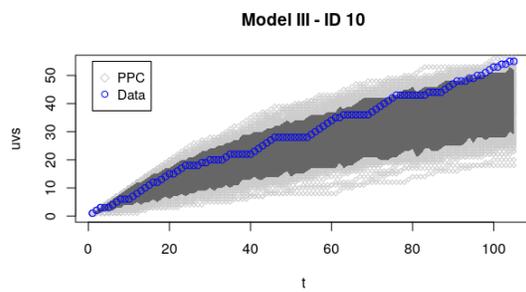
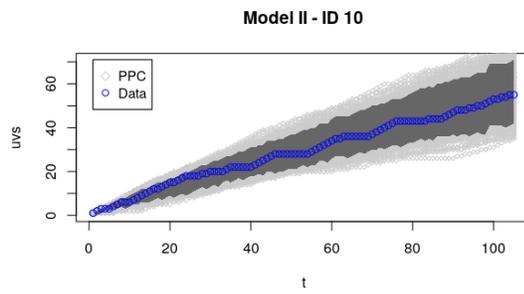
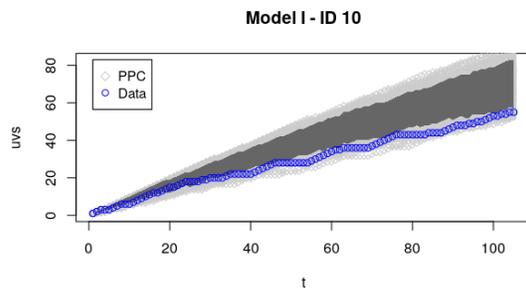
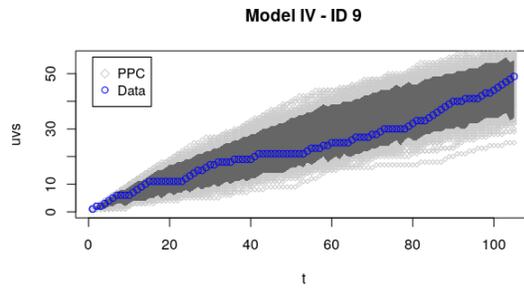
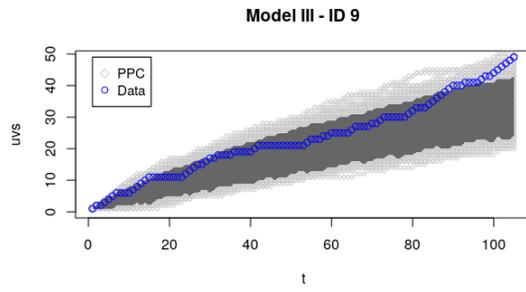
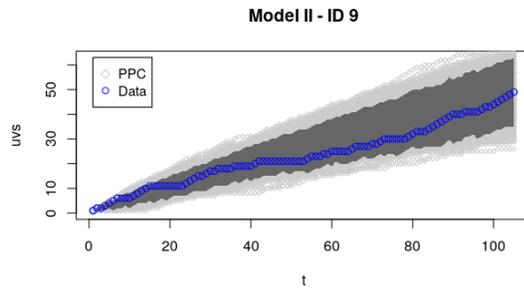
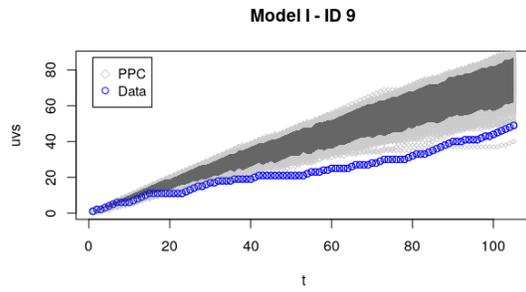
Las siguientes figuras se leen como sigue: Cada bloque de  $2 \times 2$  corresponde a la comparación entre cuatro modelos para un solo individuo. En símbolos azules el número de sitios únicos visitados observados, en símbolos gris claro la predicción hecha usando la técnica PPC sobre mil muestras independientes y en gris oscuro intervalo de credibilidad de dicha predicción al 95 %.

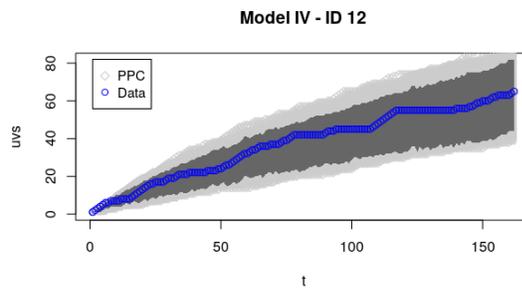
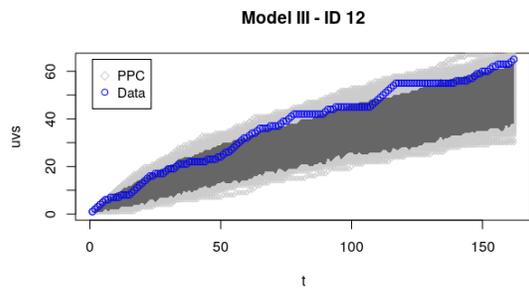
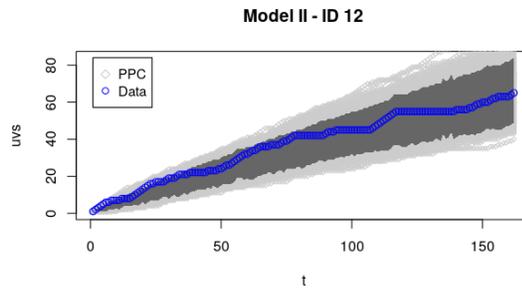
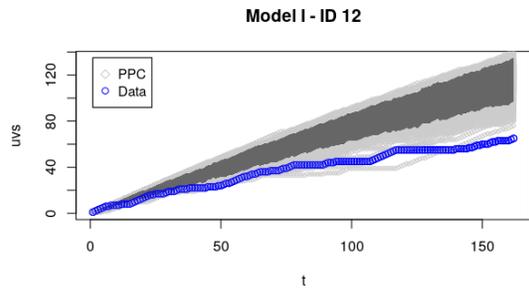
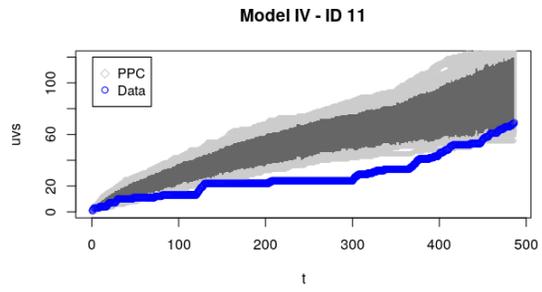
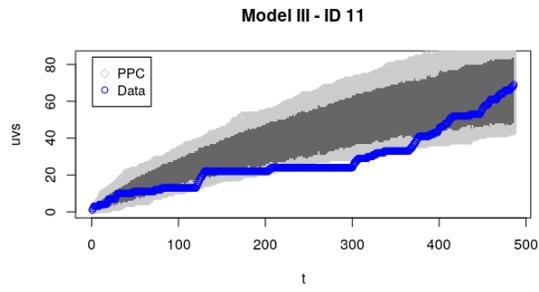
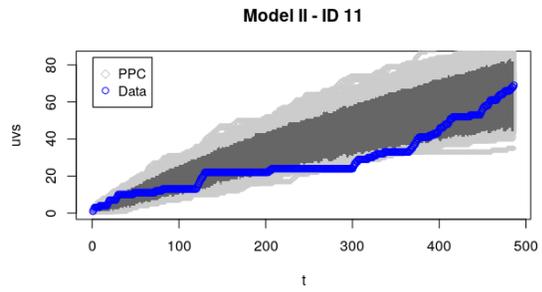
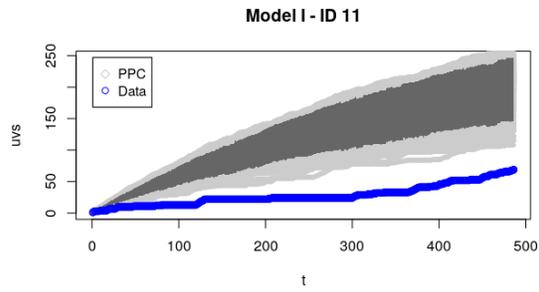


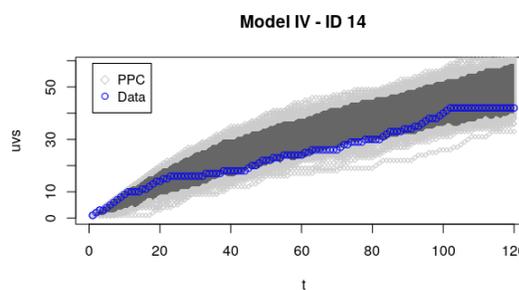
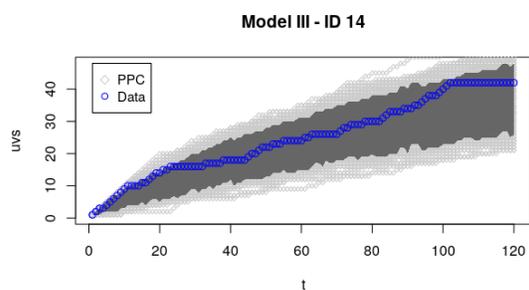
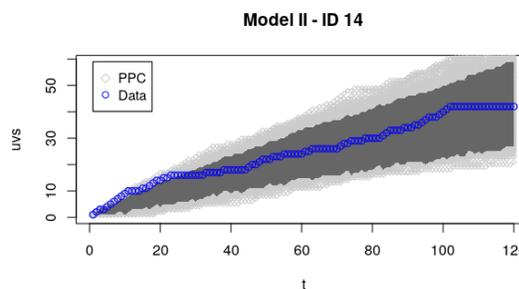
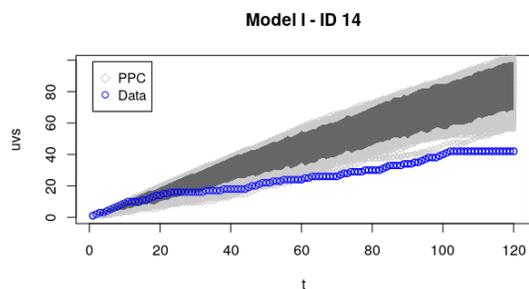
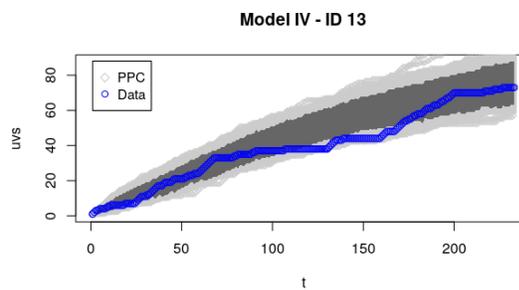
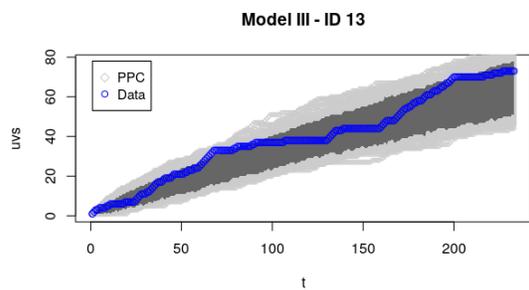
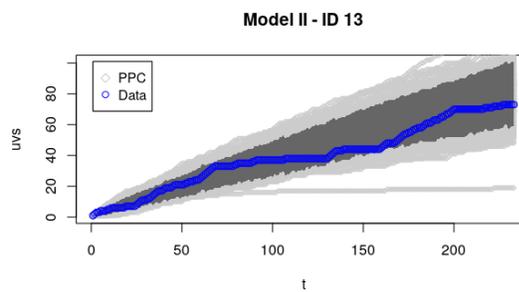
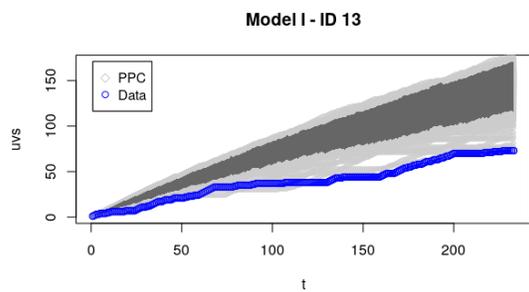


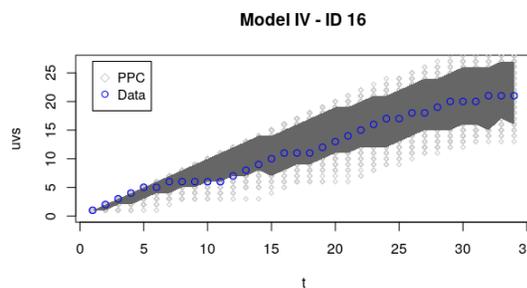
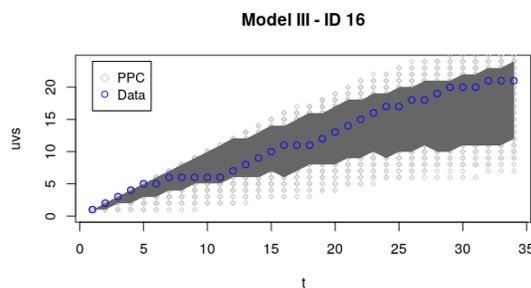
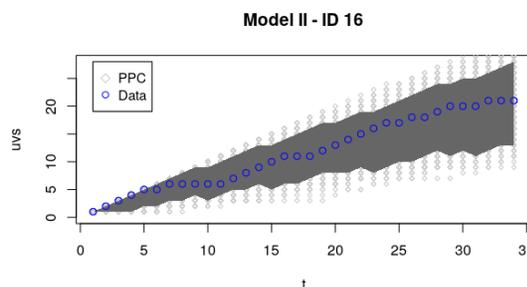
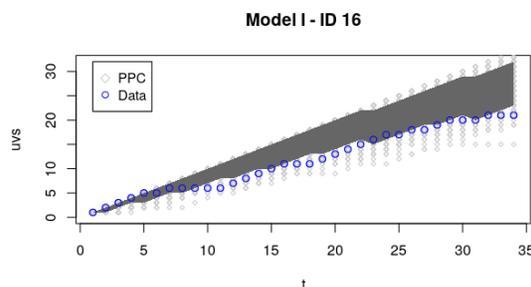
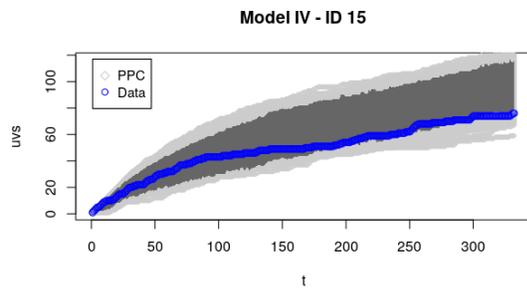
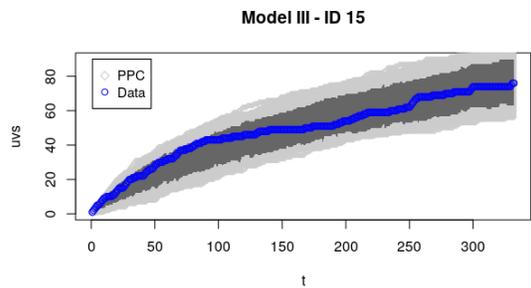
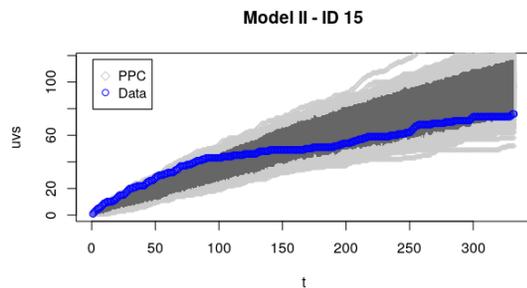
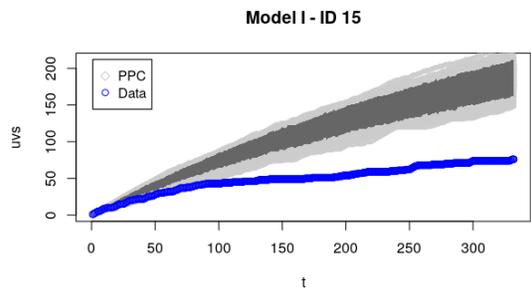




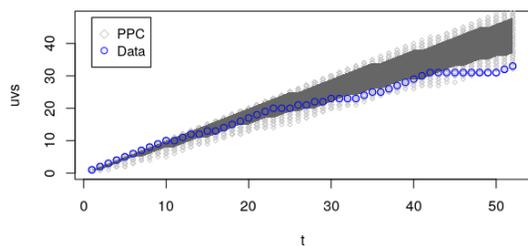




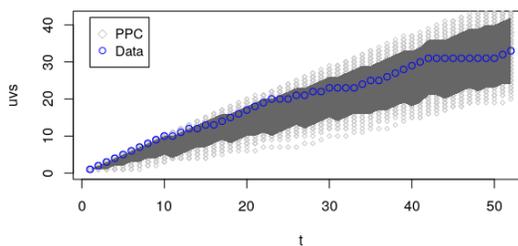




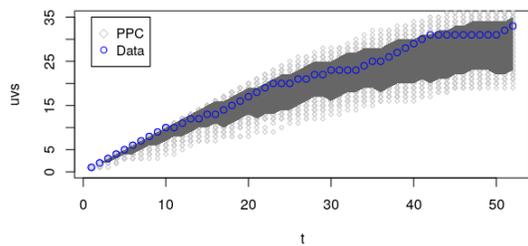
Model I - ID 17



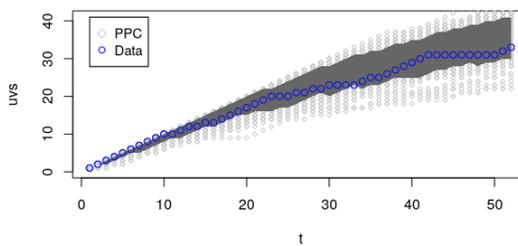
Model II - ID 17



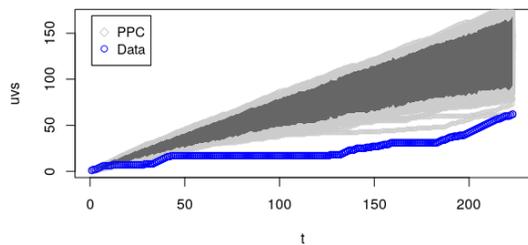
Model III - ID 17



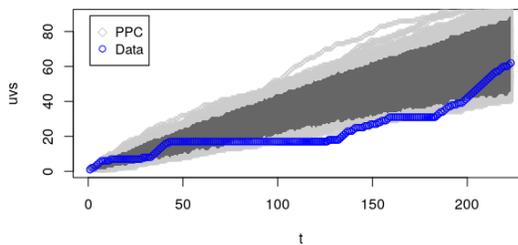
Model IV - ID 17



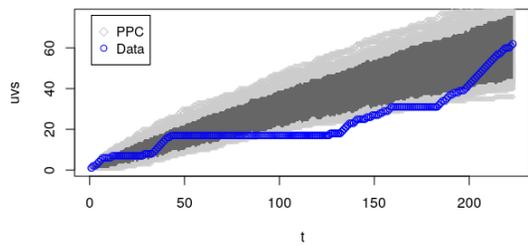
Model I - ID 18



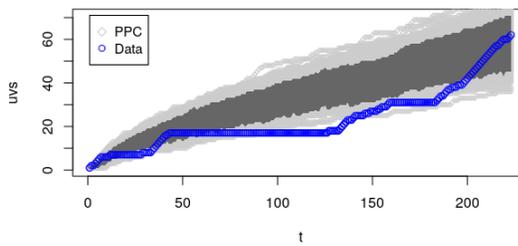
Model II - ID 18

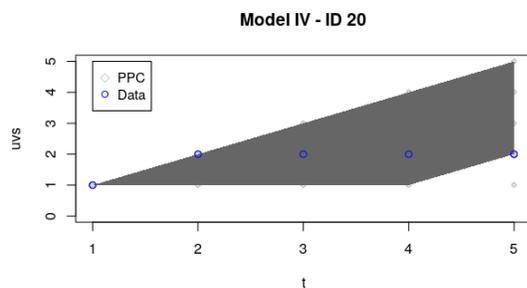
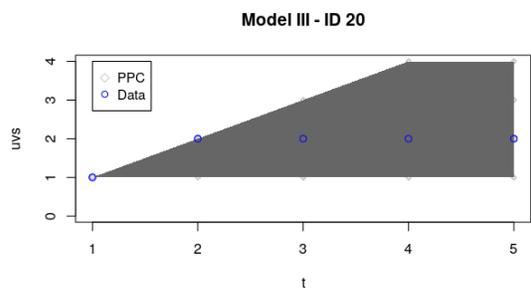
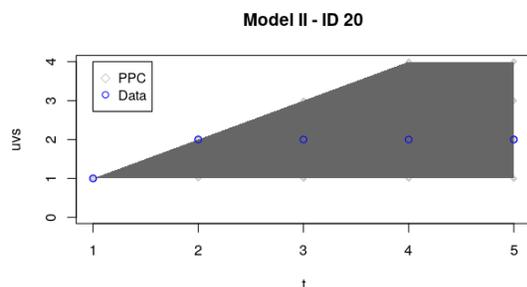
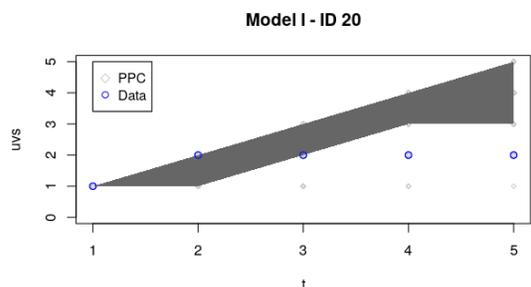
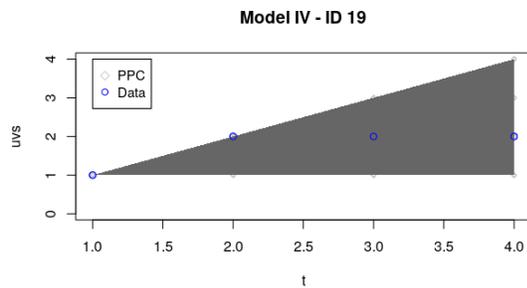
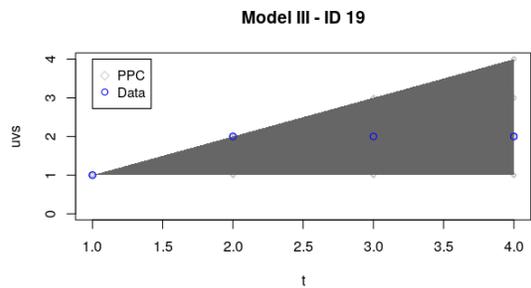
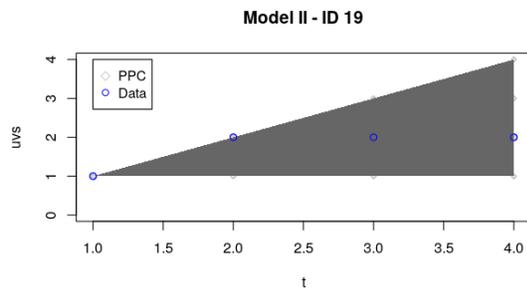
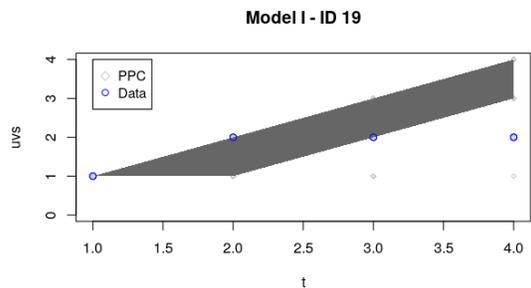


Model III - ID 18



Model IV - ID 18





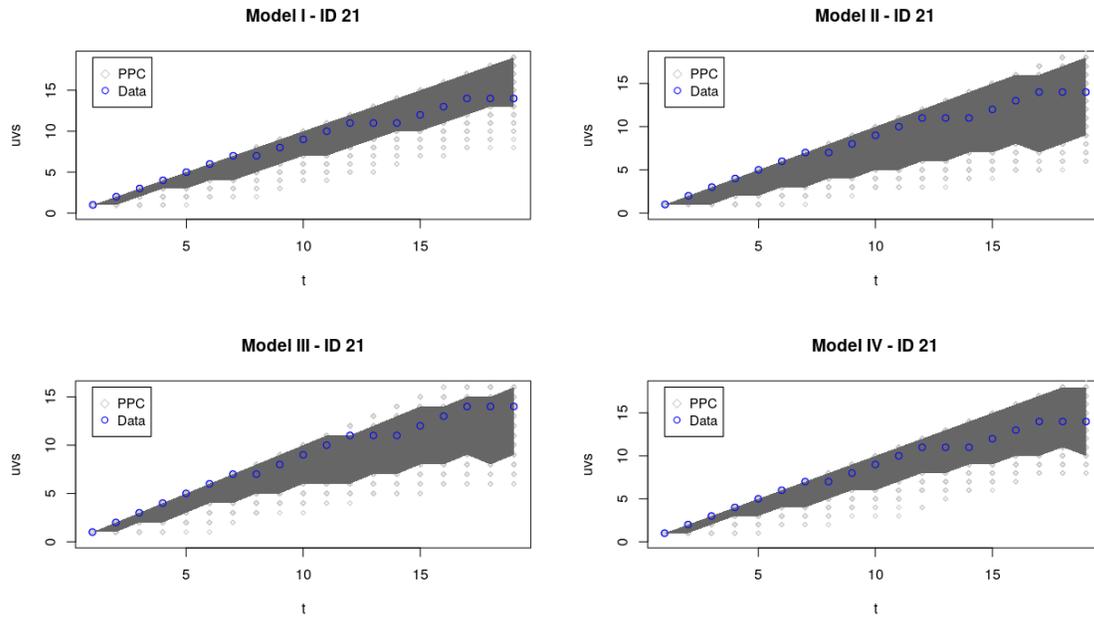


Figura 7.1: Comprobación predictiva posterior (PPC) para cada uno de los Modelos a nivel individual para todos los animales observados. El número de sitios únicos visitados (uvs) como función del tiempo  $t$  (símbolos azules). En gris claro las curvas para el PPC, en gris oscuro los intervalos de credibilidad al 95% y en azul los datos para cada una de las trayectorias reales.

## 7.2. Nivel poblacional

Las estimaciones y cálculos mostrados a continuación corresponden al ajuste poblacional, es decir, los resultados obtenidos al ejecutar el segundo paso del algoritmo de jerarquización explicado en el capítulo 5.

Modelo I

Parámetro ID	$\alpha$				$\beta$					
	Media	SD	2.5 %	97.5 %	n_eff	Media	SD	2.5 %	97.5 %	n_eff
1	1.29	0.18	1.17	1.66	3048	1.11	0.11	1.03	1.35	5676
2	0.71	0.07	0.66	0.87	4842	0.86	0.04	0.83	0.95	7500
3	1.12	0.08	1.07	1.28	6426	1.27	0.06	1.22	1.40	5360
4	1.13	0.25	0.95	1.68	4066	0.96	0.16	0.85	1.31	6445
5	1.06	0.25	0.88	1.58	4783	0.72	0.09	0.66	0.90	5690
6	1.21	0.18	1.08	1.60	4082	0.86	0.07	0.81	1.00	7204
7	0.60	0.09	0.53	0.81	3318	0.64	0.03	0.62	0.71	6281
8	1.27	0.19	1.13	1.68	3392	1.30	0.17	1.18	1.67	3628
9	1.20	0.18	1.06	1.58	3951	1.05	0.11	0.97	1.28	6272
10	1.28	0.18	1.15	1.65	3734	1.16	0.11	1.08	1.41	5181
11	0.82	0.05	0.74	0.92	5498	1.02	0.03	0.99	1.10	6679
12	1.14	0.15	1.03	1.45	5856	1.00	0.07	0.94	1.17	7087
13	0.97	0.11	0.89	1.20	6002	0.87	0.05	0.84	0.97	6756
14	0.79	0.15	0.68	1.11	4759	0.80	0.07	0.75	0.96	6534
15	1.16	0.11	1.08	1.41	5749	1.13	0.07	1.07	1.30	6270
16	0.88	0.26	0.69	1.46	4170	0.68	0.09	0.61	0.89	5422
17	1.10	0.25	0.92	1.62	4244	0.87	0.12	0.78	1.15	6756
18	0.66	0.07	0.61	0.83	4616	0.83	0.04	0.80	0.92	7165
19	1.03	0.38	0.78	1.64	266	0.76	0.39	0.48	1.58	3601
20	1.04	0.29	0.82	1.68	1375	0.91	0.33	0.65	1.65	3931
21	0.93	0.30	0.71	1.57	3694	0.69	0.13	0.59	1.00	4494
pop	1.02	0.09	0.96	1.20	2278	0.93	0.09	0.87	1.11	5271

Parámetro ID	$\lambda$					$\kappa$				
	Media	SD	2.5%	97.5%	n.eff	Media	SD	2.5%	97.5%	n.eff
1	3.60	0.26	3.40	4.17	757	0.07	0.07	0.02	0.26	7055
2	3.67	0.28	3.47	4.25	535	0.04	0.04	0.01	0.16	6995
3	3.70	0.26	3.53	4.23	586	0.03	0.03	0.00	0.12	7500
4	3.25	0.32	3.04	3.91	1013	0.15	0.14	0.04	0.50	6262
5	3.30	0.29	3.10	3.92	1416	0.10	0.10	0.03	0.38	6989
6	3.45	0.28	3.27	4.02	1020	0.06	0.06	0.01	0.22	7500
7	3.50	0.28	3.30	4.08	1039	0.05	0.05	0.01	0.21	7500
8	3.24	0.29	3.05	3.79	1001	0.14	0.13	0.04	0.50	6961
9	3.31	0.29	3.12	3.87	1566	0.07	0.07	0.02	0.25	6857
10	3.29	0.28	3.09	3.84	1144	0.07	0.07	0.02	0.29	7072
11	3.61	0.25	3.44	4.12	1019	0.02	0.02	0.00	0.09	7212
12	3.38	0.28	3.20	3.96	1353	0.05	0.05	0.01	0.18	7109
13	3.62	0.27	3.43	4.19	615	0.04	0.04	0.01	0.16	6636
14	3.41	0.27	3.23	3.96	1211	0.05	0.05	0.01	0.22	7306
15	3.81	0.27	3.60	3.45	141	0.02	0.02	0.00	0.10	7212
16	3.19	0.33	2.98	3.80	358	0.17	0.15	0.06	0.56	6598
17	3.22	0.30	3.02	3.82	570	0.13	0.12	0.04	0.22	6543
18	3.46	0.27	3.27	4.01	1336	0.04	0.04	0.01	0.15	7160
19	3.02	0.38	2.74	3.82	52	0.37	0.30	0.13	1.16	3949
20	3.10	0.32	2.87	3.73	90	0.36	0.29	0.14	1.09	4429
21	3.19	0.30	2.98	3.77	262	0.24	0.21	0.08	0.79	6013
pop	3.40	0.13	3.30	3.67	152	0.11	0.14	0.03	0.40	6595

Cuadro 7.6: Estimación para todos los parámetros del Modelo I para cada uno de los animales. Este ajuste corresponde a los resultados obtenidos con el segundo paso del algoritmo de jerarquización. Nivel poblacional.

Modelo II

Parámetro ID	$\alpha$				$\beta$					
	Media	SD	2.5 %	97.5 %	n_eff	Media	SD	2.5 %	97.5 %	n_eff
1	2.22	0.22	2.07	2.67	1451	1.37	0.16	1.26	1.71	4399
2	1.60	0.15	1.49	1.93	1983	1.01	0.06	0.97	1.14	6598
3	1.72	0.11	1.64	1.96	4534	1.47	0.09	1.41	1.66	4588
4	2.02	0.28	1.82	2.60	1770	1.21	0.21	1.05	1.67	5931
5	2.16	0.29	1.99	2.78	277	0.98	0.14	0.88	1.28	5561
6	2.49	0.22	2.32	2.98	108	1.15	0.10	1.08	1.36	6539
7	1.53	0.22	1.37	1.98	944	0.77	0.04	0.73	0.86	5886
8	2.00	0.24	1.83	2.49	3371	1.48	0.21	1.33	1.93	2445
9	2.09	0.22	1.94	2.54	1907	1.37	0.15	1.26	1.71	4211
10	1.85	0.20	1.72	2.26	3856	1.33	0.13	1.24	1.62	4952
11	2.03	0.13	1.94	2.31	5034	1.29	0.06	1.25	1.43	6236
12	2.18	0.20	2.05	2.60	2099	1.37	0.12	1.28	1.63	5139
13	1.88	0.17	1.76	2.24	4223	1.06	0.06	1.02	1.20	6659
14	1.74	0.22	1.59	2.18	2752	0.99	0.10	0.92	1.20	5874
15	1.66	0.12	1.57	1.91	3490	1.29	0.08	1.23	1.46	6365
16	1.89	0.29	1.70	2.45	2224	0.87	0.14	0.77	1.20	4027
17	1.85	0.27	1.66	2.38	2997	1.01	0.14	0.90	1.32	5804
18	1.57	0.17	1.44	1.93	1777	0.99	0.06	0.94	1.13	6378
19	1.91	0.37	1.77	2.60	110	1.03	0.35	0.79	1.71	3009
20	1.90	0.33	1.71	2.52	368	1.06	0.34	0.82	1.76	3978
21	1.81	0.32	1.59	2.45	1080	0.78	0.15	0.66	1.12	3224
pop	1.91	0.10	1.84	2.11	473	1.14	0.08	1.08	1.31	4590

Parámetro	$\lambda$					$\kappa$				
	Media	SD	2.5%	97.5%	n_eff	Media	SD	2.5%	97.5%	n_eff
1	2.77	0.29	2.57	3.36	1116	0.14	0.13	0.04	0.49	7221
2	2.81	0.29	2.60	3.42	1251	0.09	0.09	0.02	0.33	7500
3	2.87	0.29	2.66	3.45	1282	0.08	0.08	0.02	0.32	6640
4	2.58	0.31	2.37	3.21	1083	0.21	0.18	0.07	0.69	6696
5	2.59	0.31	2.38	3.20	1237	0.17	0.15	0.05	0.59	5743
6	2.70	0.29	2.50	3.29	869	0.12	0.11	0.03	0.43	6878
7	2.91	0.28	2.73	3.46	853	0.08	0.07	0.02	0.29	6953
8	2.60	0.33	2.38	3.26	734	0.19	0.17	0.06	0.65	6869
9	2.66	0.31	2.45	3.27	1561	0.10	0.10	0.03	0.36	7020
10	2.66	0.29	2.44	3.29	1286	0.09	0.09	0.03	0.34	6622
11	2.74	0.30	2.53	3.33	1529	0.08	0.08	0.02	0.30	7322
12	2.73	0.30	2.52	3.33	1193	0.08	0.08	0.02	0.31	6990
13	2.86	0.29	2.65	3.48	1061	0.08	0.08	0.02	0.30	6955
14	2.73	0.28	2.53	3.31	1424	0.10	0.09	0.03	0.37	6611
15	3.08	0.33	2.85	3.66	308	0.06	0.06	0.01	0.22	7002
16	2.55	0.32	2.31	3.21	858	0.21	0.19	0.07	0.69	6318
17	2.58	0.29	2.37	3.20	1153	0.16	0.15	0.05	0.57	6269
18	2.78	0.33	2.59	3.39	1358	0.07	0.07	0.02	0.28	6704
19	2.42	0.36	2.15	3.15	179	0.40	0.30	0.17	1.13	3849
20	2.42	0.33	2.21	3.04	165	0.42	0.31	0.17	1.15	3845
21	2.55	0.32	2.33	3.19	801	0.26	0.21	0.09	0.79	6004
pop	2.70	0.15	2.59	3.00	224	0.15	0.13	0.08	0.40	6388

Parámetro ID	$q$				
	Media	SD	2.5 %	97.5 %	n_eff
1	0.67	0.05	0.63	0.77	7108
2	0.67	0.03	0.65	0.74	7253
3	0.70	0.03	0.67	0.76	7452
4	0.48	0.09	0.41	0.66	7500
5	0.56	0.06	0.51	0.69	7383
6	0.62	0.04	0.59	0.71	7176
7	0.38	0.04	0.35	0.47	7060
8	0.48	0.09	0.42	0.66	7096
9	0.42	0.05	0.38	0.53	6957
10	0.27	0.06	0.22	0.40	6473
11	0.80	0.02	0.78	0.84	7243
12	0.49	0.04	0.45	0.58	7826
13	0.54	0.04	0.51	0.62	7500
14	0.54	0.05	0.50	0.64	7500
15	0.57	0.03	0.54	0.64	7113
16	0.32	0.09	0.25	0.50	6157
17	0.25	0.07	0.20	0.40	6127
18	0.60	0.04	0.57	0.68	7488
19	0.55	0.20	0.40	0.91	6651
20	0.61	0.18	0.49	0.93	6510
21	0.18	0.11	0.09	0.42	5226
pop	0.51	0.11	0.44	0.74	7090

Cuadro 7.7: Estimación para todos los parámetros del Modelo II para cada uno de los animales. Este ajuste corresponde a los resultados obtenidos con el segundo paso del algoritmo de jerarquización. Nivel poblacional.

## Modelo III

Parámetro	$\alpha$						$\beta$					
	ID	Media	SD	2.5%	97.5%	n_eff	Media	SD	2.5%	97.5%	n_eff	
1	2.29	0.22	2.15	2.74	1538	1.39	0.16	1.28	1.72	4288		
2	1.64	0.16	1.53	1.97	2108	1.02	0.06	0.98	1.14	6834		
3	1.75	0.11	1.67	1.99	3722	1.47	0.09	1.41	1.66	4839		
4	2.11	0.27	1.92	2.68	1792	1.25	0.22	1.07	1.70	5421		
5	2.21	0.31	2.01	2.83	338	0.98	0.14	0.88	1.30	5197		
6	2.57	0.22	2.40	3.04	165	1.15	0.09	1.09	1.36	6850		
7	1.71	0.21	1.56	2.16	1734	0.79	0.05	0.76	0.90	5573		
8	2.09	0.24	1.93	2.60	3356	1.49	0.21	1.35	1.93	2360		
9	2.21	0.23	2.05	2.68	2146	1.42	0.16	1.30	1.76	3932		
10	1.97	0.20	1.83	2.38	4716	1.38	0.14	1.28	1.69	4799		
11	2.03	0.14	1.93	2.32	5640	1.29	0.06	1.24	1.43	6762		
12	2.29	0.21	2.14	2.73	1597	1.40	0.12	1.31	1.67	4721		
13	1.98	0.17	1.86	2.32	4072	1.08	0.06	1.03	1.22	6970		
14	1.80	0.22	1.65	2.25	2753	0.99	0.09	0.92	1.19	5906		
15	1.72	0.13	1.63	1.97	3311	1.30	0.09	1.23	1.49	6128		
16	1.96	0.30	1.75	2.55	2214	0.88	0.14	0.78	1.20	3917		
17	1.93	0.28	1.74	2.49	2958	1.03	0.15	0.92	1.34	5784		
18	1.62	0.18	1.49	2.02	1317	0.99	0.06	0.95	1.13	6190		
19	2.04	0.37	1.77	2.76	142	1.06	0.34	0.82	1.75	2961		
20	2.02	0.38	1.78	2.70	321	1.08	0.32	0.85	1.70	3543		
21	1.90	0.31	1.69	2.53	1457	0.80	0.16	0.69	1.17	3502		
pop	1.99	0.10	1.92	2.20	526	1.15	0.08	1.10	1.32	4698		

Parámetro ID	$\lambda$				$\kappa$					
	Media	SD	2.5 %	97.5 %	n_eff	Media	SD	2.5 %	97.5 %	n_eff
1	2.68	0.30	2.47	3.30	1334	0.15	0.14	0.05	0.52	6735
2	2.75	0.29	2.56	3.34	998	0.09	0.09	0.02	0.34	6810
3	2.79	0.28	2.60	3.36	1295	0.09	0.09	0.02	0.35	6785
4	2.51	0.32	2.28	3.14	1173	0.23	0.19	0.08	0.72	6098
5	2.51	0.31	2.30	3.12	1120	0.17	0.15	0.05	0.58	6562
6	2.66	0.30	2.45	3.25	1005	0.12	0.11	0.03	0.43	6686
7	2.81	0.30	2.60	3.40	769	0.08	0.08	0.02	0.31	6930
8	2.51	0.30	2.31	3.11	851	0.20	0.17	0.07	0.67	6956
9	2.58	0.31	2.38	3.21	1197	0.12	0.11	0.03	0.42	6599
10	2.59	0.31	2.38	3.20	1545	0.11	0.10	0.03	0.40	6985
11	2.69	0.29	2.48	3.27	1344	0.07	0.07	0.02	0.28	6857
12	2.65	0.29	2.45	3.22	1317	0.09	0.09	0.02	0.34	6907
13	2.77	0.29	2.55	3.39	962	0.08	0.08	0.02	0.31	7158
14	2.65	0.30	2.44	3.27	968	0.11	0.11	0.03	0.41	6968
15	2.98	0.30	2.76	3.58	351	0.07	0.07	0.02	0.27	6970
16	2.46	0.32	2.25	3.06	809	0.22	0.19	0.07	0.70	5372
17	2.52	0.32	2.28	3.15	1150	0.17	0.15	0.05	0.59	7924
18	2.69	0.29	2.49	3.31	1210	0.07	0.07	0.02	0.27	6985
19	2.36	0.34	2.18	2.97	168	0.40	0.31	0.16	1.17	3668
20	2.34	0.37	2.09	3.11	179	0.42	0.31	0.17	1.17	3829
21	2.46	0.33	2.23	3.11	1010	0.26	0.22	0.09	0.84	5829
pop	2.62	0.15	2.51	2.91	236	0.16	0.12	0.08	0.40	6557

Parámetro	$\rho$					$\epsilon$				
	Media	SD	2.5%	97.5%	n.eff	Media	SD	2.5%	97.5%	n.eff
1	11.97	1.14	11.23	15.49	3	0.87	0.24	0.68	1.34	385
2	11.96	1.18	11.21	15.58	3	0.30	0.13	0.20	0.57	5639
3	11.97	1.20	11.24	15.73	3	0.52	0.15	0.42	0.84	5161
4	11.95	1.19	11.24	15.82	5	0.60	0.24	0.43	1.08	1514
5	11.96	1.18	11.25	15.68	4	0.31	0.18	0.17	0.70	5209
6	11.96	1.17	11.24	15.46	3	0.04	0.04	0.01	0.16	3899
7	11.96	1.20	11.18	15.43	4	0.07	0.13	0.00	0.50	2828
8	11.95	1.18	11.28	15.64	3	0.59	0.23	0.44	1.10	1761
9	11.96	1.19	11.24	15.58	3	0.19	0.16	0.05	0.58	3655
10	12.11	1.56	11.19	18.52	10	0.52	0.25	0.35	1.01	2927
11	6.17	0.72	5.55	6.83	7	0.11	0.03	0.09	0.20	5147
12	11.97	1.20	11.22	15.54	4	0.28	0.19	0.10	0.67	4388
13	11.98	1.20	11.23	15.83	4	0.06	0.06	0.01	0.25	3633
14	11.98	1.19	11.27	15.62	1	0.43	0.19	0.29	0.82	4899
15	12.03	1.25	11.24	15.18	6	0.43	0.18	0.31	0.79	5788
16	11.97	1.19	11.13	15.68	4	0.61	0.23	0.45	1.09	2326
17	16.96	2.01	15.84	22.49	4	0.90	0.23	0.77	1.34	317
18	11.96	1.21	11.23	15.58	3	0.03	0.03	0.01	0.11	4559
19	11.94	1.21	11.21	15.43	3	0.41	0.24	0.23	0.95	3662
20	11.97	1.21	11.26	15.59	4	0.39	0.24	0.21	0.93	3676
21	11.96	1.17	11.23	15.54	4	0.74	0.24	0.57	1.22	586
pop	11.94	1.15	11.24	15.59	3	0.40	0.08	0.34	0.56	2622

Cuadro 7.8: Estimación para todos los parámetros del Modelo III para cada uno de los animales. Este ajuste corresponde a los resultados obtenidos con el segundo paso del algoritmo de jerarquización. Nivel poblacional.

Modelo IV

Parámetro ID	$\alpha$				$\beta$					
	Media	SD	2.5%	97.5%	n_eff	Media	SD	2.5%	97.5%	n_eff
1	2.35	0.22	2.19	2.81	537	1.50	0.17	1.37	1.86	3040
2	1.47	0.16	1.36	1.83	1417	1.01	0.06	0.96	1.13	6371
3	1.64	0.11	1.57	1.86	3976	1.48	0.09	1.42	1.66	4324
4	1.98	0.29	1.77	2.56	2082	1.19	0.21	1.03	1.63	5889
5	2.11	0.26	1.92	2.75	71	1.09	0.17	0.97	1.46	5961
6	2.44	0.23	2.26	2.93	188	1.16	0.09	1.09	1.36	6729
7	1.68	0.22	1.53	2.14	2208	0.80	0.05	0.76	0.91	5430
8	1.96	0.24	1.80	2.44	3836	1.48	0.21	1.33	1.94	2588
9	2.14	0.23	1.99	2.60	1679	1.42	0.16	1.30	1.77	3791
10	1.88	0.21	1.73	2.31	4554	1.35	0.14	1.24	1.66	5175
11	1.93	0.13	1.83	2.21	5939	1.32	0.06	1.27	1.45	6343
12	2.16	0.20	2.02	2.56	2046	1.37	0.12	1.29	1.63	5135
13	1.94	0.19	1.81	2.33	4917	1.07	0.07	1.02	1.22	6729
14	1.73	0.22	1.57	2.19	3001	1.00	0.10	0.93	1.21	5187
15	1.62	0.12	1.53	1.89	2492	1.30	0.08	1.24	1.48	6182
16	1.87	0.29	1.68	2.47	2005	0.87	0.15	0.76	1.24	4008
17	1.86	0.27	1.68	2.39	3283	1.03	0.14	0.92	1.34	5897
18	1.59	0.18	1.46	1.98	2122	1.02	0.06	0.98	1.16	7840
19	1.96	0.37	1.68	2.55	100	1.07	0.35	0.84	1.77	2925
20	1.92	0.30	1.69	2.50	279	1.08	0.33	0.86	1.75	3282
21	1.84	0.31	1.62	2.43	1221	0.81	0.16	0.70	1.18	3468
pop	1.91	0.09	1.84	2.11	512	1.16	0.09	1.10	1.34	4384

Parámetro	$\lambda$					$\kappa$				
	Media	SD	2.5%	97.5%	n_eff	Media	SD	2.5%	97.5%	n_eff
1	2.99	0.29	2.79	3.60	1357	0.13	0.12	0.04	0.45	6829
2	3.05	0.29	2.85	3.64	994	0.07	0.07	0.02	0.27	6940
3	3.13	0.27	2.94	3.67	652	0.06	0.06	0.01	0.24	6691
4	2.76	0.31	2.55	3.37	1141	0.22	0.19	0.07	0.71	6454
5	2.75	0.31	2.55	3.34	1136	0.18	0.16	0.06	0.60	6456
6	2.88	0.30	2.68	3.51	902	0.12	0.11	0.03	0.43	6795
7	3.02	0.28	2.83	3.64	820	0.08	0.08	0.02	0.32	7175
8	2.76	0.31	2.55	3.35	958	0.19	0.17	0.06	0.63	6647
9	2.80	0.30	2.61	3.40	1511	0.11	0.11	0.03	0.41	6932
10	2.82	0.31	2.61	3.42	1444	0.09	0.09	0.02	0.36	7500
11	2.97	0.28	2.55	3.77	1093	0.06	0.06	0.01	0.22	6852
12	2.87	0.29	2.67	3.44	1360	0.08	0.08	0.02	0.31	7028
13	3.05	0.29	2.85	3.65	1009	0.08	0.07	0.02	0.29	6913
14	2.92	0.28	2.72	3.47	1421	0.09	0.09	0.03	0.35	6966
15	3.38	0.23	3.20	3.87	202	0.06	0.06	0.01	0.25	6758
16	2.72	0.32	2.52	3.32	604	0.20	0.18	0.06	0.69	6407
17	2.74	0.30	2.53	3.31	1043	0.17	0.15	0.05	0.57	6963
18	2.93	0.29	2.73	3.54	1259	0.07	0.07	0.02	0.27	6992
19	2.65	0.32	2.44	3.19	83	0.39	0.30	0.16	1.13	3858
20	2.61	0.34	2.36	3.37	114	0.41	0.30	0.17	1.11	3854
21	2.68	0.32	2.46	3.31	554	0.25	0.21	0.08	0.80	5525
pop	2.88	0.14	2.78	3.19	270	0.15	0.13	0.07	0.40	6684

Parámetro ID	$\rho$				$\epsilon$					
	Media	SD	2.5%	97.5%	n_eff	Media	SD	2.5%	97.5%	n_eff
1	25.87	0.55	25.66	26.99	46	0.64	0.24	0.48	1.13	1398
2	25.91	0.52	25.66	26.94	49	0.34	0.19	0.18	0.73	4456
3	25.97	0.77	25.76	26.66	69	0.39	0.22	0.22	0.82	4672
4	25.94	0.63	25.72	26.55	61	0.60	0.23	0.45	1.08	1724
5	25.97	0.61	25.67	26.85	37	0.30	0.18	0.16	0.70	5598
6	25.93	0.69	25.86	26.75	67	0.05	0.05	0.01	0.18	3733
7	25.85	0.60	25.75	26.52	37	0.14	0.24	0.00	0.84	3000
8	25.88	0.50	25.77	26.62	42	0.65	0.24	0.48	1.16	1206
9	25.94	0.50	25.78	26.74	82	0.27	0.21	0.07	0.73	4302
10	26.67	0.89	26.56	27.02	125	0.72	0.27	0.57	1.22	683
11	26.17	1.21	26.02	29.43	50	0.06	0.03	0.04	0.15	4904
12	25.90	0.57	25.67	26.57	50	0.29	0.23	0.07	0.75	4557
13	25.92	0.62	25.80	26.64	63	0.04	0.07	0.00	0.26	2898
14	25.95	0.53	25.75	26.70	83	0.38	0.21	0.22	0.79	5567
15	25.97	0.81	25.78	26.94	34	0.32	0.31	0.02	0.93	4104
16	25.94	0.50	25.80	26.64	59	0.65	0.22	0.50	1.11	1590
17	25.97	0.42	25.78	26.70	56	0.84	0.25	0.68	1.30	205
18	25.87	0.58	25.82	26.38	29	0.03	0.03	0.00	0.13	3220
19	25.93	0.69	25.70	26.76	31	0.41	0.24	0.22	0.94	3537
20	25.56	0.47	25.18	25.99	8	0.38	0.24	0.19	0.90	2560
21	25.88	0.51	25.76	26.57	41	0.70	0.26	0.50	1.25	534
pop	25.95	0.40	25.88	26.40	41	0.39	0.09	0.33	0.57	1653

Parámetro	$\nu$					$\theta$				
	ID	Media	SD	2.5%	97.5%	n.eff	Media	SD	2.5%	97.5%
1	11.00	0.53	10.66	12.16	62	0.51	0.25	0.32	1.07	3748
2	11.03	0.55	10.64	12.20	36	0.23	0.15	0.11	0.60	6073
3	11.00	0.52	10.65	12.21	43	0.16	0.10	0.09	0.41	6205
4	10.99	0.58	10.62	12.15	35	0.32	0.24	0.13	0.90	5354
5	11.04	0.54	10.70	12.40	36	0.57	0.27	0.36	1.16	2778
6	11.01	0.54	10.69	12.28	46	0.27	0.19	0.12	0.73	6435
7	11.03	0.55	10.67	12.25	35	0.37	0.22	0.20	0.90	5263
8	11.00	0.48	10.67	12.14	33	0.24	0.20	0.08	0.73	5677
9	11.07	0.56	10.72	12.28	37	0.12	0.12	0.03	0.44	5573
10	11.05	0.53	10.69	12.21	39	0.21	0.18	0.07	0.70	5843
11	11.02	0.49	10.65	12.02	30	0.35	0.17	0.21	0.72	6139
12	11.04	0.56	10.72	12.46	37	0.19	0.15	0.07	0.58	6387
13	11.03	0.52	10.67	12.18	48	0.29	0.19	0.15	0.75	6259
14	11.04	0.56	10.68	12.28	27	0.29	0.19	0.14	0.74	6026
15	11.03	0.53	10.68	12.31	46	0.07	0.06	0.02	0.22	5729
16	11.00	0.55	10.60	12.08	36	0.38	0.27	0.16	1.02	4040
17	11.00	0.53	10.68	12.26	41	0.27	0.21	0.10	0.80	6109
18	11.04	0.57	10.67	12.34	37	0.31	0.17	0.18	0.73	6410
19	11.05	0.53	10.71	12.22	33	0.41	0.28	0.19	1.07	3583
20	11.03	0.54	10.70	12.26	50	0.39	0.27	0.18	1.05	4043
21	11.01	0.55	10.64	12.11	38	0.41	0.27	0.19	1.06	3861
pop	11.02	0.45	10.74	12.06	9	0.30	0.10	0.24	0.50	4660

Cuadro 7.9: Estimación para todos los parámetros del Modelo IV para cada uno de los animales. Este ajuste corresponde a los resultados obtenidos con el segundo paso del algoritmo de jerarquización. Nivel poblacional.

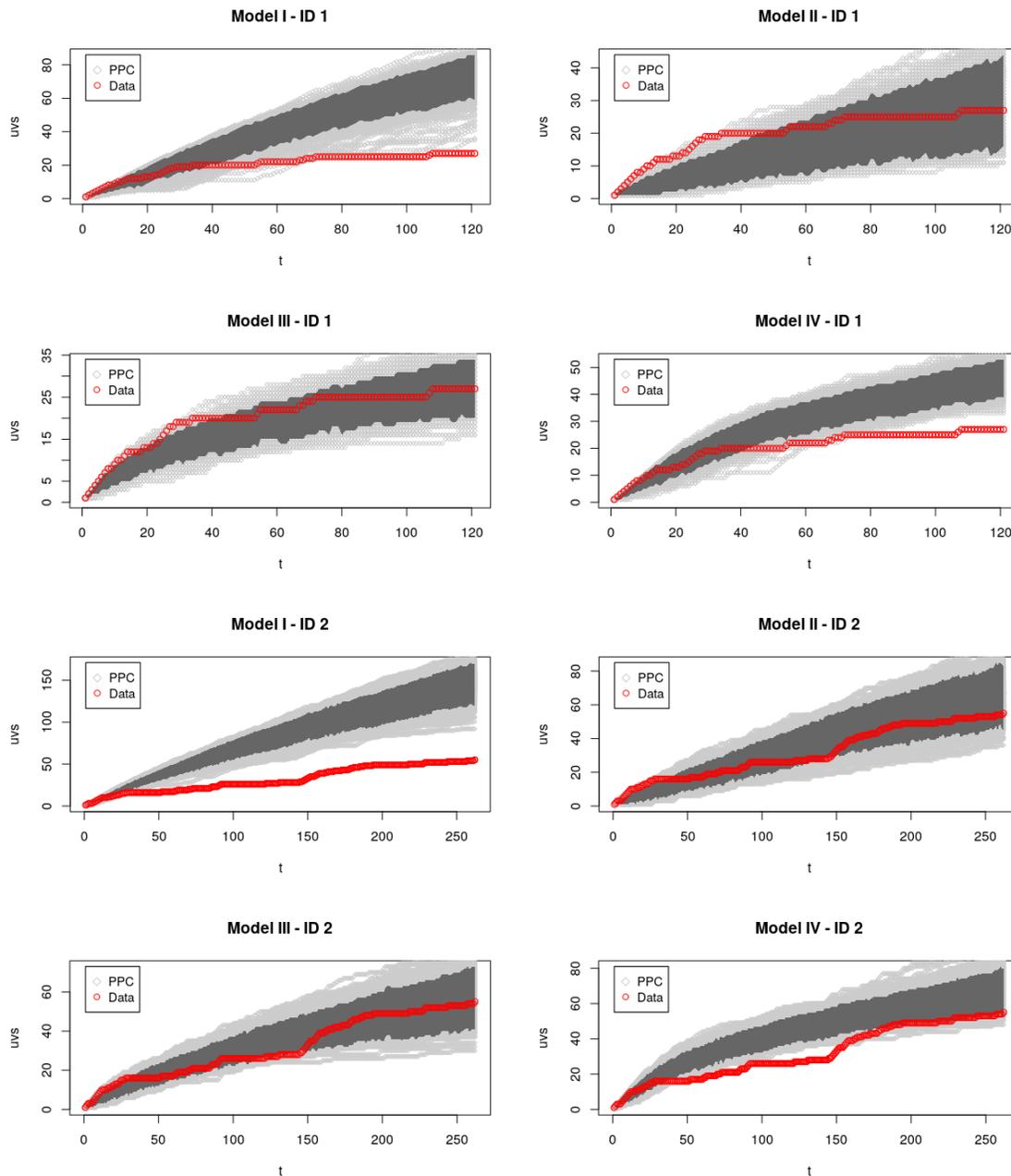
## WAIC

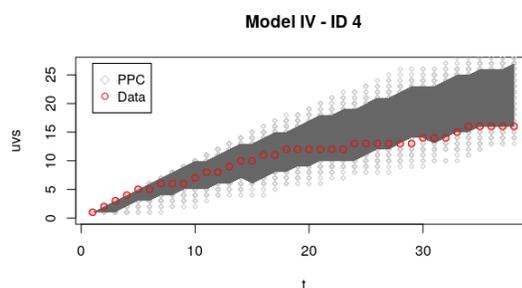
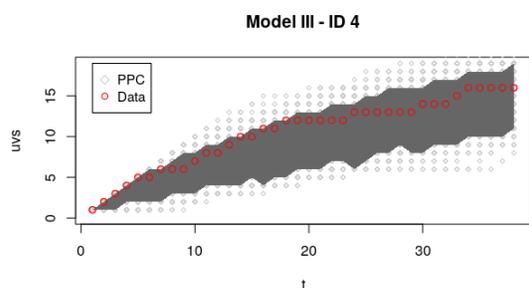
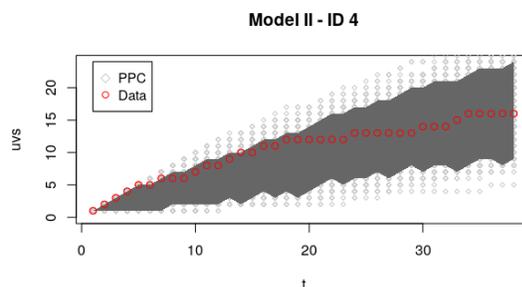
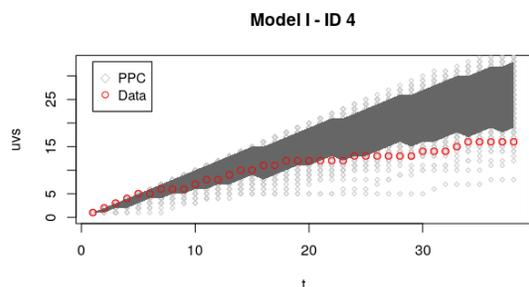
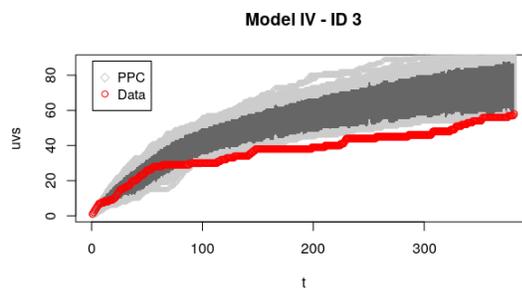
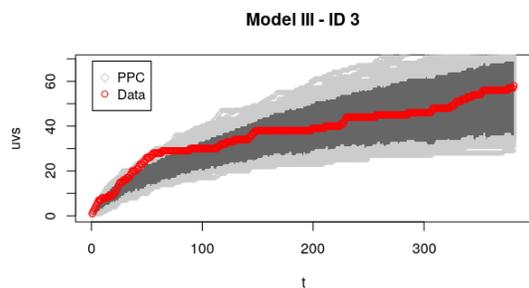
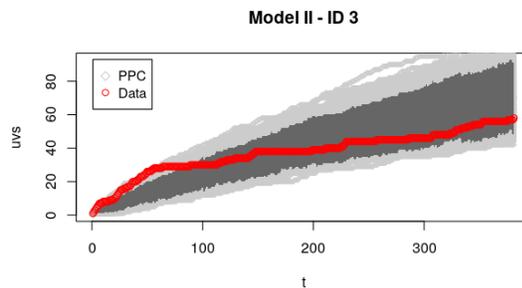
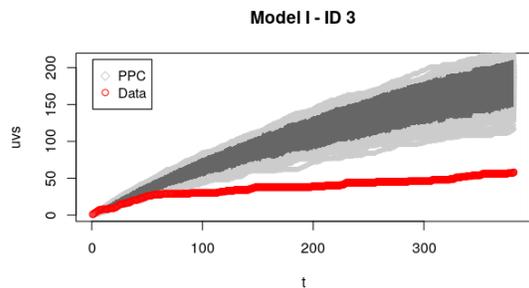
Modelo ID	Modelo I			Modelo II			Modelo III			Modelo IV		
	WAIC		P <sub>w</sub>	WAIC		P <sub>w</sub>	WAIC		P <sub>w</sub>	WAIC		P <sub>w</sub>
	Est	$\delta$	Est	$\delta$	Est	$\delta$	Est	$\delta$	Est	$\delta$	Est	
1	916.51	146.77	12.06	788.51	18.77	7.27	<b>769.74</b>	<b>0.00</b>	<b>7.56</b>	800.53	30.79	10.56
2	1864.50	220.52	8.50	1645.72	1.74	8.46	<b>1643.98</b>	<b>0.00</b>	<b>10.02</b>	1718.55	74.57	12.39
3	2751.13	340.90	5.60	<b>2410.23</b>	<b>0.00</b>	<b>5.73</b>	2413.46	3.23	11.40	2486.91	76.68	12.63
4	295.62	39.38	13.92	263.21	6.97	7.25	<b>256.24</b>	<b>0.00</b>	<b>5.40</b>	266.82	10.58	7.03
5	611.20	53.54	21.75	561.77	4.11	33.43	<b>557.66</b>	<b>0.00</b>	<b>29.81</b>	687.37	129.71	92.41
6	1161.61	176.27	21.04	994.70	9.36	31.52	<b>985.34</b>	<b>0.00</b>	<b>26.56</b>	1028.51	70.17	32.38
7	1670.40	70.08	8.93	<b>1600.32</b>	<b>0.00</b>	<b>11.42</b>	1654.47	54.15	18.00	1678.14	77.82	19.79
8	398.77	33.35	8.78	370.41	4.99	2.83	<b>365.42</b>	<b>0.00</b>	<b>1.65</b>	371.34	5.92	2.48
9	846.68	54.29	11.59	<b>792.39</b>	<b>0.00</b>	<b>9.40</b>	814.18	21.79	10.21	830.42	38.03	11.34
10	793.78	21.14	10.00	<b>772.64</b>	<b>0.00</b>	<b>5.81</b>	865.40	92.76	18.29	827.51	54.87	7.12
11	3406.23	677.88	6.71	<b>2728.35</b>	<b>0.00</b>	<b>7.34</b>	2754.21	25.86	6.56	2867.84	139.49	17.22
12	1259.08	98.78	7.88	<b>1160.30</b>	<b>0.00</b>	<b>11.72</b>	1197.11	36.81	19.09	1213.23	52.93	13.56
13	1732.02	160.37	7.32	<b>1571.65</b>	<b>0.00</b>	<b>10.58</b>	1578.03	6.38	10.05	1628.22	56.57	14.83
14	1020.30	102.70	7.39	<b>917.60</b>	<b>0.00</b>	<b>7.29</b>	923.77	6.1	7.55	984.14	66.54	9.36
15	2561.35	213.99	7.37	<b>2347.36</b>	<b>0.00</b>	<b>5.49</b>	2405.66	58.3	21.96	2446.47	99.11	17.22
16	289.98	3.03	13.39	<b>286.95</b>	<b>0.00</b>	<b>15.68</b>	291.97	5.02	14.02	299.95	13.00	20.66
17	499.80	35.39	27.65	465.15	0.74	10.51	473.24	8.83	10.20	<b>464.41</b>	<b>0.00</b>	<b>10.49</b>
18	1604.57	118.75	6.80	<b>1485.82</b>	<b>0.00</b>	<b>7.30</b>	1486.50	0.68	6.09	1539.07	53.25	9.71
19	130.98	106.64	53.14	27.41	3.07	2.84	<b>24.34</b>	<b>0.00</b>	<b>1.46</b>	27.02	2.68	2.08
20	23.02	5.23	0.78	18.05	0.26	0.69	<b>17.79</b>	<b>0.00</b>	<b>0.22</b>	19.03	1.24	0.29
21	241.79	72.34	50.91	<b>169.45</b>	<b>0.00</b>	<b>12.35</b>	176.82	7.37	13.66	176.74	7.29	15.21

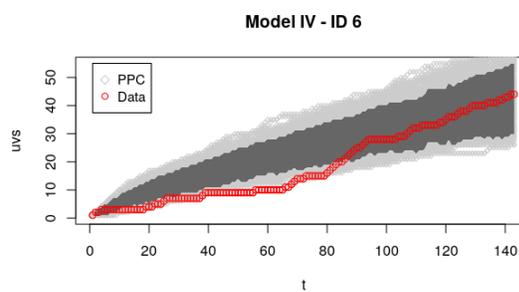
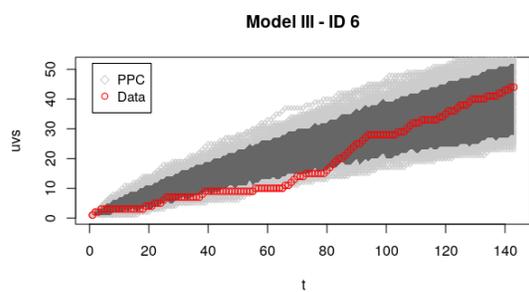
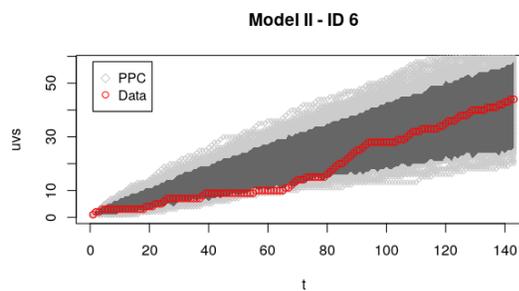
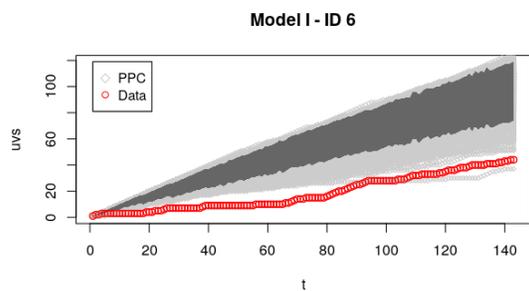
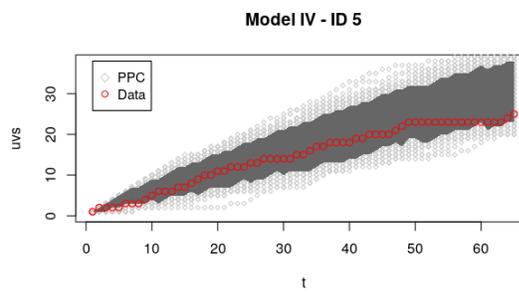
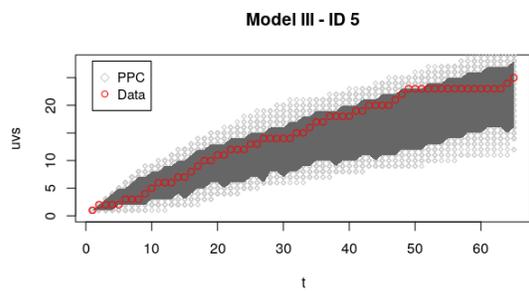
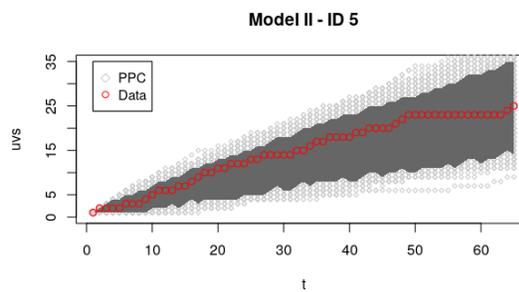
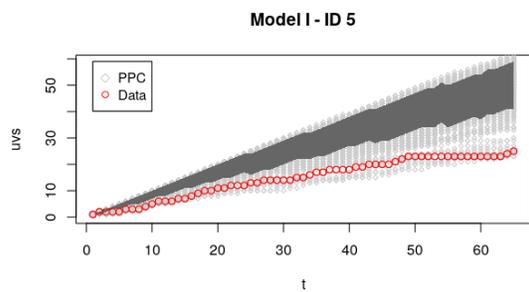
Cuadro 7.10: WAIC para el log-likelihood puntual para cada Modelo y cada animal a nivel jerárquico. La tabla muestra las estimaciones puntuales (Est) para el criterio de información WAIC, el número efectivo de parámetros (P<sub>w</sub>) y la diferencia entre los diversos valores del WAIC entre Modelos como  $\delta$ .

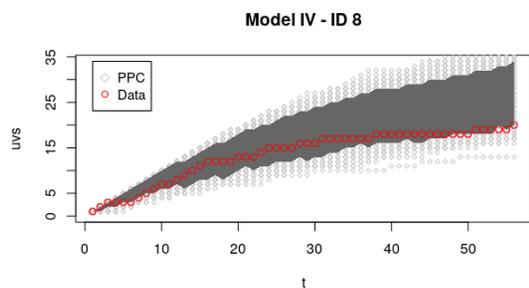
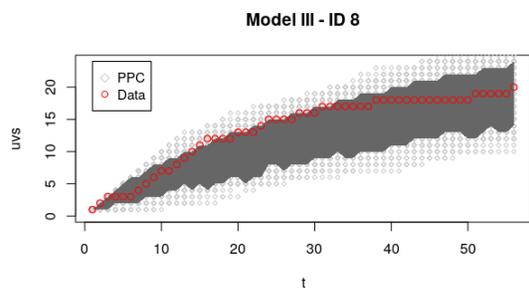
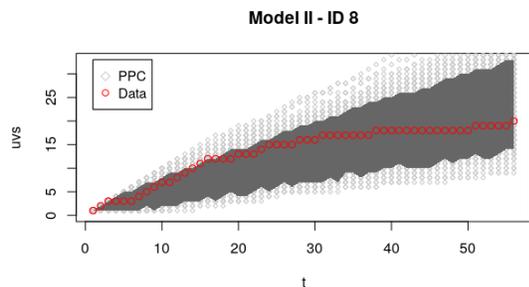
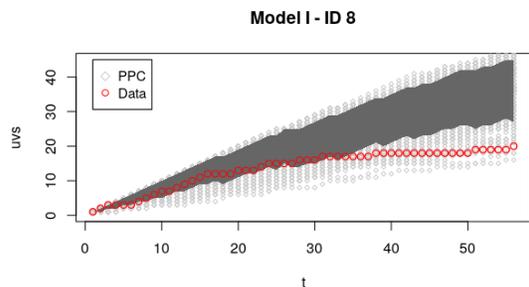
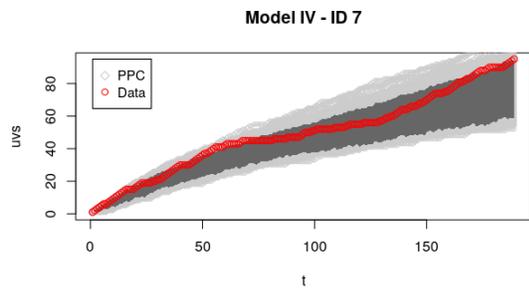
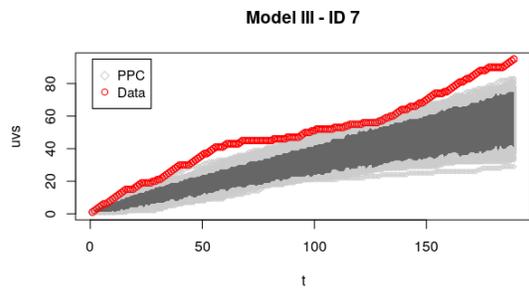
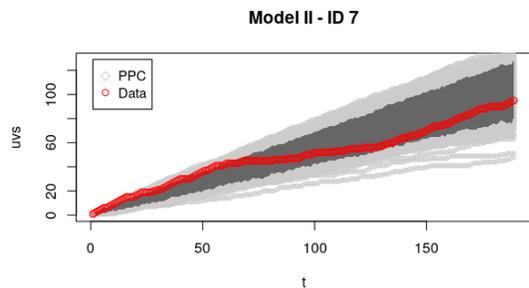
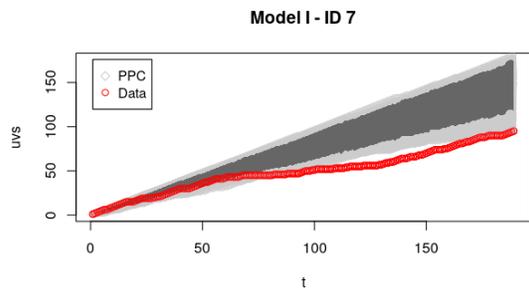
## PPC

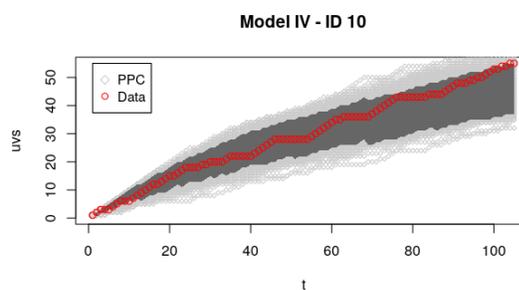
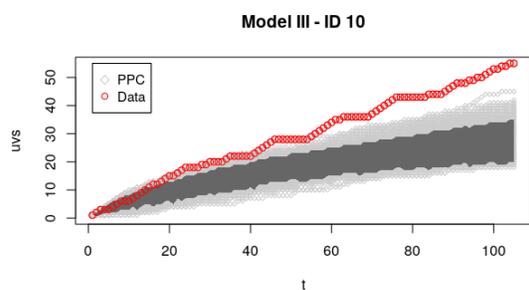
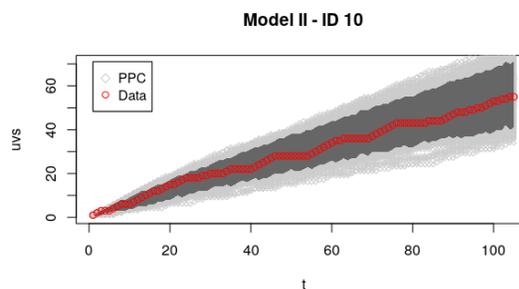
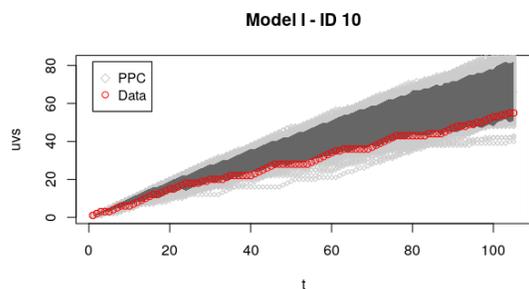
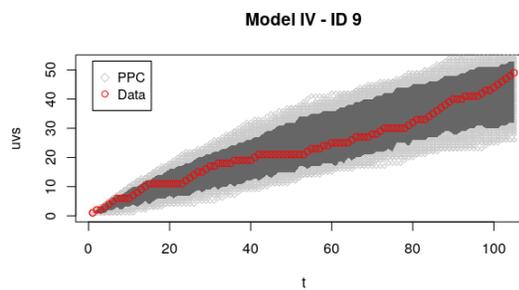
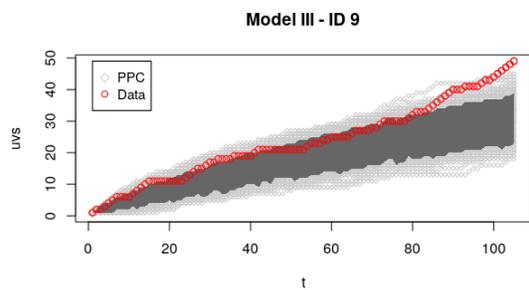
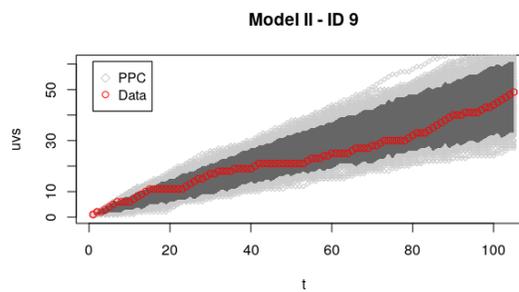
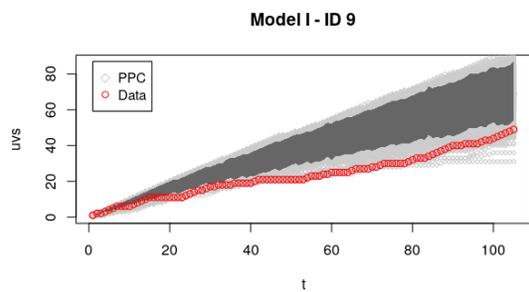
Al igual que en la sección anterior, las siguientes figuras se leen como sigue: Cada bloque de  $2 \times 2$  corresponde a la comparación entre Modelos de un solo individuo. En símbolos rojos el número de sitios únicos visitados observados, en símbolos gris claro la predicción hecha usando la técnica PPC sobre mil muestras independientes y en gris oscuro intervalo de credibilidad de dicha predicción al 95 %.

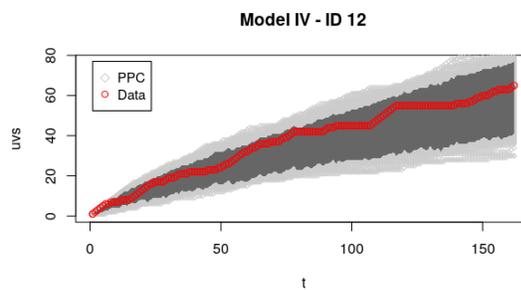
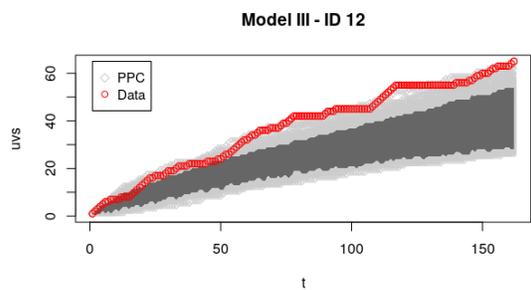
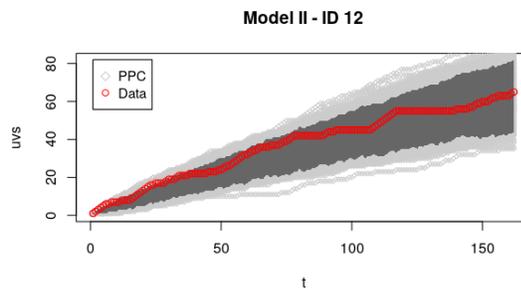
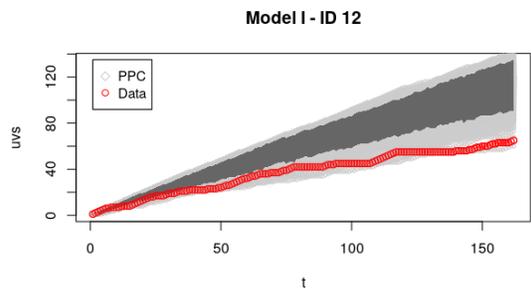
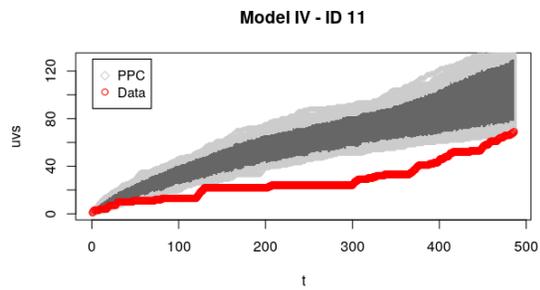
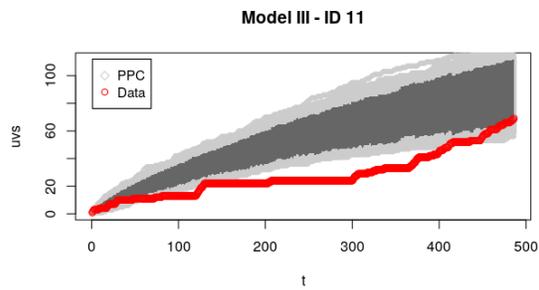
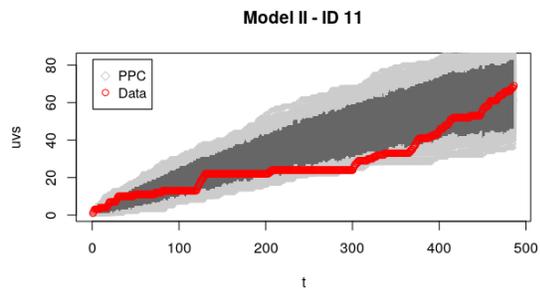
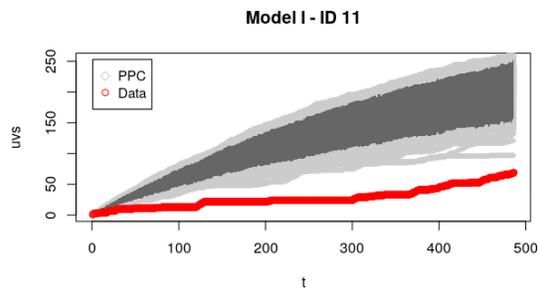




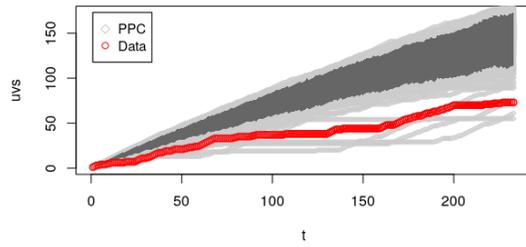




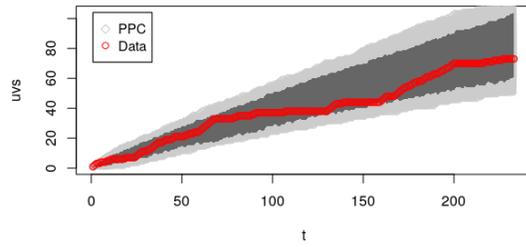




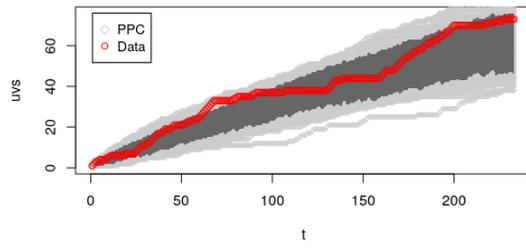
Model I - ID 13



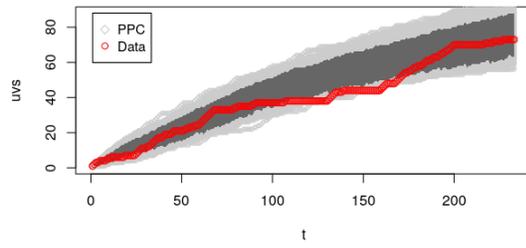
Model II - ID 13



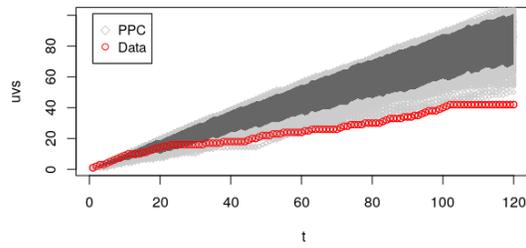
Model III - ID 13



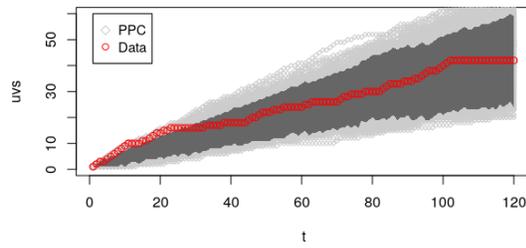
Model IV - ID 13



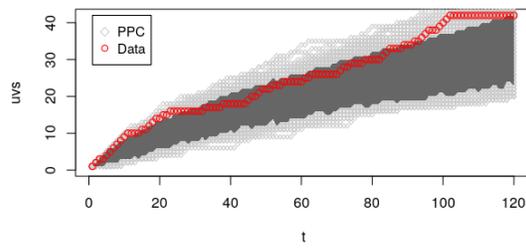
Model I - ID 14



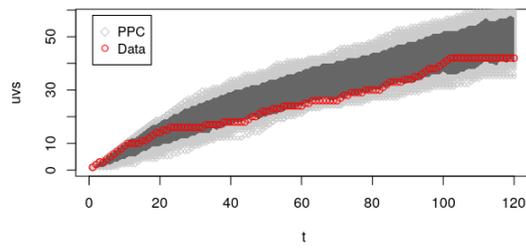
Model II - ID 14

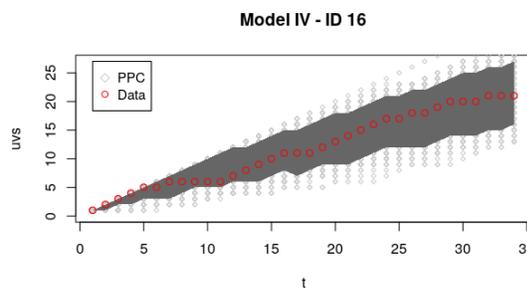
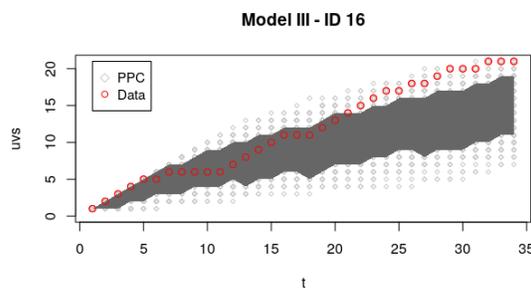
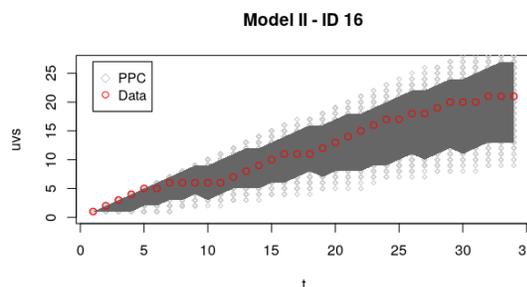
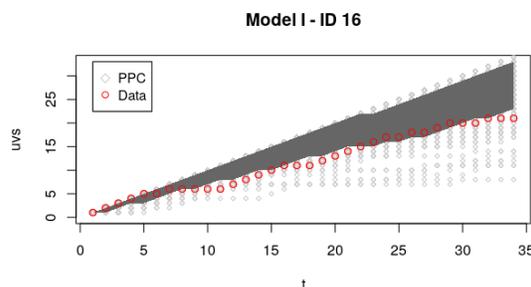
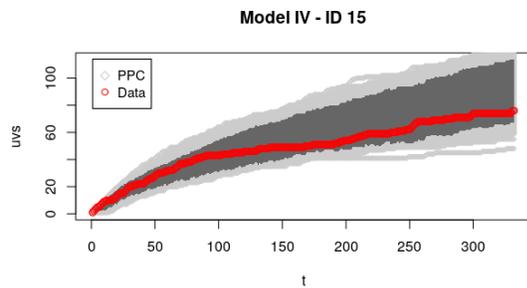
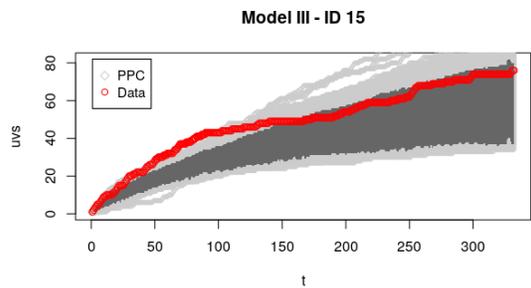
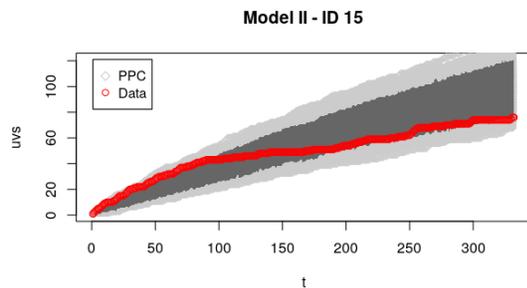
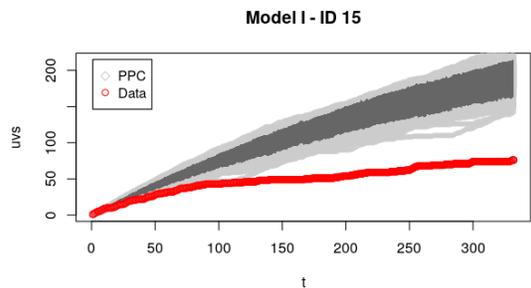


Model III - ID 14

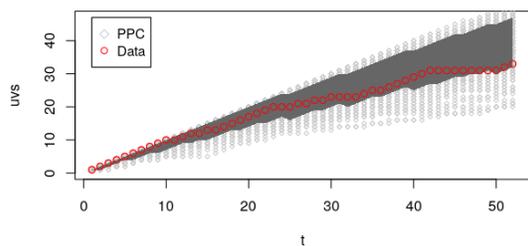


Model IV - ID 14

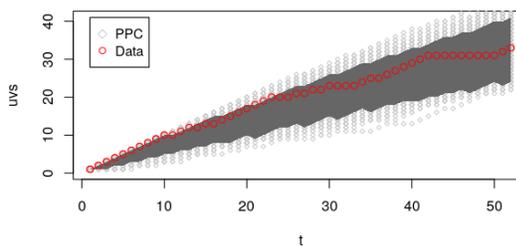




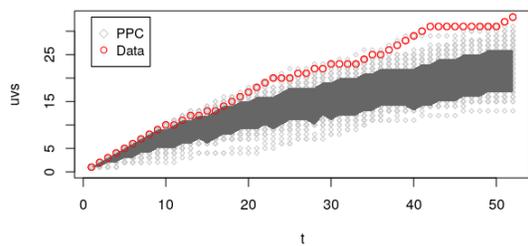
**Model I - ID 17**



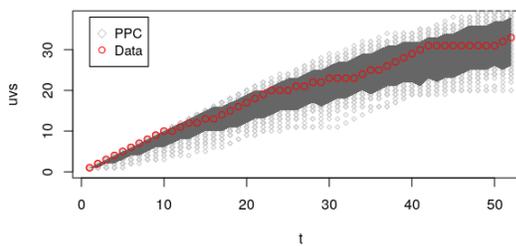
**Model II - ID 17**



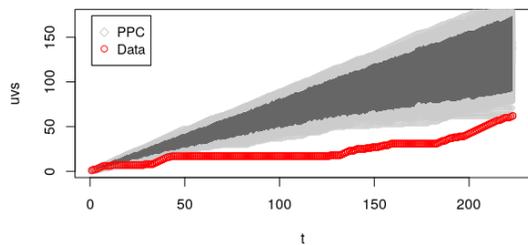
**Model III - ID 17**



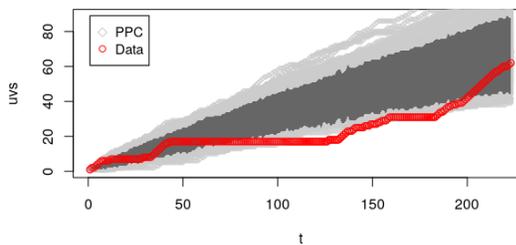
**Model IV - ID 17**



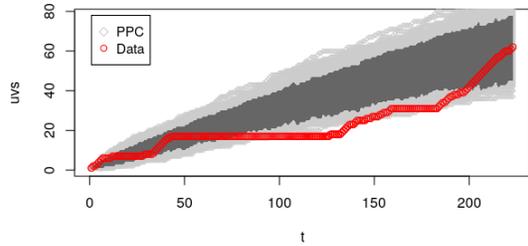
**Model I - ID 18**



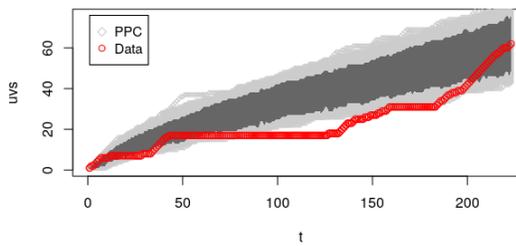
**Model II - ID 18**

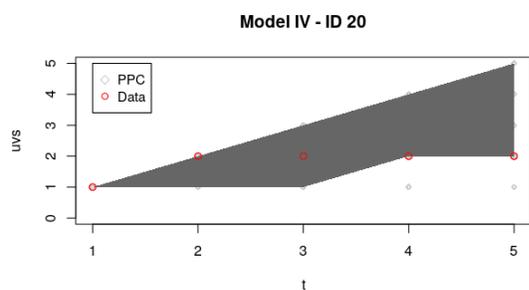
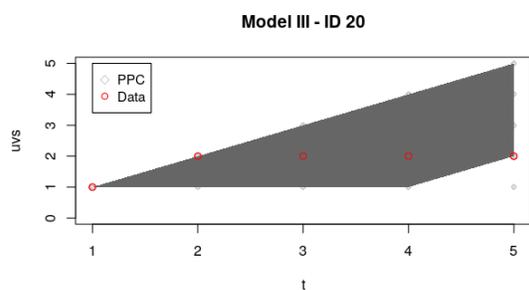
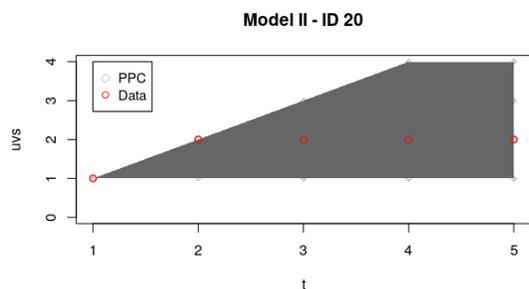
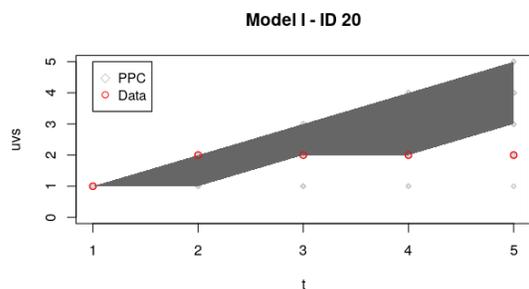
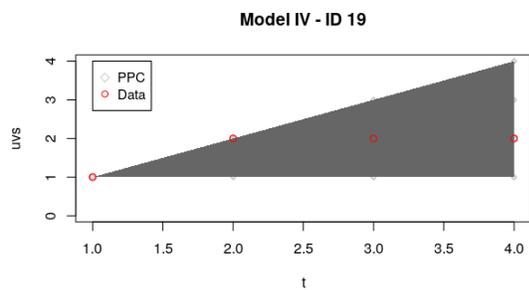
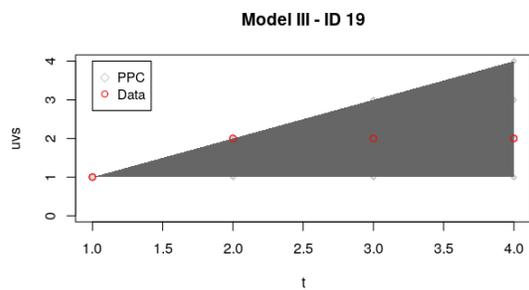
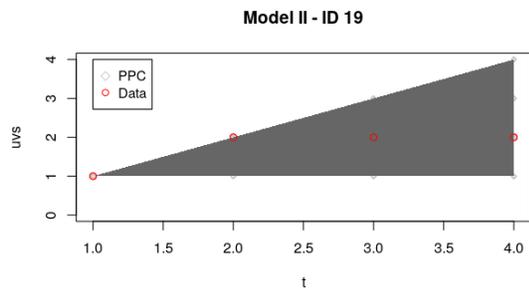
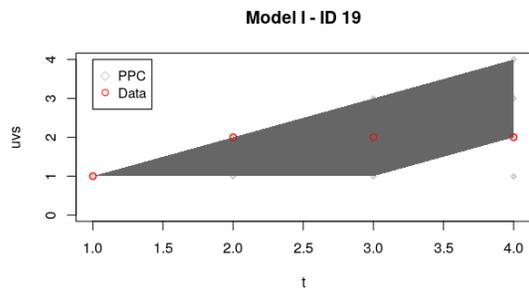


**Model III - ID 18**



**Model IV - ID 18**





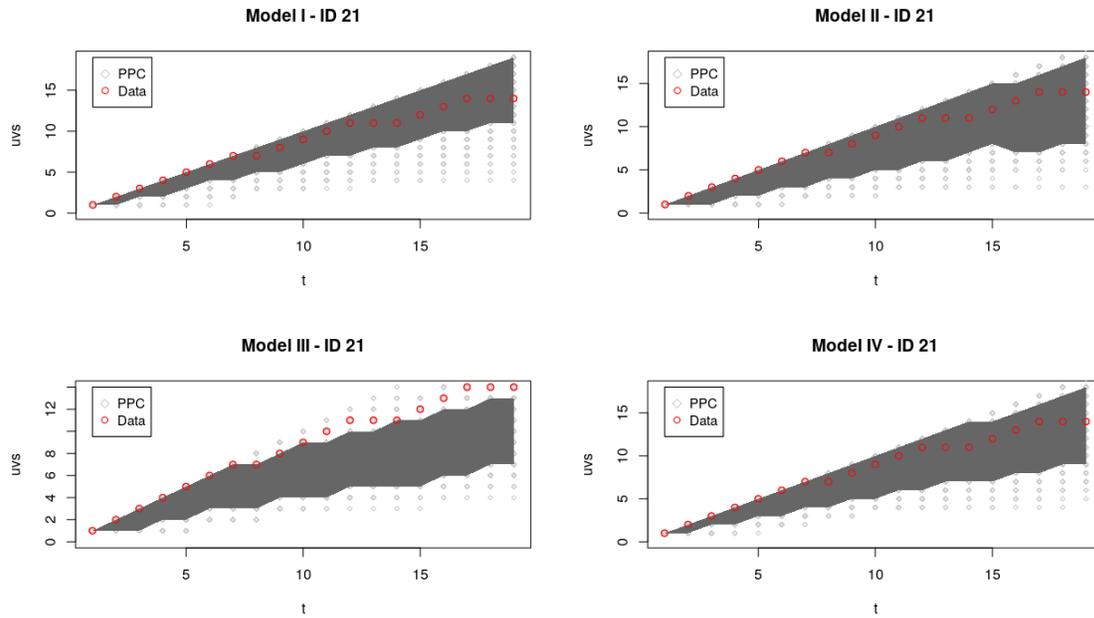


Figura 7.2: Comprobación predictiva posterior (PPC) para cada uno de los Modelos a nivel poblacional para todos los animales observados. El número de sitios únicos visitados ( $uvs$ ) como función del tiempo  $t$  (símbolos rojos). En gris claro las curvas para el PPC, en gris oscuro los intervalos de credibilidad al 95 % y en azul los datos para cada una de las trayectorias reales.

# Bibliografía

- [1] A. Falcón-Cortés, D. Boyer, L. Giuggioli, and S. N. Majumdar, “Localization transition induced by learning in random searches,” *Phys. Rev. Lett.*, vol. 119, p. 140603, 2017.
- [2] A. Falcón-Cortés, D. Boyer, and G. Ramos-Fernández, “Collective learning from individual experiences and information transfer during group foraging,” *Journal of the Royal Society Interface*, vol. 16, no. 151, p. 20180803, 2019.
- [3] D. Boyer, A. Falcón-Cortés, L. Giuggioli, and S. N. Majumdar, “Anderson-like localization transition of random walks with resetting,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2019, no. 5, p. 053204, 2019.
- [4] G. M. Viswanathan, M. G. Da Luz, E. P. Raposo, and H. E. Stanley, *The physics of foraging: an introduction to random searches and biological encounters*. Cambridge University Press, 2011.
- [5] M. F. Shlesinger, “Asymptotic solutions of continuous-time random walks,” *Journal of Statistical Physics*, vol. 10, no. 5, pp. 421–434, 1974.
- [6] P. Turchin, *Quantitative analysis of movement*. Sinauer Assoc. Inc. Sunderland MA., 1998.
- [7] R. Nathan, W. M. Getz, E. Revilla, M. Holyoak, R. Kadmon, D. Saltz, and P. E. Smouse, “A movement ecology paradigm for unifying organismal movement research,” *Proceedings of the National Academy of Sciences*, vol. 105, no. 49, pp. 19052–19059, 2008.
- [8] E. A. Codling, M. J. Plank, and S. Benhamou, “Random walk models in biology,” *Journal of the Royal Society Interface*, vol. 5, no. 25, pp. 813–834, 2008.
- [9] H. C. Berg, *Random walks in biology*. Princeton University Press, 1983.
- [10] V. Dethier, “Patterns of locomotion of polyphagous arctiid caterpillars in relation to foraging,” *Ecological Entomology*, vol. 14, no. 4, pp. 375–386, 1989.
- [11] P. Turchin, “Translating foraging movements in heterogeneous environments into the spatial distribution of foragers,” *Ecology*, vol. 72, no. 4, pp. 1253–1266, 1991.
- [12] C. R. Sparrevohn, A. Nielsen, and J. G. Støttrup, “Diffusion of fish from a single release point,” *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 59, no. 5, pp. 844–853, 2002.

- [13] G. M. Viswanathan, S. V. Buldyrev, S. Havlin, M. Da Luz, E. Raposo, and H. E. Stanley, “Optimizing the success of random searches,” *nature*, vol. 401, no. 6756, p. 911, 1999.
- [14] F. Bartumeus, F. Peters, S. Pueyo, C. Marrasé, and J. Catalan, “Helical lévy walks: adjusting searching statistics to resource availability in microzooplankton,” *Proceedings of the National Academy of Sciences*, vol. 100, no. 22, pp. 12771–12775, 2003.
- [15] G. Ramos-Fernández, J. L. Mateos, O. Miramontes, G. Cocho, H. Larralde, and B. Ayala-Orozco, “Lévy walk patterns in the foraging movements of spider monkeys (*ateles geoffroyi*),” *Behavioral Ecology and Sociobiology*, vol. 55, no. 3, pp. 223–230, 2004.
- [16] J. Merkle, D. Fortin, and J. M. Morales, “A memory-based foraging tactic reveals an adaptive mechanism for restricted space use,” *Ecology letters*, vol. 17, no. 8, pp. 924–931, 2014.
- [17] C. H. Janson and R. Byrne, “What wild primates know about resources: opening up the black box,” *Animal Cognition*, vol. 10, no. 3, pp. 357–367, 2007.
- [18] W. F. Fagan, M. A. Lewis, M. Auger-Méthé, T. Avgar, S. Benhamou, G. Breed, L. Ladage, U. E. Schlägel, W.-w. Tang, Y. P. Papastamatiou, *et al.*, “Spatial memory and animal movement,” *Ecology letters*, vol. 16, no. 10, pp. 1316–1329, 2013.
- [19] C. Song, T. Koren, P. Wang, and A.-L. Barabási, “Modelling the scaling properties of human mobility,” *Nature Physics*, vol. 6, no. 10, p. 818, 2010.
- [20] X.-W. Wang, X.-P. Han, and B.-H. Wang, “Correlations and scaling laws in human mobility,” *PloS one*, vol. 9, no. 1, p. e84954, 2014.
- [21] A. O. Gautestad and I. Mysterud, “Complex animal distribution and abundance from memory-dependent kinetics,” *Ecological Complexity*, vol. 3, no. 1, pp. 44–55, 2006.
- [22] G. M. Schütz and S. Trimper, “Elephants can always remember: Exact long-range memory effects in a non-markovian random walk,” *Physical Review E*, vol. 70, no. 4, p. 045101, 2004.
- [23] J. C. Cressoni, M. A. A. da Silva, and G. Viswanathan, “Amnestically induced persistence in random walks,” *Physical Review Letters*, vol. 98, no. 7, p. 070603, 2007.
- [24] M. Serva, “Scaling behavior for random walks with memory of the largest distance from the origin,” *Physical Review E*, vol. 88, no. 5, p. 052141, 2013.
- [25] D. Boyer and P. Walsh, “Modelling the mobility of living organisms in heterogeneous landscapes: does memory improve foraging success?,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 368, no. 1933, pp. 5645–5659, 2010.
- [26] B. Davis, “Reinforced random walk,” *Probability Theory and Related Fields*, vol. 84, no. 2, pp. 203–229, 1990.

- [27] H. Othmer and A. Stevens, “Aggregation, blowup, and collapse: The abc’s of taxis in reinforced random walks,” *SIAM Journal on Applied Mathematics*, vol. 57, p. 1044–1081, 1997.
- [28] E. Bolthausen, U. Schmock, *et al.*, “On self-attracting  $d$ -dimensional random walks,” *The Annals of Probability*, vol. 25, no. 2, pp. 531–572, 1997.
- [29] J. Choi, J.-I. Sohn, K.-I. Goh, and I.-M. Kim, “Modeling the mobility with memory,” *EPL (Europhysics Letters)*, vol. 99, no. 5, p. 50001, 2012.
- [30] R. Pemantle, “A survey of random processes with reinforcement,” *Probability Surveys*, vol. 4, pp. 1–79, 2007.
- [31] E. Abrahams, *50 years of Anderson Localization*. World Scientific, 2010.
- [32] J. G. Foster, P. Grassberger, and M. Paczuski, “Reinforced walks in two and three dimensions,” *New Journal of Physics*, vol. 11, no. 2, p. 023009, 2009.
- [33] M. Evans and S. Majumdar, “Diffusion with stochastic resetting,” *Physical Review Letters*, vol. 106, no. 16, p. 160601, 2011.
- [34] G. Theraulaz and E. Bonabeau, “A brief history of stigmergy,” *Artificial Life*, vol. 5, no. 2, pp. 97–116, 1999.
- [35] E. Bonabeau, M. Dorigo, G. Theraulaz, *et al.*, *Swarm intelligence: from natural to artificial systems*. Oxford university press, 1999.
- [36] D. Schacter, “Priming and multiple memory systems: Perceptual mechanisms of implicit memory,” *Journal of Cognitive Neuroscience*, vol. 4, no. 3, pp. 244–256, 1992.
- [37] M. Bouton, *Learning and behaviour: A contemporary synthesis*. Sunderland, MA, US: Sinauer Associates, 2007.
- [38] N. Miller and S. Shettleworth, “Learning about environmental geometry: An associative model,” *Journal of Experimental Psychology: Animal Behavior Processes*, vol. 33, no. 3, pp. 191–212, 2007.
- [39] P. C. Trimmer, J. M. McNamara, A. I. Houston, and J. A. Marshall, “Does natural selection favour the rescorla–wagner rule?,” *Journal of theoretical biology*, vol. 302, pp. 39–52, 2012.
- [40] J. M. McNamara and A. I. Houston, “Integrating function and mechanism,” *Trends in Ecology & Evolution*, vol. 24, no. 12, pp. 670–675, 2009.
- [41] W. J. Grecian, J. V. Lane, T. Michelot, H. M. Wade, and K. C. Hamer, “Understanding the ontogeny of foraging behaviour: insights from combining marine predator bio-logging with satellite-derived oceanography in hidden markov models,” *Journal of the Royal Society Interface*, vol. 15, no. 143, p. 20180084, 2018.

- [42] C. W. Clark and M. Mangel, “The evolutionary advantages of group foraging,” *Theoretical Population Biology*, vol. 30, no. 1, pp. 45–75, 1986.
- [43] J. Krause and G. Ruxton, *Living in groups*. Oxford Univ. Press, 2002.
- [44] C. C. Ioannou, F. Bartumeus, J. Krause, and G. D. Ruxton, “Unified effects of aggregation reveal larger prey groups take longer to find,” *Proceedings of the Royal Society B: Biological Sciences*, vol. 278, no. 1720, pp. 2985–2990, 2011.
- [45] C. C. Ioannou, V. Guttal, and I. D. Couzin, “Predatory fish select for coordinated collective motion in virtual prey,” *Science*, vol. 337, no. 6099, pp. 1212–1215, 2012.
- [46] D. Sumpter, *Collective animal behaviour*. University Press, 2010.
- [47] A. Berdahl, C. J. Torney, C. C. Ioannou, J. J. Faria, and I. D. Couzin, “Emergent sensing of complex environments by mobile animal groups,” *Science*, vol. 339, no. 6119, pp. 574–576, 2013.
- [48] E. Mallon, S. Pratt, and N. Franks, “Individual and collective decision-making during nest site selection by the ant *leptothorax albipennis*,” *Behavioral Ecology and Sociobiology*, vol. 50, no. 4, pp. 352–359, 2001.
- [49] L. Conradt and T. Roper, “Consensus decision making in animals,” *Trends in Ecology & Evolution*, vol. 20, no. 8, pp. 449–456, 2005.
- [50] L. Giraldeau and T. Caraco, *Social foraging theory*. Princeton Univ. Press, 2000.
- [51] T. Pitcher, A. Magurran, and I. Winfield, “Fish in larger shoals find food faster,” *Behavioral Ecology and Sociobiology*, vol. 10, no. 2, pp. 149–151, 1982.
- [52] J. C. Haney, K. M. Fristrup, and D. S. Lee, “Geometry of visual recruitment by seabirds to ephemeral foraging flocks,” *Ornis Scand.*, vol. 23, no. 1, pp. 49–62, 1992.
- [53] L.-A. Giraldeau and G. Beauchamp, “Food exploitation: searching for the optimal joining policy,” *Trends in Ecology & Evolution*, vol. 14, no. 3, pp. 102–106, 1999.
- [54] L. Conradt and T. J. Roper, “Group decision-making in animals,” *Nature*, vol. 421, no. 6919, p. 155, 2003.
- [55] G. Kerth, “Group decision-making in animal societies,” in *Animal behaviour: Evolution and mechanisms*, pp. 241–265, Springer, 2010.
- [56] C. J. Torney, A. Berdahl, and I. D. Couzin, “Signalling and the evolution of cooperative foraging in dynamic environments,” *PLoS Computational Biology*, vol. 7, no. 9, pp. 1–10, 2011.
- [57] R. Martínez-García, J. M. Calabrese, T. Mueller, K. A. Olson, and C. López, “Optimizing the search for resources by sharing information: Mongolian gazelles as a case study,” *Phys. Rev. Lett.*, vol. 110, p. 248106, 2013.

- [58] S. R. Dall, L.-A. Giraldeau, O. Olsson, J. M. McNamara, and D. W. Stephens, “Information and its use by animals in evolutionary ecology,” *Trends in ecology & evolution*, vol. 20, no. 4, pp. 187–193, 2005.
- [59] M. Lihoreau, M. A. Charleston, A. M. Senior, F. J. Clissold, D. Raubenheimer, S. J. Simpson, and J. Buhl, “Collective foraging in spatially complex nutritional environments,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 372, no. 1727, p. 20160238, 2017.
- [60] T. D. Seeley, S. Camazine, and J. Sneyd, “Collective decision-making in honey bees: how colonies choose among nectar sources,” *Behavioral Ecology and Sociobiology*, vol. 28, p. 277–290, 1991.
- [61] R. Beckers, J.-L. Deneubourg, S. Goss, and J. M. Pasteels, “Collective decision making through food recruitment,” *Insectes Sociaux*, vol. 37, pp. 258–267, 1990.
- [62] M. Hauser and P. Marler, “Food-associated calls in rhesus macaques (*macaca mulatta*): I. socioecological factors,” *Behavioral Ecology*, vol. 4, no. 3, pp. 194–205, 1993.
- [63] C. R. Brown, M. B. Brown, and M. Shaffer, “Food-sharing signals among socially foraging cliff swallows,” *Animal Behaviour*, vol. 42, no. 4, pp. 551–564, 1991.
- [64] R. Frey, A. Gebler, K. A. Olson, D. Odonkhuu, G. Fritsch, N. Batsaikhan, and I. W. Stuermer, “Mobile larynx in mongolian gazelle: Retraction of the larynx during rutting barks in male mongolian gazelle (*procapra gutturosa pallas*, 1777),” *Journal of Morphology*, vol. 269, no. 10, pp. 1223–1237, 2008.
- [65] K. McComb, D. Reby, L. Baker, C. Moss, and S. Sayialel, “Long-distance communication of acoustic cues to social identity in african elephants,” *Animal Behaviour*, vol. 65, no. 2, pp. 317–329, 2003.
- [66] N. Mathevon, A. Koralek, M. Weldele, S. E. Glickman, and F. E. Theunissen, “What the hyena’s laugh tells: Sex, age, dominance and individual signature in the giggling call of *crocuta crocuta*,” *BMC Ecology*, vol. 10, p. 9, Mar 2010.
- [67] L. J. Brent, D. W. Franks, E. A. Foster, K. C. Balcomb, M. A. Cant, and D. P. Croft, “Ecological knowledge, leadership, and the evolution of menopause in killer whale,” *Current Biology*, vol. 25, no. 6, pp. 746–750, 2015.
- [68] T. Palacios-Romo, F. Castellanos, and G. Ramos-Fernandez, “Uncovering the decision rules behind collective foraging in spider monkeys,” *Animal Behaviour*, vol. 149, pp. 121–133, 2019.
- [69] I. D. Couzin, J. Krause, R. James, G. D. Ruxton, and N. R. Franks, “Collective memory and spatial sorting in animal groups,” *Journal of theoretical biology*, vol. 218, no. 1, pp. 1–11, 2002.
- [70] D. Watts, “Mountain gorilla habitat use strategies and group movements,” *On the move. How and why animals travel in groups*, pp. 351–374, 2000.

- [71] M. B. Hooten, D. S. Johnson, B. T. McClintock, and J. M. Morales, *Animal movement: statistical models for telemetry data*. CRC press, 2017.
- [72] F. Cagnacci, L. Boitani, R. A. Powell, and M. S. Boyce, “Animal ecology meets gps-based radiotelemetry: a perfect storm of opportunities and challenges,” 2010.
- [73] M. Hooten and T. Helfey, *Bringing Bayesian Models to Life*. Chapman and Hall, 2019.
- [74] A. Gelman, H. S. Stern, J. B. Carlin, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian data analysis*. Chapman and Hall/CRC, 2013.
- [75] G. Oster and B. Heinrich, “Why do bumblebees major? a mathematical model,” *Ecological Monographs*, vol. 46, no. 2, pp. 129–133, 1976.
- [76] T. J. Valone and J. S. Brown, “Measuring patch assessment abilities of desert granivores,” *Ecology*, vol. 70, no. 6, pp. 1800–1810, 1989.
- [77] D. J. Anderson and P. J. Hodum, “Predator behavior favors clumped nesting in an oceanic seabird,” *Ecology*, vol. 74, no. 8, pp. 2462–2464, 1993.
- [78] P. Dixon and A. M. Ellison, “Introduction: ecological applications of bayesian inference,” *Ecological Applications*, vol. 6, no. 4, pp. 1034–1035, 1996.
- [79] R. Hilborn and M. Mangel, *The ecological detective: confronting models with data*, vol. 28. Princeton University Press, 1997.
- [80] J. M. Morales, D. T. Haydon, J. Frair, K. E. Holsinger, and J. M. Fryxell, “Extracting more out of relocation data: building movement models as mixtures of random walks,” *Ecology*, vol. 85, no. 9, pp. 2436–2445, 2004.
- [81] D. Coppersmith and P. Diaconis, “Random walk with reinforcement,” *Unpublished manuscript*, 1987.
- [82] R. Pemantle *et al.*, “Phase transition in reinforced random walk and rwre on trees,” *The Annals of Probability*, vol. 16, no. 3, pp. 1229–1241, 1988.
- [83] L. Börger, B. D. Dalziel, and J. M. Fryxell, “Are there general mechanisms of animal home range behaviour? a review and prospects for future research,” *Ecology Letters*, vol. 11, no. 6, pp. 637–650, 2008.
- [84] J. R. Carpenter *et al.*, “An ecological glossary,” *An ecological glossary.*, 1938.
- [85] D. Boyer and C. Solis-Salas, “Random walks with preferential relocations to places visited in the past and their application to biology,” *Physical Review Letters*, vol. 112, no. 24, p. 240601, 2014.
- [86] D. Boyer and I. Pineda, “Slow lévy flights,” *Phys. Rev. E*, vol. 93, p. 022103, 2016.
- [87] R. Metzler and J. Klafter, “The random walk’s guide to anomalous diffusion: a fractional dynamics approach,” *Physics reports*, vol. 339, no. 1, pp. 1–77, 2000.

- [88] M. Vendruscolo and M. Marsili, “Diffusion in disordered media as a process with memory,” *Physical Review E*, vol. 54, no. 2, p. R1021, 1996.
- [89] Y. G. Sinai, “The limiting behavior of a one-dimensional random walk in a random medium,” *Theory of Probability & Its Applications*, vol. 27, no. 2, pp. 256–268, 1983.
- [90] L. Kusmierz, S. N. Majumdar, S. Sabhapandit, and G. Schehr, “First order transition for the optimal search time of lévy flights with resetting,” *Phys. Rev. Lett.*, vol. 113, no. 22, p. 220602, 2014.
- [91] L. Kusmierz and E. Gudowska-Nowak, “Optimal first-arrival times in lévy flights with resetting,” *Phys. Rev. E*, vol. 92, p. 052127, 2015.
- [92] C. Mailler and G. U. Bravo, “Random walks with preferential relocations and fading memory: a study through random recursive trees,” *arXiv preprint arXiv:1810.02735*, 2018.
- [93] K. Bhattacharya and T. Vicsek, “Collective foraging in heterogeneous landscapes,” *Journal of The Royal Society Interface*, vol. 11, p. 20140674, 2014.
- [94] I. D. Jonsen, R. A. Myers, and M. C. James, “Identifying leatherback turtle foraging behaviour from satellite telemetry using a switching state-space model,” *Marine Ecology Progress Series*, vol. 337, pp. 255–264, 2007.
- [95] N. T. Hobbs and M. B. Hooten, *Bayesian models: a statistical primer for ecologists*. Princeton University Press, 2015.
- [96] J. G. C. Hopcraft, J. M. Morales, H. Beyer, M. Borner, E. Mwangomo, A. Sinclair, H. Olf, and D. T. Haydon, “Competition, predation, and migration: individual choice patterns of serengeti migrants captured by hierarchical models,” *Ecological Monographs*, vol. 84, no. 3, pp. 355–372, 2014.
- [97] C. E. Vincenot, S. Mazzoleni, K. Moriya, F. Cartenì, and F. Giannino, “How spatial resource distribution and memory impact foraging success: a hybrid model and mechanistic index,” *Ecological Complexity*, vol. 22, pp. 139–151, 2015.
- [98] W. Feller, *An introduction to probability theory and its applications*. John Wiley & Sons, New York, 2008.
- [99] L. E. Reichl, *A modern course in statistical physics*. John Wiley & Sons, 2016.
- [100] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*. Elsevier/Academic Press, Amsterdam, 2015.
- [101] G. H. Weiss, *Aspect and applications of the random walk*. Elsevier, Amsterdam, 1994.
- [102] M. Evans and S. Majumdar, “Diffusion with resetting in arbitrary spatial dimension,” *Journal of Physics A: Mathematical and Theoretical*, vol. 47, no. 28, p. 285001, 2014.

- [103] N. Goldenfeld, *Lectures on phase transitions and the renormalization group*. Addison-Wesley, Reading, MA, 1992.
- [104] S. Kotz, I. Ostrovskii, and A. Hayfavi, “Analytic and asymptotic properties of linnik’s probability densities. i.,” *Journal of Mathematical Analysis and Applications*, vol. 193, pp. 353–371, 1995.
- [105] D. Vollhardt and P. Wölfle, “Scaling equations from a self-consistent theory of anderson localization,” *Phys. Rev. Lett.*, vol. 48, pp. 699–702, 1982.
- [106] D. Vollhardt and P. Wölfle, “Diagrammatic, self-consistent treatment of the anderson localization problem in  $d \leq 2$  dimensions,” *Phys. Rev. B*, vol. 22, pp. 4666–4679, 1980.
- [107] J. Krause, G. D. Ruxton, and S. Krause, “Swarm intelligence in animals and humans,” *Trends in Ecology & Evolution*, vol. 25, no. 1, pp. 28–34, 2010.
- [108] J. Surowiecki, *The Wisdom of Crowds*. New York, NY: Anchor., 2005.
- [109] K. von Frisch, *The Dance Language and Orientation of Bees*. Harvard Univ. Press. Cambridge., 1967.
- [110] D. J. Hoare, I. D. Couzin, J.-G. Godin, and J. Krause, “Context-dependent group size choice in fish,” *Animal Behaviour*, vol. 67, no. 1, pp. 155–164, 2004.
- [111] I. D. Couzin, J. Krause, N. R. Franks, and S. A. Levin, “Effective leadership and decision-making in animal groups on the move,” *Nature*, vol. 433, pp. 513–516, 2005.
- [112] T. Sasaki and D. Biro, “Cumulative culture can emerge from collective intelligence in animal groups,” *Nature Communications volume*, vol. 8, no. 15049, 2017.
- [113] A. B. Kao, N. Miller, C. Torney, A. Hartnett, and I. D. Couzin, “Collective learning and optimal consensus decisions in social animal groups,” *PLoS Computational Biology*, vol. 10, no. 8, pp. 1–11, 2014.
- [114] A. Gautestad and I. Mysterud, “Intrinsic scaling complexity in animal dispersion and abundance.,” *The American Naturalist*, vol. 165, no. 1, pp. 44–55, 2005.
- [115] F. Aureli, C. M. Schaffner, C. Boesch, S. K. Bearder, J. Call, C. A. Chapman, R. Connor, A. D. Fiore, R. I. Dunbar, S. P. Henzi, *et al.*, “Fission-fusion dynamics: New research frameworks,” *Current Anthropology*, vol. 49, no. 4, pp. 627–654, 2008.
- [116] G. Ramos-Fernandez, D. Boyer, F. Aureli, and L. G. Vick, “Association networks in spider monkeys (*ateles geoffroyi*),” *Behavioral Ecology and Sociobiology*, vol. 63, p. 999–1013, 2009.
- [117] A. Brabási, *Network science*. Cambridge University Press., 2016.
- [118] M. Newman, *Networks: An introduction*. Oxford University Press., 2010.

- [119] M. E. Newman, S. H. Strogatz, and D. J. Watts, “Random graphs with arbitrary degree distributions and their applications,” *Phys. Rev. E*, vol. 64, p. 026118, 2001.
- [120] D. Lusseau and M. Newman, “Identifying the role that animals play in their social networks,” *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 271, p. S477–S481, 2004.
- [121] N. Pinter-Wollman, E. A. Hobson, J. E. Smith, A. J. Edelman, D. Shizuka, S. De Silva, J. S. Waters, S. D. Prager, T. Sasaki, G. Wittemyer, *et al.*, “The dynamics of animal social networks: analytical, conceptual, and theoretical advances,” *Behavioral Ecology*, vol. 25, no. 2, pp. 242–255, 2014.
- [122] P. Kanngiesser, C. Sueur, K. Riedl, J. Grossmann, and J. Call, “Grooming network cohesion and the role of individuals in a captive chimpanzee group,” *American Journal of Primatology*, vol. 73, no. 8, pp. 758–767, 2011.
- [123] F. Bartumeus, M. G. E. da Luz, G. M. Viswanathan, and J. Catalan, “Animal search strategies: a quantitative random-walk analysis,” *Ecology*, vol. 86, no. 11, pp. 3078–3087, 2005.
- [124] F. Bartumeus and J. Catalan, “Optimal search behavior and classic foraging theory,” *Journal of Physics A: Mathematical and Theoretical*, vol. 42, no. 43, p. 434002, 2009.
- [125] M. L. Platt, E. M. Brannon, T. L. Briese, and J. A. French, “Differences in feeding ecology predict differences in performance between golden lion tamarins (*Leontopithecus rosalia*) and Wied’s marmosets (*Callithrix kuhli*) on spatial and visual memory tasks,” *Animal Learning & Behavior*, vol. 24, no. 4, pp. 384–393, 1996.
- [126] R. Johnson, “Learning, memory, and foraging efficiency in two species of desert seed-harvester ants,” *Ecology*, vol. 72, no. 4, pp. 1408–1419, 1991.
- [127] K. Cheng, A. Narendra, and R. Wehner, “Behavioral ecology of odometric memories in desert ants: acquisition, retention, and integration,” *Behavioral Ecology*, vol. 17, no. 2, pp. 227–235, 2005.
- [128] L. Male and T. Smulders, “Memory decay and cache site preferences in hoarding coal tits—a laboratory study,” *Behaviour*, vol. 144, no. 6, pp. 693–710, 2007.
- [129] D. Boyer and J. Romo-Cruz, “Solvable random-walk model with memory and its relations with Markovian models of anomalous diffusion,” *Physical Review E*, vol. 90, no. 4, p. 042136, 2014.
- [130] E. A. Sánchez-Hernández, “Caminatas aleatorias con memoria: Un modelo simple para el forrajeo animal,” Master’s thesis, Universidad Nacional Autónoma de México, 2016.
- [131] A. E. Gelfand and A. F. Smith, “Sampling-based approaches to calculating marginal densities,” *Journal of the American Statistical Association*, vol. 85, no. 410, pp. 398–409, 1990.

- [132] A. W. F. Edwards, *Likelihood*. CUP Archive, 1984.
- [133] A. Azzalini, *Statistical inference based on the likelihood*. Routledge, 2017.
- [134] S. Geman and D. Geman, “Stochastic relaxation, gibbs distributions, and the bayesian restoration of images,” *IEEE Transactions on pattern analysis and machine intelligence*, no. 6, pp. 721–741, 1984.
- [135] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, “Equation of state calculations by fast computing machines,” *The journal of chemical physics*, vol. 21, no. 6, pp. 1087–1092, 1953.
- [136] W. K. Hastings, “Monte carlo sampling methods using markov chains and their applications,” *Biometrika*, vol. 51, no. 1, p. 97–109, 1970.
- [137] D. J. Lunn, A. Thomas, N. Best, and D. Spiegelhalter, “Winbugs-a bayesian modelling framework: concepts, structure, and extensibility,” *Statistics and computing*, vol. 10, no. 4, pp. 325–337, 2000.
- [138] D. Lunn, D. Spiegelhalter, A. Thomas, and N. Best, “The bugs project: Evolution, critique and future directions,” *Statistics in medicine*, vol. 28, no. 25, pp. 3049–3067, 2009.
- [139] D. Lunn, J. Barrett, M. Sweeting, and S. Thompson, “Fully bayesian hierarchical modelling in two stages, with application to meta-analysis,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 62, no. 4, pp. 551–572, 2013.
- [140] P. R. Jadhav and J. V. Gobburu, “A new equivalence based metric for predictive check to qualify mixed-effects models,” *The AAPS journal*, vol. 7, no. 3, pp. E523–E531, 2005.
- [141] A. Gelman, J. Hwang, and A. Vehtari, “Understanding predictive information criteria for bayesian models,” *Statistics and computing*, vol. 24, no. 6, pp. 997–1016, 2014.
- [142] J. M. Bernardo, “Expected information as expected utility,” *the Annals of Statistics*, pp. 686–690, 1979.
- [143] H. Akaike, “Maximum likelihood identification of gaussian autoregressive moving average models,” *Biometrika*, vol. 60, no. 2, pp. 255–265, 1973.
- [144] S. Watanabe, “Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory,” *Journal of Machine Learning Research*, vol. 11, no. Dec, pp. 3571–3594, 2010.
- [145] J. L. Frair, E. H. Merrill, J. R. Allen, and M. S. Boyce, “Know thy enemy: experience affects elk translocation success in risky landscapes,” *The Journal of Wildlife Management*, vol. 71, no. 2, pp. 541–554, 2007.

- [146] M. Wolf, J. Frair, E. Merrill, and P. Turchin, “The attraction of the known: the importance of spatial familiarity in habitat selection in wapiti cervus elaphus,” *Ecography*, vol. 32, no. 3, pp. 401–410, 2009.
- [147] J. L. Frair, E. H. Merrill, D. R. Visscher, D. Fortin, H. L. Beyer, and J. M. Morales, “Scales of movement by elk (cervus elaphus) in response to heterogeneity in forage resources and predation risk,” *Landscape ecology*, vol. 20, no. 3, pp. 273–287, 2005.
- [148] <https://www.esri.com/es-es/arcgis/about-arcgis/overview>.
- [149] B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell, “Stan: A probabilistic programming language,” *Journal of statistical software*, vol. 76, no. 1, 2017.
- [150] S. D. Team, “Rstan: the r interface to stan. r package version 2.14. 1,” 2016.
- [151] M. Betancourt, “A conceptual introduction to hamiltonian monte carlo,” *arXiv preprint arXiv:1701.02434*, 2017.
- [152] B. Van Moorter, D. Visscher, S. Benhamou, L. Börger, M. S. Boyce, and J.-M. Gaillard, “Memory keeps you at home: a mechanistic model for home range emergence,” *Oikos*, vol. 118, no. 5, pp. 641–652, 2009.
- [153] C. Kasper and B. Voelkl, “A social network analysis of primate groups,” *Primates*, vol. 50, p. 343–356, 2009.
- [154] K. D. Carter, R. Brand, J. K. Carter, B. Shorrocks, and A. W. Goldizen, “Social networks, long-term associations and age-related sociability of wild giraffes,” *Animal Behaviour*, vol. 86, no. 5, pp. 901–910, 2013.